Gene expression profiling of brain regions in inbred mouse strains reveals candidate genes for phenotypic variation.

Rickard Sandberg
Stockholm, 2000

Karolinska Institutet
Master thesis in biomedicine
Examensarbete, 20p

Performed at:
The Salk Institute for Biological Studies
The Laboratory of Genetics
10010 North Torrey Pines Road
La Jolla, CA 92037
Supervisor: Carrolee Barlow, M.D. Ph.D.

Supervisor, KI: Claes Wahlestedt, M.D. Ph.D.
Karolinska Institutet
Center for Genomic Research

Table of Contents

Appendix 1: Figures and tables for result and discussion section.

Appendix 2: Quality controls for hybridization performance.

Appendix 3: Affymetrix GeneChip algorithms

ABSTRACT

Behavioral analysis of mouse strains generated using transgenic technology is becoming increasingly common as a method for understanding the molecular basis of behavior. However, the phenotype of mutants maintained on different genetic backgrounds is variable and studies in non-isogenic strains can be difficult to interpret. We used gene expression profiling of multiple brain regions in two commonly used inbred mouse strains to determine if differing gene expression patterns might contribute to the phenotypic variability in these strains. The arrays used monitor approximately 10,000 genes (13,069 probe sets). We found that 7,169 of the probe sets (~55%) were expressed in at least one brain region. Of these, 73 (73/7,169 or 1.0%) were differentially expressed between the two strains. If the percent of differentially expressed genes identified using these arrays is representative, then approximately 1,000 genes, out of the estimated 100,000, would be differentially expressed due to changes in genetic background. This suggests that gene expression studies in non-isogenic transgenic strains of mice could create a significant amount of false positive data not related to the particular mutation, but rather to inherited differences in gene expression between strains. These results lay a foundation for interpreting gene expression profiling when mutants differ in genetic background. In addition, these data provide candidate genes that can be evaluated to investigate their role in modulating the distinct behavioral phenotypes between these strains of mice. Finally, these types of analysis will allow us to apply a systems approach for asking how multiple, subtle molecular changes act in concert to give rise to a particular phenotype.

Abbreviations

ADC	Average Difference Change
Ag	Amygdala
Cb	Cerebellum
Cx	Cortex
Ec	Entorhinal cortex
Hp	Hippocampus
Mb	Midbrain
MEF	Mouse Embryonic fibroblasts

Introduction

Recent advances in mouse genetics have opened up new avenues in the field of neurobehavioral genetics. By generating targeted mutations in genes (null mutants) or overexpressing genes (transgenics), many novel behavioral phenotypes have been observed. Much of the recent focus has been on how single gene defects result in specific alterations in behavior. Mice generated in these studies have a variety of phenotypes including decreased or enhanced learning and memory (Abeliovich et al., 1993; Mayford et al., 1995; Tang et al., 1999), difficulty performing motor tasks (Barlow et al., 1996) and differing sensitivity to drugs of addiction (Phillips, 1997).

The neurobehavioral phenotype of a particular mouse results not only from the specific alteration induced by a targeted mutation, the mis-expression of a particular gene or the administration of a particular drug, but also from the effects of genetic modifiers, which may differ significantly based on genetic background. Examples of this come from studies focused less on single gene defects and more on the constellation of genetic arrangements that account for significant differences between inbred strains of mice. For example, neurogenesis after exposure to an enriched environment differs substantially between the C57BL/6 and 129SvJ mouse strains (Kempermann et al., 1997; Kempermann et al., 1998). Other studies showed that despite similar seizure susceptibility, various inbred strains exhibited large differences in neuronal cell death after seizures (Schauwecker and Steward, 1997). At the level of behavior, inbred strains of mice vary greatly in their behavioral response to drugs of addiction, such as ethanol (McClearn and Rodgers, 1959; Metten et al., 1998) and also show marked differences in some types of behavioral testing, such as prepulse inhibition (Crawley et al., 1997). Finally, it has been demonstrated that single gene mutations, “knockouts”, can result in substantially different phenotypes depending on the background genetic strain on which the mutation is maintained (for a review see Gerlai, 1996). These studies suggest that effects reported in gene targeting studies might be due to the genetic background of the hybrids with the induced mutation rather than the particular genetic mutation alone (Crawley, 1996; Crusio, 1996; Dawson et al., 1996; Gerlai, 1996; Lathe, 1996; Morrison et al., 1996; Watanabe et al., 1996). Because behaviors are influenced by many factors ranging from the environment to specific gene interactions, it is increasingly important to consider candidate genes or mutations in light of the multitude of potential modifiers.

Embryonic stem cells are derived from 129 strains, most commonly 129Sv (see Simpson et al., 1997; Threadgill et al., 1997 for a review on the revised nomenclature of 129 strains) and are part of the genetic background of most mutants generated using homologous recombination. C57BL/6 is the strain most commonly used for outcrossing, the background strain of many spontaneous mutants, and is used in many drug and neurobehavioral studies. With the advent of gene expression arrays (Lockhart et al., 1996; Wodicka et al., 1997; Lipshutz et al., 1999), it is now possible to study inbred strains of mice and ask the question: what is the interacting array of genes that might account for the differences between inbred mouse strains? We have used gene expression profiling of multiple brain regions and control mouse embryonic fibroblasts from these two commonly used inbred strains, C57BL/6 and 129SvEv.

Aim

Use gene expression profiling of multiple brain regions in two commonly used inbred mouse strains (C57BL/6 and 129SvEv) to determine if differing gene expression patterns might contribute to the phenotypic variability in these strains.

Materials and methods

Mice

Male C57BL/6 and 129SvEv mice were purchased from Taconic (New York) at an age of 7 weeks and housed individually for 1 week before sacrifice. Dissections were carried out between 14.00-17.00 hour. After removal of the brain, further dissection was performed on a paraffin covered petri dish filled with wet ice. The cortex dissection included the entire cortex except the olfactory bulbs (therefore, cortex in this dissection includes the medial temporal lobe structures). The cerebellum was dissected free of the brainstem. The midbrain was dissected free of the cortex and the brainstem. The hippocampus was removed after cutting the cortex sagitally and removing the entire structure with a paintbrush. The remaining tissue was discarded. Cortex (Cx), cerebellum (Cb), midbrain (Mb) and hippocampus (Hp) were prepared in duplicates from two different mice of each strain. In order to obtain sufficient tissue for RNA purification from amygdala (Ag) and entorhinal cortex (Ec) the microdissected regions of 7 animals were pooled from each strain respectively. The area lateral to and spanning from 3 mm ventral to the bifurcation of the external capsule was used as a landmark to define the borders of dissection for the entorhinal cortex. The area enclosed between the bifurcation of the external capsule was used to demarcate the dissection plane for the amygdala. The dissected tissues were directly frozen on dry ice and stored at –80°C. Mouse embryonic fibroblasts were prepared according to standard protocols from 6 embryos at day 13.5 for each strain (Hogan et al., 1994).

High density oligonucleotide arrays

High density oligonucleotide arrays (GeneChip, Affymetrix) is a direct and highly parallel approach to monitor gene expression levels. The arrays contain up to hundreds of thousands of different oligonucleotides spatially patterned on a small glass surface. The oligonucleotides are synthesized directly on the glass surface using photolithography and solid-phase DNA synthesis. Nucleotides are covalently linked to a photolabile protecting group. Light is directed through a photolithographic mask to specific areas of the array to produce localized photodeprotection (Figure 1). Nucleotides with the covalent linker are incubated with the array surface and chemical coupling occurs at the photodeprotected sites. This chemical cycle is repeated. This approach enables, in theory, the synthesis of a complete set of 4^N oligonucleotides with the length of N nucleotides to be synthesized in $4 \times N$ cycles.

A fluorescently labeled mRNA population is incubated with the array surface and mRNA transcripts hybridize to complementary oligonucleotides on the array. The level of bound mRNA transcripts are measured using a confocal scanner (Lockhart et al., 1996; Wodicka et al., 1997; Lipshutz et al., 1999).

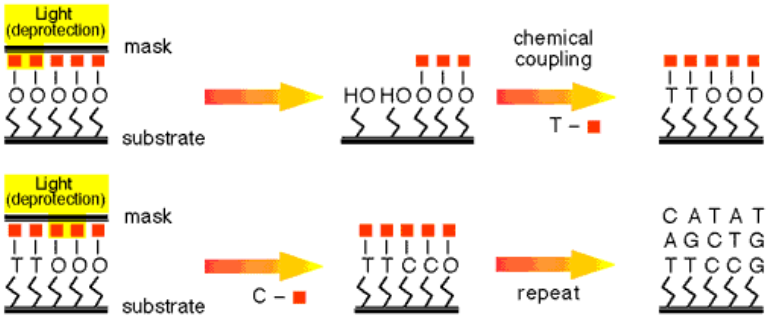

Figure 1. Oligonucleotide synthesis is performed directly on the array's glass surface. Light is directed through a mask and deprotects specific areas of the array. In the next step, nucleotides are added and chemically react with the deprotected oligonucleotides. This process is repeated. Adapted from Lipshutz et al., 1999.

Total RNA purification from mouse tissues

Stored tissues were rapidly placed into TRIzol (GIBCO-BRL) (added at approximately 1 ml per 100 mg tissue to the frozen tissues) and homogenized (Polytron, Kinematica) at maximum speed for 90-120 sec. The subsequent steps were done according to the manufacturer's instructions (TRIzol, GIBCO-BRL). RNA was resuspended in RNase-free water at a concentration of 1 mg/ml.

Preparation of Labeled Targets for Hybridization

To prepare cRNA for hybridization, 10μg of total RNA was denatured at 70°C for 10 minutes with 10 μM T7-tagged oligo-dT primer (GGCCAGTGAATTGTAATACG-ACTCACTATAGGGAGGCGG-T(24)), cooled on ice 5 minutes, then heated to 42°C for 2 minutes in 1X first strand buffer, 10 mM DTT, and 0.5 mM dNTP (each). Reverse transcription was performed with 400 U SuperScript II at 42°C for 1 hour in a total volume of 20 μl (all reagents GIBCO-BRL). Second strand cDNA was synthesized by adding 30 μl 5X second strand buffer, 3 μl 10mM dNTP (each), 10 U E. *coli* DNA Ligase, 40 U E. *coli* DNA pol I, 2 U E. *coli* RNase H, and DEPC-treated H_2O to 150 μl total volume and incubated at 16°C for 2 hours (all reagents GIBCO-BRL). The cDNA was blunt-ended with 10 U T4 DNA pol (GIBCO-BRL) at 16°C for 5 minutes then stopped with addition of 10 μl 0.5 M EDTA on ice. The synthesized cDNA was extracted once with an equal volume of phenol:chloroform and aqueous phase recovered using Phase-Lock Gel (5 Prime-3 Prime, Inc., Boulder, CO). Double stranded cDNA was ethanol precipitated and resuspended in 20 μl of distilled water. 10 μl of cDNA was used for *in vitro* transcription using ENZO BioArray Labeling kit according to the manufacturer's instructions. Biotinylated cRNA was purified using RNeasy columns (Qiagen). Biotinylated cRNA quality was checked with both a spectrophotometer and agarose gel electrophoresis. Biotinylated cRNA was fragmented in 40 mM Tris-acetate pH 8.1, 100 mM KOAc and 30 mM MgOAc for 35 minutes at 94°C. Fragmented cRNA was then brought up to a volume of 300 μl of hybridization solution with final concentrations of 0.05 μg/μl cRNA, 0.1 mg/ml herring sperm DNA (Fisher) and 0.5 mg/ml acetylated BSA (GIBCO-BRL) in 1X hybridization buffer (1M NaCl, 100mM MES pH 6.5, 0.01% triton X-100).

Array Hybridization

The biotinylated samples were denatured for 5 minutes at 99°C, incubated 5 minutes at 45°C and centrifuged at 16,000 x g for 5 minutes to pellet debris. Two different oligonucleotide arrays (GeneChip, Affymetrix) were used that together represent 13,069 probe sets corresponding to more than 10,000 unique genes and ESTs (Mu11ksubA and Mu11ksubB). The array cartridges were prehybridized in 1X hybridization buffer with 0.5 mg/ml acetylated BSA and 0.1 mg/ml Herring Sperm DNA for 15 minutes at 45°C, 60 rpm on a rotisserie (rotating hybridization oven from Affymetrix). Prehybridization solution was then removed and 200 µl of the sample was added to each cartridge and hybridized for 16 hours at 45°C, 60 rpm. After hybridization, the sample was recovered and the cartridges were washed with wash solution (6X SSPE, 0.01% Triton x-100) on a fluidics station (Affymetrix). The cartridges were rinsed and incubated with 200 µl high stringency wash buffer (0.1M NaCl, 100 mM MES, 0.01% Triton x-100) for 30 minutes at 45°C, 60 rpm. After removing the solution, 200 µl of staining solution (1X hybridization buffer, with 2.5 mg/ml acetylated BSA (Sigma) and 10 µg/ml streptavidin R-phycoerythrin (Molecular Probes, Eugene, OR) were added and the cartridge incubated 15 minutes at 37°C, 60 rpm. After staining, cartridges were washed on a fluidics station and incubated with 200 µl antibody solution (1X hybridization buffer with 0.5 mg/ml acetylated BSA (Sigma) and 1 µg/ml goat biotinylated-Anti-streptavidin antibody (Vector Labs) for 30 minutes at 37°C, 60 rpm. The cartridges were then washed with wash solution on the fluidics station, then incubated for 15 minutes at 37°C rotating at 60 rpm in 200 µl staining solution. Again the cartridges were washed on the fluidics station. Arrays were scanned using a Hewlett-Packard GeneArray confocal scanner using GeneChip 3.1 software (Affymetrix).

Quality controls in sample preparation and array hybridization

The process from total RNA to labeled and fragmented cRNA includes quality control checkpoints. The quality of the total RNA is a key factor for successful gene expression analysis and is checked by electrophoresis and by measuring absorbance in a spectrophotometer. The stained RNA on the gel should show no sign of degradation and have a 2:1 ratio between 28S and 18S rRNA. The absorbance was measured both in H_2O and in TE, because we believe H_2O is more accurate for determining concentration and TE for the purity of the RNA. Only RNA with a absorbance ratio (260/280) over 2.0 (in TE) were further used in this study.

After the *in vitro* transcription the cRNA was checked using electrophoresis and a size distribution between 500 and 2000 bp is expected (Figure 2).

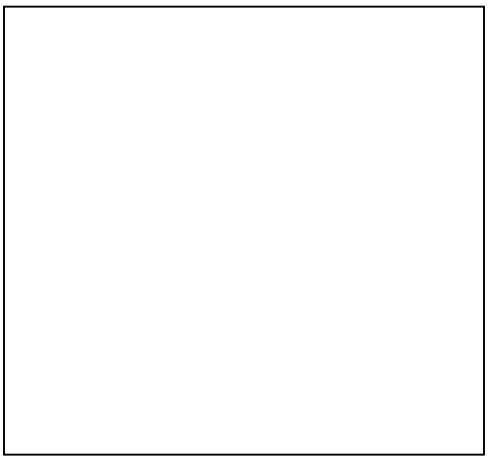

Figure 2. cRNA run on an agarose gel.
Expected size: 500 - 2000 bp. Lane 1,2,3,4 and 6 are examples of good quality cRNA, but in lane 5 it is degraded (lacking its larger fragments). Marker, 1kb plus DNA ladder (Gibco-BRL), are present in lane 7. Arrows indicate desired range.

The cRNA is fragmented in order to increase efficiency in binding to the array 25 oligomers. Fragmented cRNA is checked with polyacrylamide gelelectrophoresis and expected size distribution ranges from 20 to 70 bp (Figure 3).

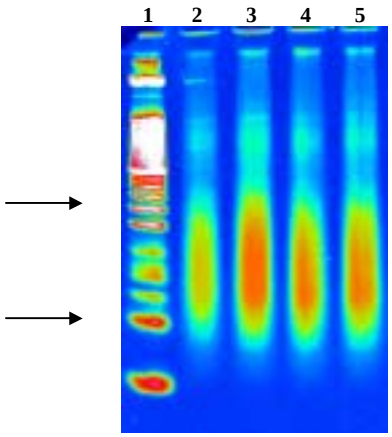

Figure 3. Fragmented cRNA on denaturing PAGE gel.
Lane 1: 10 bp marker (GIBCO-BRL). Lane 2-5 samples. Expected size: ~20-70 bp.
Arrows indicates the range from 20-70bp.

All samples that fulfilled the above criteria were applied to a ”test array” (Test2, Affymetrix) in order to further determine that the labeled samples were of good quality. The ”test array” consists of only control oligos and is used to determine the hybridization performance of a labeled sample. Even though the sample passed previous checkpoints it has happened that a sample performs poorly when hybridizing to an array (no clear cause). In order to evaluate hybridization performance, analysis of hybridization parameters is performed: Qraw(noise), background, standard deviation of background, scalar factor, degradation controls (Actin and GAPDH 3’/5’ ratio). All parameters are defined in appendix 2. The result from this analysis on the Mu11ksubA and Mu11ksubB arrays are presented in Appendix 2.

Data Analysis

General

Data analysis was performed using GeneChip version 3.1 (Affymetrix) and NFueGGo 2.1c (Lockhart and Lockhart, Genomics institute of the Novartis research Foundation). We used the GeneChip software global scaling algorithm in order to compare all 24 samples (48 total arrays-24 SubA and 24 SubB arrays). In global scaling, the output of any experiment is multiplied by a factor (the Scaling Factor, SF) to make the average fluorescence intensity across the entire array (after subtraction of background) equal to a target intensity set by the user. Scaling normalizes a number of experiments to one target intensity, allowing comparison between any two experiments. In our analysis, we scaled all samples to a target intensity of 200. A target intensity of 200 has been shown to correspond to 3-5 transcripts per cell (Wodicka et al., 1997 and unpublished data). This permits correlation of the hybridization signal to the transcript copy number in a sample. To ensure the quality of the experiment, any array that required a scaling factor of greater than 2 standard deviations from the mean were not used and the

experiment was repeated on a replacement array. All strain variation analysis were performed by comparing C57BL/6 to 129SvEv. Therefore, all x-fold changes are calculated as the ratio of expression between C57BL/6 to 129SvEv, hence a positive value indicates higher expression in the C57BL/6 relative to the 129SvEv sample(s) and a negative value indicates higher expression in the 129SvEv relative to the C57BL/6 sample(s).

In order to increase confidence in the results, as well as the quantitative accuracy, all experiments were performed in duplicate (independent region sources, independent sample preparations and independent arrays). When comparing two different samples to identify differentially expressed genes, duplicate comparisons (e.g., A vs. B and A' vs. B') were performed, and only differences that were consistent in both comparisons were considered further. The magnitude of the difference or ratio is calculated as the arithmetic mean of the values obtained in the duplicate measurements. Based on the analysis of replicate measurements (independent mice, independent sample preparations, independent arrays) the number of "false positives", defined as genes scoring incorrectly as "increase" or "decrease" with a two-fold change or greater, was very low (2/13,069 or 0.017%).

Strain variation throughout all brain regions

All C57BL/6 samples were compared to all 129SvEv samples (12 comparisons-6 regions per strain performed in duplicate). The criteria used to detect differences in gene expression were a difference call of "increase", "marginal increase", "decrease", "marginal decrease" in 8/12 comparisons, a fold change >= 1.8 in 8/12 comparisons and an average difference change (ADC) >=50 in 8/12 comparisons.

Brain region specific differences between mouse strains

C57BL/6 samples were compared to the corresponding 129SvEv brain region. The criteria used to detect differences in gene expression were a difference call of "increase", "marginal increase", "decrease", "marginal decrease"; a fold change >= 1.8 and an ADC >=50 in both comparisons.

Brain region specific gene expression

To identify genes with region-restricted gene expression, we performed an analysis based on the absolute analysis of each brain region. Genes were classified as "present" in a region if the gene had an absolute call of "present" in at least three out of four samples. Similarly, we used an absolute call of "absent" in four out of four brain regions and these were classified as clearly not detected (absent or expression at levels below the threshold of detection). The average difference from one brain region was then compared to all other brain regions and genes with significant differences were included (p=0.05 using a student's T-test). This data was used to generate Venn diagrams representing overlapping and non-overlapping gene expression patterns (Figure 5 in Appendix 1).

To detect region specific variation (both restriction and enrichment) in gene expression, the following criteria (Table 1) were used with the additional criteria that at least one sample must be scored as "present".

Comparison	Difference Call {I,MI,D,MD}	Fold Change	ADC
Cx / Cb, Mb	7/8	>1.8 in 7/8	>50 in 7/8
Cb / Cx, Mb, Hp, Ag, Ec	18/20	>1.8 in 18/20	>50 in 18/20
Mb / Cx, Cb, Hp, Ag, Ec	18/20	>1.8 in 18/20	>50 in 18/20
Hp / Cb, Mb, Ag, Ec	14/16	>1.8 in 14/16	>50 in 14/16
Hp / Cb, Mb	7/8	>1.8 in 7/8	>50 in 7/8
Hp / Ag, Ec	7/8	>1.8 in 7/8	>50 in 7/8
Ag / Cb, Mb, Hp, EC	14/16	>1.8 in 14/16	>50 in 14/16
Ag / Cb, Mb	7/8	>1.8 in 7/8	>50 in 7/8
Ag / Hp, Ec	7/8	>1.8 in 7/8	>50 in 7/8
Ec / Cb, Mb, Hp, Ag	14/16	>1.8 in 14/16	>50 in 14/16
Ec / Cb, Mb	7/8	>1.8 in 7/8	>50 in 7/8
Ec / Hp, Ag	7/8	>1.8 in 7/8	>50 in 7/8

Table 1. The criteria's used for detecting tissue specific gene expression. Comparison indicates the comparisons done for the specific tissue and for each tissue comparison there are 8 individual samples compared. The algorithm for difference call, fold change and average difference change can be found in appendix 3.

Genes were then classified as 1) "restricted/highly enriched" if all other regions had an absolute call of "absent" 2) "enriched" if detected in all other regions but higher levels in the region in question 3) "decreased" if detected in all other regions but lower in the region in question. and 4) "not detected" if the region scored with an absolute call of "absent" in all four samples and another region scored as "present" in all four samples.

Results

General analysis

Of the 13,069 probe sets analyzed, 7,169 (55%) gave a hybridization signal consistent with a call of "present" in at least one brain region in the 24 samples analyzed. This suggests that at least 55% of the genes covered on the murine arrays (Mu11ksubA and Mu11ksubB) are expressed in one or more areas of the adult male mouse brain. We compared expression profiles of cortex, cerebellum and midbrain within the same strain and found that, on average, only a small number of genes (70/13,069 or 0.54%) showed a clear difference in gene expression between these brain regions. In contrast, 13.6% (1,780/13,069) of the monitored genes were differentially expressed between brain and MEFs, even though the two very different types of cell populations express a similar overall number of genes. This indicates, as might be expected, that various brain regions are more similar to each other than to non-CNS tissue.

To estimate how different the 129SvEv expression profiles are from those of C57BL/6 in a particular brain region, we performed the following analysis: To determine the variation in replicate samples ("false positives") we identified the number of genes that scored as

differentially expressed in comparisons of duplicate strain samples. On average only two genes (2/13,069 or 0.017%) were identified as described in the materials and methods. If the comparison is done between a C57BL/6 region and the corresponding 129SvEv region there is on average a ten-fold increase in the number of genes differentially expressed. This ranges from 23-39 genes depending on the region or 0.17%-0.3% of the 13,069 probe sets (range is 0.45% to 0.73% of the total number of genes present in the particular region). In a similar comparison using MEFs, the strain variation in gene expression was 115/13,069 or 0.88%. Combining the data from the six different brain regions the total number of genes differentially expressed between C57BL/6 and 129SvEv was 73, or 1.08% of the genes expressed in the adult male mouse brain.

Strain specific variation throughout all brain regions

To determine which genes were differentially expressed in multiple brain regions between C57BL/6 and 129SvEv mice, all C57BL/6 brain samples were compared to all 129SvEv samples region by region. The criteria used to assess differences in gene expression are described in the materials and methods. We chose this approach because, although the very strict criteria will lead to some loss in sensitivity, it is most important at this stage to keep the false positives to a minimum. Using this method, we identified 24 genes which were consistently differentially expressed in all six brain regions of C57BL/6 as compared to all six brain regions from 129SvEv (see Figure 4 and Table 2 in appendix 1). In Figure 4, the accession number for the gene is listed in order to simplify the figure. The name, relative expression level and the number of times the gene was detected in a specific brain region are shown in Table 2. As shown, the highly differentially expressed murine leukemia virus gene was only detectable in C57BL/6 mice (Figure 4, AA097626 and Table 2). This is an expected finding as the oligonucleotide probes on the array were derived from a C57BL/6 endogenous retrovirus not expressed in other inbred strains (Kubo et al., 1994). The expected result serves as a positive control for the validity of the approach. In addition, the *Gas5* gene (Figure 4, X59728 and Table 2) is disrupted by a frameshift mutation which decreases the RNA stability in several inbred strains of mice including in the 129 strain but not the C57BL/6 strain (Muller et al., 1998). It is plausible that this accounts for the average overall 1.9 fold difference in expression level detected between the two strains. Interestingly, differences in the abundance of *Gas5* message are correlated with strain specific sensitivity to hyperthermia-induced exencephaly (Vacha et al., 1997).

Brain-region specific differences between mouse strains

To determined which genes were differentially expressed in specific brain regions between the two strains of mice. This allowed us to correlate changes in gene expression in specific regions with behavioral manifestations. For example, if a gene is differentially expressed in the cerebellum but shows no difference in the hippocampus, then this gene is an unlikely candidate to account for the differential strain sensitivity to seizure induced hippocampal neuronal death.

In this analysis we found that a total of 73 genes were differentially expressed in at least one brain region between the two strains. Twenty-four of these 73 genes were already identified and described above. The remaining 49 are shown in Table 2. In general, genes differentially expressed between the strains in one brain region either showed a consistent trend in all other

regions or were not detected in either strain (see Table 3 in appendix 1). Only two genes showed a pattern that was different in different regions. One was glutathione peroxidase which was decreased by approximately nine fold in the midbrain of C57BL/6 compared to the 129SvEv midbrain. By contrast, in the C57BL/6 cerebellum the level of glutathione peroxidase was increased >1.5 fold in comparison to 129SvEv. The second gene is a gene of unknown function ("novel"). The RNA for this gene was decreased by approximately eight fold in entorhinal cortex of C57BL/6 as compared to 129SvEv. In contrast, it was increased by >1.5 fold in the cerebellum of C57BL/6.

Brain region specific gene expression

We also used this data to identify genes that were uniquely expressed or highly enriched in one region as compared to other regions. This information is useful when trying to address the questions of how region specific gene expression may influence brain function. In addition, the identification of genes with unique expression profiles could help identify regulatory elements that could be further exploited to drive gene expression in specific cell types or tissues in animal models. Initially we used the absolute analysis of the data for each brain region. We determined the number of genes with an absolute call of "present" in at least three out of four of the replicate brain region specific samples and classified those as expressed. Similarly we used an absolute call of "absent" in four out of four brain regions and classified those as clearly not expressed or expressed at levels below the threshold of detection. We then used this data to generate Venn Diagrams representing overlapping and non-overlapping gene expression patterns (see Figure 5 in appendix 1).

Several interesting findings emerged. First, the cerebellum appears to be the most unique region of those tested. Twenty-three (0.3%) genes were expressed in the cerebellum but were not detected in other regions. Another 28 were not expressed in cerebellum but were present in other brain regions. Importantly, genes such as *PCP-2*, a known cerebellar specific gene, and *NMDA NR2C*, a known cerebellar specific NMDA receptor subunit, were identified as being specifically expressed in the cerebellum, thereby validating the approach. The midbrain was interesting in that, although there were ten genes uniquely expressed, no genes were exclusively "absent" in the midbrain.

The structures of the medial temporal lobe are known to be similar in their biological importance for learning and memory. By comparing profiles between the three structures (hippocampus, amygdala and entorhinal cortex) we noted that, as might have been predicted, the regions show extremely similar expression profiles. Only eight genes (0.1%) were unique to one region or another (Figure 5B). We then tested how many of these genes were also found in midbrain or cerebellum (Figure 5B numbers in parenthesis). The cortex was not included in this analysis because, as described in the materials and methods, the dissection methods for the cortex included the structures of the medial temporal lobe. We identified only seven genes (0.1%) uniquely expressed in the hippocampus (six of which were also expressed outside of the medial temporal lobe), one in the amygdala and none in the entorhinal cortex in a comparison of the three structures.

We used a second more conservative analysis to compare one brain area to multiple other brain areas (see Table 4 in appendix 1) to detect region specific variation (both restriction and enrichment). Comparison of the gene expression profiles of the cerebellum to all other brain regions showed that 142 genes (2.0%) were differentially expressed (13 restricted/highly

enriched, 64 enriched, 52 decreased 13 not detected). A similar comparison of midbrain showed that only 12 genes (0.2%) were differentially expressed (2 restricted/highly enriched, 9 enriched, 1 decreased and 0 not detected) suggesting that the genes expressed outside of the midbrain were also expressed in the midbrain. Comparison of cortex to midbrain and cerebellum showed that 55 genes (0.8%) were differentially expressed (3 restricted/highly enriched, 33 enriched, 15 decreased and 4 not detected). The gene names and accession numbers for genes that were restricted/highly enriched or not detected are shown in Table 4 for cerebellum, cortex and midbrain comparisons.

A similar strategy was used for the structures of the medial temporal lobe (see Table 5). We found that ten genes (0.1%) were differentially expressed in the hippocampus (0 restricted/highly enriched, 8 enriched, 0 decreased and 2 not detected). The same comparison of amygdala identified only three genes (0.04%) (1 restricted/highly enriched, 0 enriched, 2 decreased and 0 not detected). In entorhinal cortex, ten genes (0.1%) (0 restricted/highly enriched, 5 enriched, 3 decreased and 2 not detected) were differentially expressed. The complete list of genes detected in this analysis is available on the web including information on expression in the cortex for the medial temporal lobe analysis and MEF for all analyses.

Discussion

Biological significance of strain specific variations in gene expression

It has not been established how the constellations of modifying genes within various inbred genetic backgrounds modulate phenotypes. Highly parallel gene expression approaches allow one to look at the global interactions of genes and modifiers and their effects, and will greatly enhance our ability to define the role of developmental alterations, mutations, and compensatory mechanisms in causing or modifying particular behaviors. The expression results can serve as a framework to begin to understand how the large differences in gene expression found in these strains is responsible for the variation in phenotypes including behaviour, drug sensitivity and neurotoxic-induced cell death. The data are also important for understanding how subtle changes in gene expression may give rise to pleiotropic effects. We chose to use these strains because of the ongoing controversy surrounding neurobehavioral studies using non-isogenic mutant mouse strains generated using transgenic technology (Crawley, 1996; Crusio, 1996; Gerlai, 1996; Lathe, 1996).

It is interesting to first comment on the genes described in Table 2 and Figure 4. These 24 genes are differentially expressed in all 12 samples of C57BL/6 as compared to 129SvEv, suggesting that global regulatory mechanisms might account for these changes. In support of such an hypothesis are the findings that the mRNA for the murine leukemia virus gene (derived from the endogenous retrovirus isolated from a C57BL/6 derived cDNA library) was only expressed in C57BL6; and the observation that the *Gas5* gene in the 129 strain of mice is known to have mutations which alter RNA stability (Muller et al., 1998) and, hence, likely accounts for the ~2 fold decrease in expression in 129SvEv as compared to C57BL/6.

What insights can be gained from looking at genes which are differentially expressed between the strains in all brain regions? Virtually all of the known genes that we observed to be differentially expressed have previously defined roles in the central nervous system (CNS). Several genes are worth further comment with respect to studies that identified linked chromosomal regions that contain one or more genes that contribute to strain differences in CNS

phenotypes. These quantitative trait loci (QTL) for phenotypes ranging from seizure susceptibility to abnormal acute ethanol responses, have been mapped to several specific chromosomes (Ferraro et al., 1997; Crabbe et al., 1999; Demarest et al., 1999; Ferraro et al., 1999) in studies using C57BL/6 mice and other inbred strains. QTL analysis is powerful for mapping susceptibility loci to chromosome intervals but many genes reside in these large intervals, and extensive additional work is required to identify which specific gene or genes are implicated. The use of gene expression profiling between these strains may prove useful in identifying candidate genes responsible for the quantitative trait.

In addition, this data may be useful in understanding how modifier genes, whose expression may vary substantially between the strains, might influence a given phenotype. For example, five of the genes we detected as highly differentially expressed have been mapped to specific chromosomal locations. Of these five, two are on chromosome 1. *GIRK3* (Figure 4-U11860) is interesting in that it is located on chromosome 1 in a region that has been shown to contain one or more of the QTLs that contribute to strain differences for free running period and locomotor activity (Mayeda and Hofstetter, 1999), aspects of fear conditioned response (cued and contextual) (Caldarone et al., 1997; Wehner et al., 1997), open field emotionality (Flint et al., 1995), as well as acute pentobarbital induced seizures (Buck et al., 1999). This gene is known to play a role in maintaining resting potential and in controlling excitability of the cell (Kubo et al., 1993) and should be considered a potential candidate for involvement in modulating multiple CNS phenotypes. Another interesting candidate is *PAM* (Figure 4-U79523) that is present in the 129SvEV brain but not detectable in C57BL/6. PAM is a bifunctional key enzyme in the activation of neuropeptides (Ouafik et al., 1992). The gene encodes two different enzymes, peptidylglycine alpha-hydroxylating monooxygenase (PHM) and peptidyl-alpha-hydroxyglycine alpha-amidating lyase (PAL). These enzymes function sequentially in a two step pathway of peptide amidation. This gene maps to 57.5 cM and an ethanol induced loss of righting reflex QTL has been mapped to chromosome 1 between 43-59 cM (Markel et al., 1997). Interestingly, changes in several neural peptides, such as neurotensin, have been linked to ethanol sensitivity, providing a potential link between PAM and modifications of peptides involved in mediating ethanol responses (Duncan and Erwin, 1992).

Another two genes, *I2RF5* and a G-protein subunit, differentially expressed between the strains are located on distal mouse chromosome 4. This region of chromosome 4 has been linked to QTLs for alcohol drinking preference, saccharin and sucrose preference (Bachmanov et al., 1997; Bachmanov et al., 1996; Blizard et al., 1999; Tarantino et al., 1998), and methyl beta-carboline-3-carboxylate seizure susceptibility (Martin et al., 1995). I2RF5 (Figure 4-U31908) is found in post-mitotic neurons but not astrocytes, and functions as a subunit for the shaker type potassium channels. Mutations in this class of voltage-sensitive K+ channels are involved in a number of diseases including seizure disorders. The G-protein specific subunit (Figure 4-U29055) is one of the three G_{β} subunits, $G_{\beta 1}$, of guanine nucleotide-binding G proteins, a large family of proteins that act as signal transducers between transmembrane receptors and cellular effectors, which are widely used in the nervous system.

Several other genes could be considered as strong candidates accounting for the phenotypic variation between strains based on the clear strain differences in gene expresssion. Three of these genes are “novel” or have no known function. Of the genes with known functions, neither *CAP* nor *spi2/eb4*, are detected in C57BL/6 but are detected in 129SvEv. *CAP* (Figure 4-L12367) is an adenylyl cyclase binding protein thought to enhance ras/adenylyl cyclase

interactions in yeast and *spi2/eb4* (Figure 4-M64086) is involved in the cellular response to injury (Inglis et al., 1991). Mounting evidence suggests that increased expression of spi2/eb4 protein is detrimental and is associated with the long term reactive astrocytosis that destroys surrounding brain tissue after ionizing radiation (Chiang et al., 1997). It is interesting to speculate that the enhanced sensitivity to neurotoxic insults in 129SvEv is due to a difference in the regulation of this enzyme.

The *Ste20*-like kinase (Figure 4-AA120636) and a highly similar protein kinase (Figure 4-W51229) were expressed at higher levels in C57BL/6 as compared to 129SvEv. The *Ste20*-like kinase gene is an upstream critical component of the signal transduction cascade that is activated by oxidative induced changes in intracellular calcium in response to insults that generate free radicals, including both hypoglycemia and anoxia. These genes may be ideal candidates for analysis in these strains as the differential sensitivity to cell death may be mediated by changes in the expression (Schauwecker and Steward, 1997).

Finally, we detected several other genes, which were differentially expressed in at least one brain region between the two strains (Table 3). Several of these genes are interesting in light of the known phenotypic variation between these strains of mice. For example, the metabotropic glutamate receptor 1 shows higher expression in the hippocampus of 129SvEv. It is known that administration of antagonists of this receptor diminishes excitoxic and hypoglycemic induced neuronal cell death in the hippocampus (Pellegrini-Giampietro et al., 1999). The decreased expression of this receptor in C57BL/6 hippocampus may account for the decreased susceptibility to cell death (Schauwecker and Steward, 1997). Another such correlation is the difference in expression of *B2m*. B2m is a 12 kDa protein that associates with the Class I products of the H2 major histocompatibility complex such as the H2-K, H2-D, H2-Q, and H2-T antigens. *B2m* is closely linked to *H3* and *H42* on Chromosome 2. Knockout mice for *B2m* are more susceptible to acute encephalitis (Drake and Lukacher, 1998; Lavi et al., 1999). Differences in the expression level of this gene may be important in strain sensitivity to infection, tumor formation and autoimmune diseases (Dawe et al., 1987).

Although these data can only be correlative, it provides potential candidate genes for further study to determine their role in mediating strain specific phenotypes. For example, the task of identifying a gene(s) underlying a QTLs is typically accomplished using standard genetic techniques to narrow the chromosomal region, followed by an attempt to identify the specific gene or genes responsible for the phenotype. However, this latter step is often extremely difficult and time-consuming. With the increase in the number of genes discovered, the main focus is on testing candidate genes rather than discovery of new genes in mapped regions. Given the difficulties in identifying the genes responsible for the phenotype, gene expression profiling may be extremely useful in identifying or establishing the role of a particular set of genes mapping to the region.

Issues surrounding the use of non-isogenic mice for gene expression studies

It is clear that the consequences of a specific mutation often depend on the genotypes at other loci. Only recently has the problem of studying behavior in non-isogenic strains begun to emerge due to the increasing number of mice generated with null mutations with interesting phenotypes. Most of these mutants are not maintained on an inbred background, rather on a mixed background of 129Sv and C57BL/6. Using non-isogenic mouse strains to study behavior or gene expression is likely to produce situations where differences may be identified; but it will

be difficult to conclude with certainty whether the differences are due to the null mutation or the genetic background. We can now estimate how different the 129SvEv gene expression profile is from that of other strains commonly used in transgenic experiments such as C57BL/6. Based on our findings, if the entire expression profiles were monitored using these methods, over ~500-1,000 genes would be differentially expressed between the two inbred mouse strains (estimated based on the range of differences from ~0.56-1.0% of an estimated 100,000 genes in the mouse genome). This study will aid in the interpretation of data arising from gene expression profiling of mice derived from the C57BL/6 and 129SvEv mouse strains.

Use of gene expression profiling for brain molecular mapping

These data can also be used as a guide for determining patterns of gene expression unique to specific brain regions. Studies of the regulatory elements for these genes may be useful in identifying promoters, which could be used to drive expression in specific cell types or tissues in animal models. The paucity of site-specific tools in the mouse makes this an important use of the expression database. The level of consistency between our expression data and published results indicate that array based parallel expression profiling can be a sensitive and accurate method for detecting expression patterns. For example, as shown in Table 4, 13 genes were highly enriched or restricted to the cerebellum. Of those, 11 are known genes. We searched the literature to determine the known expression patterns of these genes. The regional expression patterns we observed were entirely consistent with published findings for ten of the 11 genes. Only *MB-IRK2* was inconsistent in that we were unable to detect mRNA for *IRK2* in any region except the cerebellum, whereas published reports using in situ hybridization were consistent with expression in the cortex and hippocampus with higher levels in the cerebellum (Karschin et al., 1996). Because this was the only gene whose expression pattern was not consistent with published results, we speculate that there may be a specific splice variant represented on the array that is uniquely expressed in cerebellum. Regardless of the explanation for this discrepancy, the >90% concordance with published results suggests that the genes are accurately identified. As it becomes possible to use this technology for nuclei or even small cell populations in the CNS, much higher resolution, region specific and cell type specific information will be gained. This data will also be useful for the identification of genes with restricted expression patterns that can be further studied to define regulatory elements useful for cell specific gene expression. For example, the ARP1 was only detected in the amygdala but not in hippocampus, entorhinal cortex or elsewhere in the brain regions tested (Table 5). Using surgical lesions in rats, it has been suggested that the amygdala is required for memory consolidation in response to epinephrine or glucocorticoids and the role of the hippocampus in this process is not entirely understood (McGaugh, 2000). It may be possible to use the promoter region of this gene to perform molecular lesions in the amygdala without affecting the hippocampus and thereby establish the molecular mechanisms that may elucidate the role of the amygdala in memory consolidation. In addition, the list in Table 4 and 5 was generated using criteria as described in the materials and methods, which excludes genes with low expression levels, but nevertheless may be good candidates for further study. For example, in the analysis used to generate the Venn diagram, *C/EBP-delta* was identified as being uniquely expressed in the hippocampus. It has been demonstrated, that in the brain, the gene is expressed in the hippocampus, is critical for development of long-term facilitation and may be important in the transcription-dependent phase of memory formation (Alberini et al., 1994; Kuo et al., 1990; Yukawa et al., 1998). The

availability of the complete raw and processed data presented in this study will allow investigators in neurobiology and the broader research community to perform further analysis and/or comparison to other analyses performed using this technology (ftp://ftp.gnf.org/pub/papers/brainstrain).

Summary

The ability to simultaneously monitor the expression level of thousands of genes will greatly impact our ability to understand the brain. This study demonstrates the feasibility and accuracy of brain region expression profiling and lays the foundation for asking system-level questions. There is no doubt that the continuing advances in gene targeting technology combined with robust behavioral analysis and gene expression arrays will provide new avenues for studying the brain and further our ability to understand complex “brain systems”.

Acknowledgments

I thank Rie Yasuda for her part on this project (we did it together), she did all the mouse work including housing, dissection and the initial total RNA purification step.
Carrolee Barlow, my supervisor at Salk, for her encouraging enthusiasm, support and friendship and for the parts of the manuscript we wrote together.
Claes Wahlestedt, my supervisor at the Karolinska Institutet, for help on the thesis report.
David J. Lockhart for guidance in gene expression profiling and for sharing a little bit of his genius in interesting discussions.
Lisa Wodicka, and the Lockhart group at GNF (Genomics Institute of the Novartis Foundation), for assistance with gene expression profiling.
Members of the Barlow laboratory for helpful discussions.

References

Abeliovich, A., Paylor, R., Chen, C., Kim, J. J., Wehner, J. M., and Tonegawa, S. (1993). PKC gamma mutant mice exhibit mild deficits in spatial and contextual learning. Cell *75*, 1263-71.

Alberini, C. M., Ghirardi, M., Metz, R., and Kandel, E. R. (1994). C/EBP is an immediate-early gene required for the consolidation of long- term facilitation in Aplysia. Cell *76*, 1099-114.

Bachmanov, A. A., Reed, D. R., Ninomiya, Y., Inoue, M., Tordoff, M. G., Price, R. A., and Beauchamp, G. K. (1997). Sucrose consumption in mice: major influence of two genetic loci affecting peripheral sensory responses. Mamm Genome *8*, 545-8.

Bachmanov, A. A., Tordoff, M. G., and Beauchamp, G. K. (1996). Ethanol consumption and taste preferences in C57BL/6ByJ and 129/J mice. Alcohol Clin Exp Res *20*, 201-6.

Barlow, C., Hirotsune, S., Paylor, R., Liyanage, M., Eckhaus, M., Collins, F., Shiloh, Y., Crawley, J. N., Ried, T., Tagle, D., and Wynshaw-Boris, A. (1996). Atm-deficient mice: a paradigm of ataxia telangiectasia. Cell *86*, 159-71.

Blizard, D. A., Kotlus, B., and Frank, M. E. (1999). Quantitative trait loci associated with short-term intake of sucrose, saccharin and quinine solutions in laboratory mice. Chem Senses *24*, 373-85.

Buck, K., Metten, P., Belknap, J., and Crabbe, J. (1999). Quantitative trait loci affecting risk for pentobarbital withdrawal map near alcohol withdrawal loci on mouse chromosomes 1, 4, and 11. Mamm Genome *10*, 431-7.

Caldarone, B., Saavedra, C., Tartaglia, K., Wehner, J. M., Dudek, B. C., and Flaherty, L. (1997). Quantitative trait loci analysis affecting contextual conditioning in mice [see comments]. Nat Genet *17*, 335-7.

Chiang, C. S., Hong, J. H., Stalder, A., Sun, J. R., Withers, H. R., and McBride, W. H. (1997). Delayed molecular responses to brain irradiation. Int J Radiat Biol *72*, 45-53.

Crabbe, J. C., Phillips, T. J., Buck, K. J., Cunningham, C. L., and Belknap, J. K. (1999). Identifying genes for alcohol and drug sensitivity: recent progress and future directions. Trends Neurosci *22*, 173-9.

Crawley, J. N. (1996). Unusual behavioral phenotypes of inbred mouse strains. Trends Neurosci *19*, 181-2; discussion 188-9.

Crawley, J. N., Belknap, J. K., Collins, A., Crabbe, J. C., Frankel, W., Henderson, N., Hitzemann, R. J., Maxson, S. C., Miner, L. L., Silva, A. J., Wehner, J. M., Wynshaw-Boris, A., and Paylor, R. (1997). Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. Psychopharmacology (Berl) *132*, 107-24.

Crusio, W. E. (1996). Gene-targeting studies: new methods, old problems. Trends Neurosci *19*, 186-7; discussion 188-9.

Dawe, C. J., Freund, R., Mandel, G., Ballmer-Hofer, K., Talmage, D. A., and Benjamin, T. L. (1987). Variations in polyoma virus genotype in relation to tumor induction in mice. Characterization of wild type strains with widely differing tumor profiles. Am J Pathol *127*, 243-61.

Dawson, V. L., Kizushi, V. M., Huang, P. L., Snyder, S. H., and Dawson, T. M. (1996). Resistance to neurotoxicity in cortical cultures from neuronal nitric oxide synthase-deficient mice. J Neurosci *16*, 2479-87.

Demarest, K., McCaughran, J., Jr., Mahjubi, E., Cipp, L., and Hitzemann, R. (1999). Identification of an acute ethanol response quantitative trait locus on mouse chromosome 2. J Neurosci *19*, 549-61.

Drake, D. R., 3rd, and Lukacher, A. E. (1998). Beta 2-microglobulin knockout mice are highly susceptible to polyoma virus tumorigenesis. Virology *252*, 275-84.

Duncan, C. C., and Erwin, V. G. (1992). Neurotensin modulates K(+)-stimulated dopamine release from the caudate- putamen but not the nucleus accumbens of mice with differential sensitivity to ethanol. Alcohol *9*, 23-9.

Ferraro, T. N., Golden, G. T., Smith, G. G., Schork, N. J., St. Jean, P., Ballas, C., Choi, H., and Berrettini, W. H. (1997). Mapping murine loci for seizure response to kainic acid. Mamm Genome *8*, 200-8.

Ferraro, T. N., Golden, G. T., Smith, G. G., St Jean, P., Schork, N. J., Mulholland, N., Ballas, C., Schill, J., Buono, R. J., and Berrettini, W. H. (1999). Mapping loci for pentylenetetrazol-induced seizure susceptibility in mice. J Neurosci *19*, 6733-9.

Flint, J., Corley, R., DeFries, J. C., Fulker, D. W., Gray, J. A., Miller, S., and Collins, A. C. (1995). A simple genetic basis for a complex psychological trait in laboratory mice. Science *269*, 1432-5.

Gerlai, R. (1996). Gene-targeting studies of mammalian behavior: is it the mutation or the background genotype? [see comments] [published erratum appears in Trends Neurosci 1996 Jul;19(7):271]. Trends Neurosci *19*, 177-81.

Hogan, B., Beddington, R., Costantini, F., and Lacy, E. (1994). Manipulating the mouse embryo: a laboratory manual.,, Second Edition Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Inglis, J. D., Lee, M., Davidson, D. R., and Hill, R. E. (1991). Isolation of two cDNAs encoding novel alpha 1-antichymotrypsin-like proteins in a murine chondrocytic cell line. Gene *106*, 213-20.

Karschin, C., Dissmann, E., Stuhmer, W., and Karschin, A. (1996). IRK(1-3) and GIRK(1-4) inwardly rectifying K+ channel mRNAs are differentially expressed in the adult rat brain. J Neurosci *16*, 3559-70.

Kempermann, G., Brandon, E. P., and Gage, F. H. (1998). Environmental stimulation of 129/SvJ mice causes increased cell proliferation and neurogenesis in the adult dentate gyrus. Curr Biol *8*, 939-42.

Kempermann, G., Kuhn, H. G., and Gage, F. H. (1997). More hippocampal neurons in adult mice living in an enriched environment. Nature *386*, 493-5.

Kubo, Y., Baldwin, T., Jan, Y., and Jan, L. (1993). Primary structure and functional expression of a mouse inward rectifier potassium channel. Nature *362*, 127-33.

Kubo, Y., Nakagawa, Y., Kakimi, K., Matsui, H., Higo, K., Wang, L., Kobayashi, H., Hirama, T., and Ishimoto, A. (1994). Molecular cloning and characterization of a murine AIDS virus-related endogenous transcript expressed in C57BL/6 mice. J Gen Virol *75*, 881-8.

Kuo, C. F., Xanthopoulos, K. G., and Darnell, J. E., Jr. (1990). Fetal and adult localization of C/EBP: evidence for combinatorial action of transcription factors in cell-specific gene expression. Development *109*, 473-81.

Lathe, R. (1996). Mice, gene targeting and behaviour: more than just genetic background. Trends Neurosci *19*, 183-6; discussion 188-9.

Lavi, E., Das Sarma, J., and Weiss, S. R. (1999). Cellular reservoirs for coronavirus infection of the brain in beta2- microglobulin knockout mice. Pathobiology *67*, 75-83.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. Nat Genet *21*, 20-4.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays [see comments]. Nat Biotechnol *14*, 1675-80.

Markel, P. D., Bennett, B., Beeson, M., Gordon, L., and Johnson, T. E. (1997). Confirmation of quantitative trait loci for ethanol sensitivity in long- sleep and short-sleep mice. Genome Res *7*, 92-9.

Martin, B., Clement, Y., Venault, P., and Chapouthier, G. (1995). Mouse chromosomes 4 and 13 are involved in beta-carboline-induced seizures. J Hered *86*, 274-9.

Mayeda, A. R., and Hofstetter, J. R. (1999). A QTL for the genetic variance in free-running period and level of locomotor activity between inbred strains of mice. Behav Genet *29*, 171-6.

Mayford, M., Wang, J., Kandel, E. R., and O'Dell, T. J. (1995). CaMKII regulates the frequency-response function of hippocampal synapses for the production of both LTD and LTP. Cell *81*, 891-904.

McClearn, G. E., and Rodgers, D. A. (1959). Differences in alcohol preference among inbred strains of mice. Q J Stud Alcohol *20*, 691-695.

McGaugh, J. L. (2000). Memory-a century of consolidation. Science *287*, 248-251.

Metten, P., Phillips, T. J., Crabbe, J. C., Tarantino, L. M., McClearn, G. E., Plomin, R., Erwin, V. G., and Belknap, J. K. (1998). High genetic susceptibility to ethanol withdrawal predicts low ethanol consumption. Mamm Genome *9*, 983-90.

Morrison, R. S., Wenzel, H. J., Kinoshita, Y., Robbins, C. A., Donehower, L. A., and Schwartzkroin, P. A. (1996). Loss of the p53 tumor suppressor gene protects neurons from kainate- induced cell death. J Neurosci *16*, 1337-45.

Muller, A. J., Chatterjee, S., Teresky, A., and Levine, A. J. (1998). The gas5 gene is disrupted by a frameshift mutation within its longest open reading frame in several inbred mouse strains and maps to murine chromosome 1. Mamm Genome *9*, 773-4.

Ouafik, L. H., Stoffers, D. A., Campbell, T. A., Johnson, R. C., Bloomquist, B. T., Mains, R. E., and Eipper, B. A. (1992). The multifunctional peptidylglycine alpha-amidating monooxygenase gene: exon/intron organization of catalytic, processing, and routing domains. Mol Endocrinol *6*, 1571-84.

Pellegrini-Giampietro, D. E., Peruginelli, F., Meli, E., Cozzi, A., Albani-Torregrossa, S., Pellicciari, R., and Moroni, F. (1999). Protection with metabotropic glutamate 1 receptor antagonists in models of ischemic neuronal death: time-course and mechanisms. Neuropharmacology *38*, 1607-19.

Phillips, T. J. (1997). Behavior genetics of drug sensitization. Crit Rev Neurobiol *11*, 21-33.

Schauwecker, P. E., and Steward, O. (1997). Genetic determinants of susceptibility to excitotoxic cell death: implications for gene targeting approaches. Proc Natl Acad Sci U S A *94*, 4103-8.

Simpson, E. M., Linder, C. C., Sargent, E. E., Davisson, M. T., Mobraaten, L. E., and Sharp, J. J. (1997). Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. Nat Genet *16*, 19-27.

Tang, Y. P., Shimizu, E., Dube, G. R., Rampon, C., Kerchner, G. A., Zhuo, M., Liu, G., and Tsien, J. Z. (1999). Genetic enhancement of learning and memory in mice [see comments]. Nature *401*, 63-9.

Tarantino, L. M., McClearn, G. E., Rodriguez, L. A., and Plomin, R. (1998). Confirmation of quantitative trait loci for alcohol preference in mice. Alcohol Clin Exp Res *22*, 1099-105.

Threadgill, D. W., Yee, D., Matin, A., Nadeau, J. H., and Magnuson, T. (1997). Genealogy of the 129 inbred strains: 129/SvJ is a contaminated inbred strain. Mamm Genome *8*, 390-3.

Vacha, S. J., Bennett, G. D., Mackler, S. A., Koebbe, M. J., and Finnell, R. H. (1997). Identification of a growth arrest specific (gas 5) gene by differential display as a candidate gene for determining susceptibility to hyperthermia-induced exencephaly in mice. Dev Genet *21*, 212-22.

Watanabe, Y., Johnson, R. S., Butler, L. S., Binder, D. K., Spiegelman, B. M., Papaioannou, V. E., and McNamara, J. O. (1996). Null mutation of c-fos impairs structural and functional plasticities in the kindling model of epilepsy. J Neurosci *16*, 3827-36.

Wehner, J. M., Radcliffe, R. A., Rosmann, S. T., Christensen, S. C., Rasmussen, D. L., Fulker, D. W., and Wiles, M. (1997). Quantitative trait locus analysis of contextual fear conditioning in mice [see comments]. Nat Genet *17*, 331-4.

Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997). Genome-wide expression monitoring in Saccharomyces cerevisiae. Nat Biotechnol *15*, 1359-67.

Yukawa, K., Tanaka, T., Tsuji, S., and Akira, S. (1998). Expressions of CCAAT/Enhancer-binding proteins beta and delta and their activities are intensified by cAMP signaling as well as Ca2+/calmodulin kinases activation in hippocampal neurons. J Biol Chem *273*, 31345-51.

Appendix 1: Figures and tables for result and discussion sections.

see also complete gene lists at ftp://ftp.gnf.org/pub/papers/brainstrain

Gene Name	FC	B6	129	Acc. #	MEF
Murine leukemia virus (pol)	~40	12P	0P	AA097626	A
Novel	~9,0	2P	0P	C77761	A
Pituitary tumor transforming gene protein (*PTTG*)	~8.5	12P	2P	AA711028	=
sim. Ste20-like kinase	5.4	8P	0P	W51229	A
Potassium channel beta 2 subunit (*I2RF5*)	~5,0	12P	9P	U31908	A
Novel	3.8	12P	12P	AA409826	*
Ste20-like kinase	3.4	12P	9P	AA120636	*
Novel	~3,0	12P	3P	W35693	*
Dynactin subunit p25	2.3	12P	12P	AA110732	*
Phosphatidylethanolamine binding protein	2.1	12P	12P	AA049118	*
Kinesin heavy chain (*kif5b*)	2.1	12P	12P	AA072168	*
Growth arrest specific protein-5 (*Gas5*)	1.9	12P	9P	X59728	*
Erythroid differentiation regulator	~-17	0P	1P	AA538477	A
Spi2 proteinase inhibitor (*spi2/eb4*)/alpha-1-antichymotrypsin-like protein EB22/4	~-17	0P	10P	M64086	A
Adenylyl cyclase-associated protein (*CAP*)	~-12	0P	12P	L12367	*
Novel	~-10	9P	12P	AA138388	*
Peptidylglycine alpha-amidating monooxygenase (*PAM*)	~-8.4	0P	10P	U79523	A
Novel	~-5.8	0P	9P	AA689927	*
Novel	-4.9	10P	12P	AA114725	*
G protein beta 36 subunit	-4.7	12P	12P	U29055	=
G protein coupled inward rectifier K+ channel 3 (*GIRK3*)	-2.9	2P	12P	U11860	A
Beta-1 globin	-2.6	12P	12P	V00722	A
Beta-globin complex DNA (*y, bh0, bh1, b1, b2* genes & *bh2, bh3* pseudogenes)	-2.3	12P	12P	X14061	A
Novel	-2.3	12P	12P	AA674148	*

Table 2. Strain specific variation consistent in all brain regions

The average fold change (FC) indicates the mean ratio of expression in C57BL/6 relative to 129SvEv in all comparisons (a positive number indicates a higher level of expression in C57BL/6, and a negative number, a higher level in 129SvEv). The tilda (~) indicates that the value is an approximation because the numerator or denominator in one of the comparisons was small relative to the noise. Blue indicates genes with increased expression in C57BL/6 compared to 129SvEv and red, genes with increased expression in 129SvEv. The column labelled B6 represents the number of times a gene scored as present in the analysis of C57BL/6 samples; the number of times a gene scored as present in the absolute analysis of 129SvEv samples is shown in the column labelled 129; The column labelled "MEF" gives a comparison of the expression pattern of the genes when comparing C57BL/6 MEF to 129SvEv MEF. "A" indicates absent, * indicates a similar trend in MEFs compared to that found in the brain, and = indicates no change in expression level between the two samples.

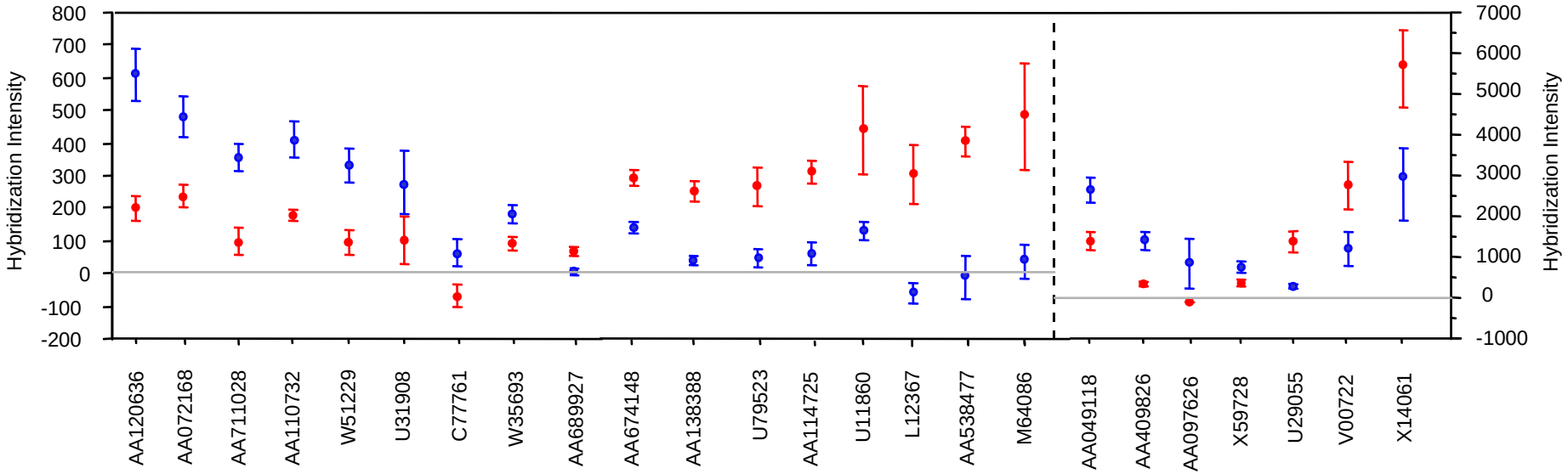

Figure 4. Global gene expression differences between C57BL/6 and 129SvEv mouse strains.
Shown are the hybridization intensity signals of the 24 genes differentially expressed in all brain regions between C57BL/6 and 129SvEv mouse strains. Each gene is represented by a mean value based on the hybridization intensity from the 12 individual samples from each strain (six brain regions done in duplicate) (blue dots ● represent C57BL/6 and red dots ● 129SvEv). The Y-axis is labeled with the hybridization intensities ranging from –200 to 800 (left side of graph) and –1,000 to 7,000 (right side of graph) separated by a hatched black line. The gray dotted line indicates the noise level. Error bars represent the 95% confidence interval derived from the 12 values from different brain regions for each strain.

Gene Name	Cb	Cx	Mb	Hp	Ag	Ec	MEF	Acc.Nr.
Murine leukemia virus (pol)	~49	*	*	*	*	~13	A	AA087673
c-fos oncogene	~11	~6.1	=	=	A	A	A	V00727
Novel	8.4	8.4	*	*	*	*	=	AA154646
Ataxia telangiectasia mutated (*Atm*)	~4.9	*	*	*	*	*	*	U43678
sim. rat mitochondrial enoyl-CoA hydratase (e-83)	4.3	*	=	=	=	=	=	AA270965
Novel	4.2	*	*	*	A	A	=	AA277082
ADP-ribosylation factor 5 (*ARF5*)	3.1	*	*	*	*	*	A	D87902
sim. human beta tubulin 4 (e-35)	2.2	=	=	=	=	=	=	AA030364
Novel	2.2	*	=	*	*	*	*	AA276848
Novel	2.2	=	*	*	*	=	=	AA530176
Novel	2.2	1.9	=	*	*	*	=	R75193
Synaptotagmin 4	2.2	=	=	*	*	*	A	U10355
Novel	2	*	*	*	*	2.4	=	C76063
Novel	1.9	=	=	=	=	=	=	W34733
T-complex testis-specific protein 1 (*Tctex-1*) / Dynein light chain	-1.9	*	*	=	=	*	*	W15873
Novel	-5.2	*	*	*	*	*	A	AA023065
sim. bovine b17.2 subunit of mitochondrial NADH:ubiquitine oxidoreductase complex (e-108)	*	~17	*	*	*	*	=	W90880
Neuronal Protein *NP25*	*	~15	*	*	*	2.2	A	AA118297
Novel	*	~13	*	*	*	*	A	AA288448
Groucho-related gene 1 (*Grg1*)	A	5.3	*	=	=	*	=	U61362
Myocyte specific enhancer factor 2 (*MEF-2C*)	A	2.4	A	~9.8	*	=	A	L13171
Novel	*	1.9	*	2.2	2.3	*		AA035993
T-complex testis-specific protein 1 (*Tctex-1*) / Dynein light chain	*	*	-2.9	=	*	*	*	M25825
Novel	*	*	-2.7	*	=	*	*	R74815
Plasma glutathione peroxidase (*MUSPGPX*)	+	*	~-9.1	=	=	*	*	U13705
Novel	=	=	=	2.9	*	2.0	*	AA289858
Phosphodiesterase I / nucleotide pyrophosphatase (*PDNP2*)	=	=	*	2.4	=	=	A	AA059550

Novel	=	*	*	2.4	*	*	*	AA048853
Kinesin heavy chain (*Kif5b*) / pancreatic beta-cell kinesin heavy chain	*	=	*	1.8	*	*	*	U86090
Glutmate receptor 1 / AMPA1 (alpha1)	=	*	=	-1.9	*	*	A	X57497
Nuclear factor I (NfiA2-protein, splice variant)	*	*	*	-3.0	*	*	A	ET63137
Novel	A	*	*	*	~18	*	A	W47728
Phosphatidylethanolamine binding protein	*	*	*	*	4.1	3.1	*	W35778
sim. rat microvascular endothelial differentiation gene 1 (*MDG1*) (e-109)	=	=	=	=	3.2	=	=	AA673251
Novel	=	=	=	=	2	=	=	AA237797
Novel	=	=	=	*	-6.3	*	A	C81189
Acidic nuclear phosphoprotein *pp32* / Leucine rich acidic nuclear protein (*Lanp*)	=	=	=	=	~-24	*	A	U73478
sim. to phosphatidylinositol 4-phosphate 5-kinase type I-alpha (e-72)	=	=	=	=	-3.1	=	A	AA623242
Beta2-microglobulin (*B2m*)	*	*	*	*	*	~19	=	AA059700
sim. rat peroxisomal membrane protein PMP 70 (e-47)	=	=	*	*	*	2.1	=	AA028386
Novel	*	*	*	*	*	1.8	=	AA198947
Alpha-globin	=	*	=	*	-2.8	-2.7	A	V00714
Alpha-globin	=	*	=	*	-3.1	-2.1	A	C77409
Alpha-globin	=	*	*	*	*	-2.5	A	C79775
Novel	*	*	*	=	=	-2	=	D18279
Novel	*	*	*	=	=	-2.5	*	D18376
SRY-box containing gene 11 (*SOX-11*)	=	*	*	=	*	~-3.1	=	AF009414
Novel	*	*	*	=	=	-6.3	*	AA666918
Novel	=	+	=	*	=	-7.6	*	AA285607

Table 3. Brain-region specific differences between mouse strains
Values represent the fold change in comparisons of C57BL/6 to 129SvEv for Cerebellum (Cb), Cortex (Cx), Midbrain (Mb), Hippocampus (Hp), Amygdala (Ag) and Entorhinal cortex (Ec). MEF indicates a comparison between the C57BL/6 MEF and 129SvEv MEF. Symbols are as follows: "A "indicates absent or below the level of detectability, * indicates similar trend to that found in other brain regions and = indicates no change in expression level between the region in C57BL/6 as compared to 129SvEv. Genes known to be involved in transcription are shown in green, in vesicular transport/synaptic transmission in red and signal transduction in blue. Several "novel" genes and genes with unknown function were also identified (black).

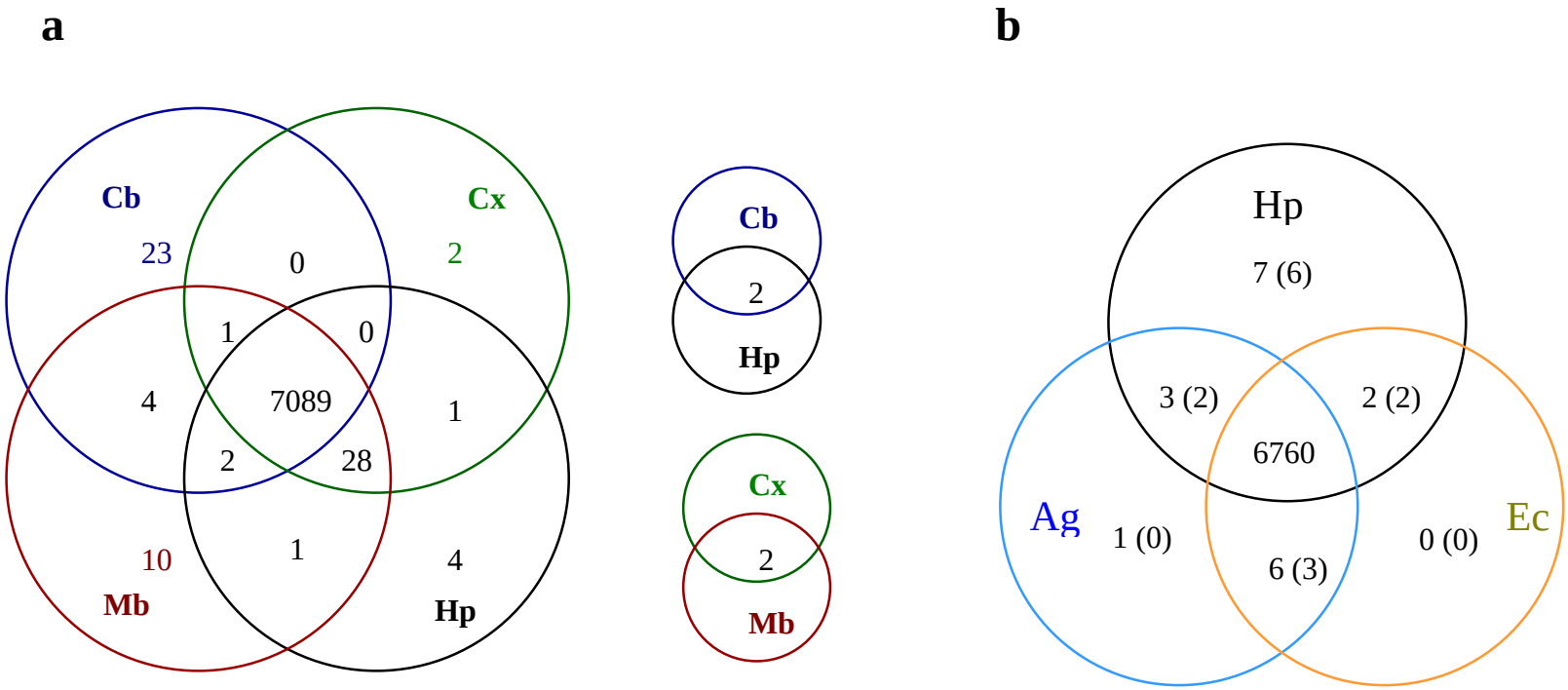

Figure 5. Venn diagrams representing overlapping and non-overlapping gene expression in a subset of adult mouse brain regions in both strains of mice. Region dependent expression patters for cerebellum (Cb), cortex Cx), midbrain (Mb) and hippocampus (Hp) are represented as color-coded circles. The Diagram represents the number of genes with indicated expression patterns. A) Comparison of cerebellum, cortex, midbrain and hippocampus (left side). For clarity, extra circles for areas not captured in the main diagram, because of dimensional restrictions, are shown on the right. B) A separate Venn diagram from an analysis of profiles in hippocampus, amygdala and entorhinal cortex. The number in parenthesis represents the subset of the genes identified, which were also expressed in midbrain and/or cerebellum.

Table 4. Gene list of region-specific gene expression patterns

Cerebellum "Restricted/highly enriched"			
Acc. #	Gene Name	Avg. FC	MEF
AA183544	Novel	3,5	A
AA212550	Novel	8,8	P
L35029	N-methyl-D-aspartate receptor subunit NR2C (NMDA2C) gene	4,1	A
M21532	PCD-5	45	A
M32299	D-amino acid oxidase	~33	A
M90388	protein tyrosine phosphatase (70zpep)	~156	A
M60596	Murine GABA-A receptor delta-subunit gene, exon 9	~22	A
Z38118	Synaptonemal complex protein 1	5,7	A
X80417	MB-IRK2	6,7	A
M90365	Plakoglobin	~36	A
L00919	protein 4.1	~26	A
X61397	Carbonic anhydrase-related polypeptide.	~140	A
D13266	Glutamate receptor channel delta 2 subunit	11	A

Cerebellum "Enriched"			
Acc. #	Gene Name	Avg. FC	MEF
AA034800	Novel	7,9	A
AA123934	Novel	2,7	P
AA270913	Novel	2,5	A
AA274696	Novel	3,2	P
AA289572	Novel	~12	P
AA444931	SNF 1 rel. kinase [rat] e-113 X89383	2,3	A
AA444931	SNF 1 rel. kinase [rat] e-113 X89383	3,9	P
AA472865	Novel	5,3	P
AA473309	ribosomal prot. Kinase S6 (rsk)	2,4	P
AA597258	Novel	3	P
AF004294	Myelin transcription factor 1 .	4,5	A
AF016697	Chemokine receptor gene.	3,9	A
AF035683	p21	4,3	A
D31898	for protein tyrosine phosphatase, PTPBR7	3,9	A
D32167	Zic protein	~100	P
D83262	Neuronal glutamate transporter EAAT4.	12	P
L02241	Protein kinase inhibitor (testicular isoform)	4,8	A
L12147	Early B-cell factor (EBF)	~9.4	A
L12705	Engrailed protein (En-2)	8,2	A
L16846	BTG1	5,5	P
L22144	S100 beta protein exons 1-3	2,8	P
M21531	Calbindin (PCD-29)	4	A
M28489	Ribosomal protein S6 kinase (rsk)	3,5	P
R74641	Novel	16	A
R74735	Novel	3,1	P
U05245	BALB/c invasion inducing protein (Tiam-1)	2,6	P
U19860	Growth arrest specific , clone 3544	3,6	P
U24703	Reelin	7	A
U28068	Neurogenic differentiation factor (neuroD)	~48	A
U33630	Myeloid ecotropic viral integration site-1b (Meis1b) .	5,5	P
U37091	Carbonic anhydrase IV gene	2,8	P
U44725	Mast cell growth factor (Mgf)	3,3	P
U53456	Protein phosphatase 1cgamma (PP1cgamma)	2,4	P
ET61440	Trp-related protein 3 , partial cds.	~12	A

W46015	Histone binding prot NASP	2,7	P
AA059527	p21	2,4	P
AA008502	p21	3,8	A
Y08640	RORalpha 4	4,1	A
X56007	Na/K-ATPase beta 2 subunit gene	3,3	A
X59382	Parvalbumin (small transcript)	19	A
W09791	Zebrin II / p20 cerebella / aldolase C	2,6	A
L16846	BTG1	4,9	P
W41032	hom to homo sapiens MVP gene	5,7	P
W45964	Novel	3,5	P
W77105	Novel	2,9	P
W82359	Novel	3,6	P
X13605	Replacement variant histone H3.3.	3,6	P
X15373-2	Cerebellum for P400 protein.	9,2	P
X51438	Vimentin.	5,2	P
X51986	GABAA receptor alpha-6 subunit.	39	A
X61431	Diazepam-binding inhibitor.	2,7	P
X63963	Pax-6 for paired box protein.	4,1	P
X67141	Pva for parvalbumin.	18	A
X67141	Pva for parvalbumin.	~34	A
X69063	Ank-1 for erythroid ankyrin.	3,1	A
X70398	P311 .	15	P
X73985	Calretinin.	4,8	P
X83202	11beta-hydroxysteroid dehydrogenase/carbonyl reductase.	2,8	A
X98014	Alpha-2,8-sialyltransferase.	~31	A
Y00864	c-kit .	3	A
X70398	P311	14	P
AA105564	SNF-1 rel kinase	6,4	A
W10037	M-cadherin	4,2	A
X61448	D3 clone.	12	P

Cerebellum "Decreased"

Acc. #	Gene Name	Avg. FC	MEF
AA028770	CRP2 Cysteine rich protein [rat] D17512	-4,8	P
AA028770	CRP2 Cysteine rich protein [rat] D17512	~-37	P
AA175767	hom to a focal adhesion related domain (e-12) AF063890	~-34	P
AA221937	lymphocyte antigen 6H (e-102)	-13	A
AA230776	homolgy to thymosin beta 10 (e-24)	-2,4	P
AA245242	MRP MARCKS-rel. prot. (e-71) S65597	-2,7	P
AA285931	Novel	-8,2	A
AA537404	thymosin beta 10 [rat] e-99 M58405	-2,8	P
AA673405	huntington assoc. prot interacting prot (HAPIP) e-79 [human]	-12	A
AA689048	hom to guanine reg prot (ABR) [human] e-22	-2,8	P
AB006191	Cornichon-like protein.	-4,6	A
AB006191	Cornichon-like protein.	-6,3	A
AF026124	Schwannoma-associated protein (SAM9) .	-4,2	P
AF033655	Pftaire-1 .	-4,4	P
D67016	Heat shock protein 105 kDa alpha	-2,3	P
D83206	P24 protein.	-3,7	A
L01695	Calmodulin-dependent phosphodiesteras	-5,2	P
L34214	Glucocorticoid regulated endocrine protein (RESP18)	-4,2	P
M55669	Kex2 homologue	-3,1	A
M59470	cystatin C .	-2,6	P
R74842	N-copine e-138 AB008893	-5,6	A
R75152	Neurochondrin 1&2 e-175 AB019041	-2,6	P
R75531	Novel	-6,4	P

U17259	p19	-6,2	A
U23184	Carboxypeptidase E (Cpe)	-2,3	P
U29088	Nervous system-specific RNA binding protein Mel-N1	-3,7	P
U48797	Astrotactin	-3,9	A
U59418	Protein phosphatase 2A B'alpha3 regulatory subunit , partial cds.	-2,6	P
U86338	Zinc finger protein Png-1 (Png-1) .	-3,5	A
M83749	D-type cyclin (CYL2)	~-12	P
N28171	Novel	-11	A
AA050852	nucleoside diphosphatase kinase A / tumor metastatic process assoc. prot NM23	-2,4	P
W50975	no seq	-4,8	A
X04663	beta tubulin isotype Mbeta 5	-2,1	P
W57404	no seq	-2,1	P
W63974	b-regulatory subunit of protein phosphatase 2A	-3,7	A
AA031158	Neuronal tissue enriched acidic prot NAP-22	-56	P
X51468	preprosomatostatin gene	~-110	A
X59520	CCK gene for cholecystokinin, exon 1	-9,9	A
AA059763	beta tubulin	-2,3	P
W40709	mitochondrial carrier homolg 1 isoform b	-2,3	A
w76777	Novel	-8,4	P
X03151	gene for Thy-1 antigen.	-2,5	P
X07751	c-erbA alpha2 for thyroid hormone receptor.	~-11	A
X51468	preprosomatostatin gene.	~-100	A
X58861	Complement subcomponent C1Q alpha-chain.	~-10	P
X59520	CCK gene for cholecystokinin, exon 1.	-9,5	P
Y00964	Beta-hexosaminidase.	-3,1	P
Z31269	hom to NAP-22 / and hom to a estrogen induced clone	-20	P
AA103457	LIM only 4 prot	~-25	P
L13171	myocyte-specific enhancer factor 2 (MEF-2C) sequence	~-30	A
W63974	Novel	6,8	A

Cerebellum "Absent"

Acc. #	Gene Name	Avg. FC	MEF
AA183623	Novel	~-44	A
AA220788	Novel	~-45	A
AA607353	Novel	~-61	A
L42463	Rho-GDI3	~-15	A
U58887	SH3-containing protein SH3P13	-3	A
U06483	BALB/c telencephalin precursor .	-8,4	A
U28217	protein tyrosine phosphatase STEP61 m	-11	A
U36760	brain factor-1 (Hfhbf1) , class 2	~-46	P
U39738	P21 activated kinase-3 (mPAK-3)	~-24	A
U92565	fractalkine .	~-130	A
U92565	fractalkine .	~-85	A
U56649	cyclic nucleotide phosphodiesterase (PDE1A2)	-9,1	A
AA017811	C kinase substrate calmodulin binding prot (RC3)	~-32	A

Cortex "Restricted/highly enriched"

Acc. #	Gene Name	Avg. FC	MEF
U68058	Frizzled	~12	A
L13171	Myocyte specific enhancer factor 2 (MEF-2C)	~33	A
W64596	Novel	~18	A

Cortex "Enriched"

Acc. #	Gene Name	Avg. FC	MEF

X51468	Preprosomatostain	~50	A
X59520	CCK, cholesystokinin	7,2	P
AA017811	Neurogranin/RC3	~24	A
X51468	Preprosomatostain	~49	A
X53532	protein kinase C beta-II	3,5	A
X59520	CCK, cholesystokinin	7,4	P
Y09257	NOV protein	6	P
AA028770	CRP2 (cysteine rich protein 2)	4,4	P
AA166452	Novel	2,9	A
AA172864	CDCREL-1 homolog	2,2	P
AA175767	sim. Focal adhesion kinase	~19	P
AA183623	Novel	~36	A
AA184871	Novel	2,6	P
AA204034	Novel	2,6	P
AA242333	Novel	2,1	P
AA289338	cAMP regulated phosphoprot. (ARPP-19)	4	P
AA289972	Novel	2,1	A
AA409164	Novel	2,2	P
AA673405	Homo sapiens huntington assoc. protein interacting prot (HAPIP)	12	A
c78582	sim. Zinc finger domain containg prot	~9.8	P
L77867	MEP .	4,3	A
m19436	atrial/fetal myosin alkali light chain (Myla) , clone pCL10.4	5,6	P
M96163	(clone 2) serum inducible kinase (SNK) , sequence	2,8	P
R75030	sim. Homo sapiens BAP-2 alpha prot	4,9	A
U05252	nuclear matrix attachment DNA-binding protein SATB1	2,6	P
U06483	BALB/c telencephalin precursor .	5,5	A
U20372	voltage-dependent calcium channel beta-3 subunit (CCHB3)	2,9	P
u29086	neuronal helix-loop-helix protein NEX-1 (nex-1) , complete cds	7,1	A
U36760	brain factor-1 (Hfhbf1) , class 2	~34	P
U49251	putative cerebral cortex transcriptional regulator T-Brain-1 (Tbr-1)	7	A
U92478	SrcSH3 binding protein , partial cds.	~6.2	P
u92565	fractalkine .	~76	A
u92565	fractalkine .	~50	A

Cortex "Decreased"

Acc. #	Gene Name	Avg. FC	MEF
X04017	Cysteine rich glycoprotein SPARC	-2,8	P
M69042	Protein kinase C delta	-4,8	P
AA008502	p21	-3,2	A
X56007	Na/K-ATPase beta 2 subunit	-2,8	A
X04017	Cysteine rich glycoprotein SPARC	-2,8	P
X73985	Calretinin	-4	P
W17473	Angiotensinogen	~-47	A
AA106347	Angiotensinogen	~-16	A
AA035912	Novel	-3	P
AA409750	Novel	-2,1	P
D32167	zic for Zic protein	~-43	P
m35131	neurofilament component (NF-H) , complet	-2,5	A
M72414	microtubule-associated protein 4 (MAP4)	-3	P
M74570	aldehyde dehydrogenase II	-2,5	A
R74641	Novel	-5,5	A

Cortex "Absebt"

Acc. #	Gene Name	Avg. FC	MEF
D78572	House mouse; Musculus domesticus for membrane glycoprotein	~-12	A
AA002979	Na/K-ATPase beta 3 subunit	-2,6	P

U61751	Versicle associated membrane prot VAMP-1	~-13	A
W13136	Angiotensinogen	-5,4	A

Midbrain "Restricted/highly enriched"

Acc. #	Gene Name	Avg. FC	MEF
AA106347	Angiotensinogen	~19	A
X70393	for inter-alpha-inhibitor H3 chain.	~12	A

Midbrain "Enriched"

Acc. #	Gene Name	Avg. FC	MEF
D16847	Stromal cell derived protein - 1	~10	P
U64572	myelin/oligodendrocyte glycoprotein	~14	A
U81317	myelin-associated/oligodendrocyte basic protein (Mobp)	~66	A
M69042	protein kinase C delta	6,1	A
W13136	Angiotensinogen	5,7	A
W17473	Angiotensinogen	~55	A
X56518	for acetylcholinesterase.	8,9	A
X60304	for protein kinase C-delta.	~18	P
U13705	domesticus C57BL/6J plasma glutathione peroxidase (MUSPGPX)	~19	A

Midbrain "Decreased"

Acc. #	Gene Name	Avg. FC	MEF
L28035	protein kinase C-gamma	~-22	A

Appendix 2: Quality controls for hybridization performance.

Hybridization parameters

% Present	The number of probe sets with an absolute call of ”Present” devided by the total number of probe sets. Indicates the global quality of hybridization. Low % Present could result from low signal, high background or high noise level.
Background	Average of the intensities in the lowest 2% of probe cells.
Stdev	Standard deviation of the background values calculated for different areas of the array.
Qraw (Noise)	Standard deviation of the pixel intensities in the lowest 2% of probe cells. High noise cause low % Present.
Scalar Factor	The factor used to make the average fluorescence intensity across the entire array (after subtraction of background) equal to a target intensity set by the user. Scaling normalizes a number of experiments to one target intensity, allowing comparison between any two experiments. High scalar factor indicates low signal and problem with either hybridization conditions or the labelled sample.
Degradation	Actin and GAPDH probe sets for different regions of the transcript are present on the chip (5’ UTR, middle region and 3’ UTR). By calculating the 3’/5’ ratio possible degradation is detected.

Mu11ksubA arrays - hybridization parameters

Sample	% P	Bkgd	Stdev.	Qraw	SF	Actin	GAPDH
129 Cb1	43%	150	5,6	4,27	1,415	1,09	0,92
B6 Cb1	49%	146	5,8	4,17	1,193	1,07	0,93
129 Cx1	44%	105	3,4	3,33	1,465	1,42	0,91
B6 Cx1	47%	112	4	3,42	1,145	1,24	0,97
129 Mb1	46%	104	4,2	3,23	1,582	0,88	0,95
B6 Mb1	50%	129	10,6	3,57	1,058	1,20	0,97
129 Hp1	50%	102	3,6	3,21	1,04	1,30	0,95
B6 Hp1	46%	96	2,4	3,04	1,643	1,56	1,05
129 MEF	51%	114	3,5	3,43	0,82	1,65	1,16
B6 MEF	48%	113	3,1	3,34	1,13	1,35	1,00
129 Ag1	43%	124	4,3	3,75	1,152	1,42	0,96
B6 Ag1	47%	127	3,7	3,71	1,122	1,36	0,94
129 Ec1	49%	121	3	3,62	1,032	1,21	0,93
B6 Ec1	47%	131	4,3	3,91	0,97	1,21	0,95
129 Cb2	44%	132	4,5	4,06	1,21	1,17	0,91
B6 Cb2	48%	117	3,5	3,77	1,18	1,18	1,00
129 Cx2	44%	112	3,5	3,53	1,45	1,26	0,93
B6 Cx2	45%	119	4,3	3,68	1,24	1,54	0,99
129 Mb2	50%	125	3,7	3,92	0,86	1,21	0,97
B6 Mb2	48%	113	3	3,54	1,18	1,19	0,91
129 Hp2	48%	120	4,2	3,55	1,08	1,34	1,01
B6 Hp2	50%	123	4,2	3,67	0,85	1,42	0,97
129 Ag1:2	49%	98	4,9	3,15	1,28	1,70	1,02
B6 Ag1:2	41%	119	4,9	3,56	1,5	1,10	0,97
129 Ec1:2	46%	107	3,6	3,31	1,53	1,25	0,89
B6 Ec1:2	41%	131	8,5	3,85	1,25	1,20	0,88
129 MEF 2	54%	103	7,3	3,16	0,73	1,80	1,41
B6 MEF 2	48%	126	4,6	3,88	0,98	1,25	0,98

Mu11ksubB arrays - hybridization parameters

Sample	% P	Bkgd	Stdev.	Qraw	SF	Actin	GAPDH
129 Cb1	26%	141	4,6	4,2	1,84	1,03	0,86
B6 Cb1	26%	129	4,6	4,06	1,79	1,06	0,90
129 Cx1	22%	106	4	3,44	2,62	1,39	0,95
B6 Cx1	26%	112	3,6	3,5	1,93	1,21	0,94
129 Mb1	27%	103	3,1	3,28	2,46	1,19	0,96
B6 Mb1	28%	214	12,6	5,29	2,06	1,20	1,00
129 Hp1	27%	110	4	3,43	1,89	1,27	0,96
B6 Hp1	25%	106	4,8	3,24	2,61	1,53	0,95
129 MEF	28%	112	2,7	3,53	1,2	1,81	1,28
B6 MEF	27%	108	9,1	3,32	1,73	1,33	1,04
129 Ag1	27%	107	3,2	3,41	1,64	1,47	0,98
B6 Ag1	24%	113	2,5	3,51	2,16	1,48	0,98
129 Ec1	25%	103	1,9	3,26	2,2	1,12	0,87
B6 Ec1	26%	132	3,4	3,91	1,78	1,18	0,94
129 Cb2	24%	119	4	4,28	2,29	1,19	0,93
B6 Cb2	25%	137	6,2	4,14	1,58	1,03	0,89
129 Cx2	23%	116	5,3	3,59	2,86	1,20	0,95
B6 Cx2	24%	137	3,4	4,21	2,2	1,36	1,03
129 Mb2	26%	130	3,8	3,93	1,6	1,15	0,89
B6 Mb2	28%	123	2,8	3,77	2,04	1,21	0,95
129 Hp2	23%	110	3,9	3,45	2,31	1,35	0,93
B6 Hp2	28%	102	2,9	3,19	1,96	1,37	1,00
129 Ag1:2	24%	113	2,7	3,64	1,75	1,57	0,99
B6 Ag1:2	27%	109	2,8	3,39	1,68	1,23	0,94
129 Ec1:2	21%	123	6	3,97	2,77	1,14	0,87
B6 Ec1:2	27%	104	2,7	3,37	1,54	1,18	0,97
129 MEF 2	30%	102	3	3,36	1,32	1,82	1,41
B6 MEF 2	32%	130	3,2	4,11	1,05	1,22	1,10

Appendix 3: Affymetrix GeneChip algorithms.

Reference: Affymetrix Gene Expression Manual (Dec 1999).

Introduction

This appendix will define the GeneChip algorithms used for establishing the criterias in this study. Firstly, the basic terms and the basic representation on the arrays are described, then the algorithms of the absolute analysis and then comparative analysis. All algorithms have been developed as a result of empirical adjustments based on numerous experiments with known amounts of target transcripts conducted at Affymetrix.

Basic Terms

Probe: a single stranded DNA oligonucleotide complementary to a specific sequence. On Mu11ksubA and Mu11ksubB arrays all probes are 25 bases long.

Probe cell: a single square-shaped feature on any array containing one type of probe (typically 50 or 24μm). Each cell contains millions of probe molecules.

Perfect match: (PM) probes that are complementary to a reference sequence.

Mismatch: (MM) probes that are complementary to the sequence of interest except for a homomeric base mismatch at the central (13^{th}) position. Mismatch serves as a control for cross-hybridization.

Probe pair: two probe cells, a PM and its corresponding MM. On the array the PM cell is located directly above the MM cell.

Probe set: a set of probes designed to detect one transcript. A probe set usually consists of 16-20 probe pairs. For example, a 20 probe pair set is made up of 20 PM and 20 MM for a total of 40 probe cells.

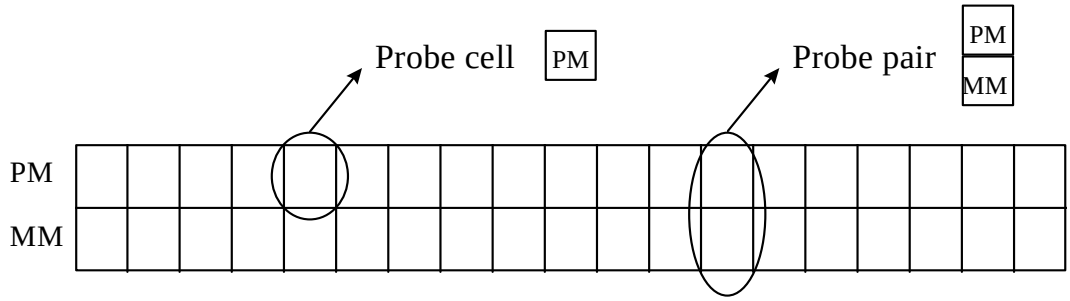

Fig 6. Illustration of a probe set.

Absolute analysis

The goal of the absolute analysis is to determine if a transcript is present in sample or not detectable, based on the observed hybridization intensities. The MM probe sets serves as a sequence dependent background control. The whole process is similar to a courthouse were the judge calls a transcript ”Present” or ”Absent”(or ”not clearly detected”) by the input of twenty different jury members (the probe pairs). If the intensity in the PM cell is higher than the MM cell by more than 2 times the noise (Q) then the probe pair is called a positive probe pair (**Pos PP**). If the MM showes greater than 2 times the noise (Q) higher intensity than PM the probe pairs is called a negative (**Neg PP**). All the probe pairs have one vote either for the presence of transcript (Pos PP) or the absence of detectable transcript (Neg PP). When the difference in intensity is lower than 2 times the noise in any direction, the probe pair will ”pass” because no clear difference was observed.

The **Absolute Call** (AC) could be ”Present”, ”Absent” and”Marginal” (inbetween ”Present” and ”Absent”). Three different factors are used in the absolute call calculation.

1. Positive / Negative Ratio (Pos PP / Neg PP).
2. Positive Fraction (Pos PP / total # of probe pairs (normally 20)).
3. Log Average Ratio (10 * [$\sum$ log (PM / MM)] / (total # of probe pairs).

These factors make up a decision matrix from which the empirically calculated algorithms make the absolute call.

The relative indicator **Average Difference** (Avg Diff) is calculated by taking the difference between the PM and MM of every probe pair and averaging the probe pairs over the entire probe set.

Average Difference = [$\sum$ (PM-MM)] / (total # of probe pairs).

Average diffence correlates with the expression level and is later used for estimating the change in expression level between two experiments.

Comparative analysis

The goal of the comparative analysis is to determine the change in expression level between experiments (and arrays). (Before any comparisons are done the data must be normalized or scaled; The scaling procedure in this study are described in the materials and methods.) You define one of your arrays as the baseline to which you compare the other array, the experiment. The GeneChip software determines a **Difference Call** of ”increase”, ”marignal increase”, ”decrease”, ”marginal decrease” or ”Not Changed”. The software also calculates a fold change, the relative change in transcript abundancy, between two experiments.

Firstly, it determines the number of **Increased Probe pairs** (Inc PP) and **Decreased Probe pairs** (Dec PP). A probe pair is called Increased if:

1. $(PM\text{-}MM)_{exp} - (PM\text{-}MM)_{base} >= 2 * \max(Q_{exp}, Q_{base})$

AND

2. $[(PM\text{-}MM)_{exp} - (PM\text{-}MM)_{base}] / (PM\text{-}MM)_{base} >= PCT / 100$

Percent Change Threshold (PCT): defined by user (default 80).
Likewise, the two criterias reversed must be true in order to call a probe pair ”Decreased”.

The **Difference Call** is determined by 4 factors:

1. Inc PP / total # PP

2. Inc PP / Dec PP

3. Log Avg Ratio Change = $\text{Log Avg}_{exp} - \text{Log Avg}_{base}$

4. Dpos - Dneg Ratio = $(\#\text{Pos PP}_{exp} - \#\text{Pos PP}_{base}) - (\#\text{Neg PP}_{exp} - \#\text{Neg PP}_{base}) / \text{total \# PP}$

In order to determine the change in expression level an **Average Difference Change** (ADC) is calculated:

$$ADC = \text{Avg Diff}_{exp} - \text{Avg Diff}_{base}$$

A relative measure of the change in transcription level is the **Fold Change** (FC) calculation:

$$FC = (ADC) / [\max(\min(\text{Avg Diff}_{exp}, \text{Avg Diff}_{base}), 2.1 * (\max(Q_{exp}, Q_{base}))]$$

$+ 1$ if $\text{Avg Diff}_{exp} >= \text{Avg Diff}_{base}$
$- 1$ if $\text{Avg Diff}_{exp} < \text{Avg Diff}_{base}$

If the maximum value of the noise parameters (Q) is greater than the minimum Avg Diff value the fold change will have a tilde ” ~”, indicating that it only is an approximation because one of the signal was lower than the noise.