APPLICATION OF MICROARRAY, LASER CAPTURE AND TRANSGENIC TECHNOLOGIES TO THE STUDY OF NEURAL DIVERSITY

Thesis by

Mariela Zirlinger

In partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(Defended May 30, 2002)

2002

Mariela Zirlinger

Felizmente, no nos debemos a una sola tradición; podemos aspirar a todas. Mis limitaciones personales y mi curiosidad dejan aquí su testimonio.

Jorge Luis Borges (Buenos Aires, 1899- Geneva, 1986)

ACKNOWLEDGMENTS

First of all, I would like to thank Caltech for accepting me in the biology graduate program. Caltech has provided an enjoyable atmosphere in which to conduct neurobiological research.

Next, I must thank my advisor, David Anderson. He has been an ideal scientific role model because of his enthusiasm, rigor, and vast knowledge. He has also shown me creative ways to think about science.

Kai Zinn, Henry Lester, José Alberola-Ila, and Erin Schuman, members of my thesis committee, also guided me with useful suggestions.

My family in Argentina, mom, dad and Andrés were very supportive. Despite the distance, I felt very close to them during these years.

Gabriel Kreiman, my husband, has been not only supportive in our daily life but also provided a lot of helpful advice, particularly with computational data analysis. He has been a great father to our baby, which allowed me to work round the clock at the end of my doctoral work.

I would also like to acknowledge members of the Anderson lab, past and present, for constructive criticism and for teaching me various molecular biology techniques. Special thanks to Sherry Perez, Jae Kim, Sebastian Gerety and Mark Zylka.

ABSTRACT

A major quest in modern neurobiology is to understand how the brain controls behavior. To this end, the convergence of two traditionally separate fields, systems neuroscience and molecular neuroscience, is required. The delineation of brain regions responsible for different behaviors, and in particular, their underlying neural circuits should be accompanied by the appreciation of the molecules that compose such circuits.

I have taken two approaches toward unraveling the molecular signatures of specific neural structures. First, I conducted microarray-based RNA expression analyses to search, in a large scale and with no *a priori* constraints, for differentially expressed gene products in several brain regions, including the amygdala, cerebellum, hippocampus, olfactory bulb and periaqueductal gray. Interestingly, only 0.3% of the genes characterized to date showed restricted expression in distinct brain areas. Further characterization by *in situ* hybridization was performed for genes enriched in the amygdala, a structure that modulates emotional behavior. Remarkably, this revealed that most region-specific genes possessed expression domains whose limits respected subnuclear boundaries defined by classical cytoarchitectonic criteria. These analyses were not only informative about the molecular composition of distinct brain areas, but also provided tools to genetically dissect the role of different brain nuclei in specific behaviors.

Second, I have used a genetic strategy to label all cellular derivatives of neural crest precursor cells expressing a particular gene, *Ngn2*. Such lineage tracing study uncovered a segregated cellular subpopulation in the developing peripheral nervous system, which was strongly biased for the generation of sensory rather than autonomic neurons. Despite this fate bias, *Ngn2*-derived cells in the dorsal root ganglion were equally likely to give rise to neurons or glia. This suggests that some neural crest cells become restricted to sensory or autonomic sublineages before becoming committed to neuronal or glial fates. In general, visualization of the behavior of neural progenitors during the formation of the nervous system may further our understanding of the generation of specific neuronal subtypes and, eventually, neuronal connections that shape the functioning brain.

The combination of strategies here described will enable the characterization of brain regions at the molecular level on a broad, systems-based approach.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The adult brain contains around 100 billion neurons that make 100,000 billion synaptic contacts (Kandel et al., 2000). Our brain's exquisite pattern of connections and the way different brain regions transduce signals underlie the control of voluntary and involuntary body actions as well as feelings or emotions. Understanding brain function at the molecular level poses one of the most fascinating challenges in modern science.

How the brain controls behavior is being studied from several flanks at once, among others, by molecular biologists and system neuroscientists. The former examine in detail genes and signals that modulate neural function, usually concentrating on simplified model systems amenable to experimental manipulation. System neuroscientists, on the other hand, analyze the intricacies of the nervous system by looking at the entire brain (or large regions of it) as a whole, often treating the brain as a black box disregarding neuronal diversity. Much progress has been made recently in both domains, reviewed in (Albright et al., 2000); we are now entering an exciting era in which both fields can be integrated to provide a more comprehensive view of the brain.

The mammalian brain is subdivided into cytoarchitectonically and physiologically distinct regions. Functional magnetic resonance imaging (fMRI) and lesion studies have suggested that this anatomical parcellation reflects a modular functional organization (Kandel et al., 2000). A major goal then is to elucidate the functional roles of such brain modules, and of the neuronal subtypes that comprise them, in mediating specific behaviors. An important first step in applying the tools of molecular biology to this goal is to identify molecular markers for these subregions. This point is well illustrated by recent work (Abel et al., 1997; Rotenberg et al., 1996; Tsien et al., 1996 (a); Tsien et al.,

1996 (b)) that combined mouse genetics and behavioral paradigms. There, insight about the action of particular molecules in defined neuronal circuits was only possible thanks to the ability to manipulate genes in specific neuronal subpopulations, namely in the hippocampus. To date, very few genes have been reported to be specifically expressed in different brain regions, however, subtractive hybridization experiments have suggested that such brain subregion-restricted genes do exist (Gautvik et al., 1996). Examples of these genes include orexin-A and –B, neuropeptides produced exclusively by a well-defined group of neurons in the lateral hypothalamus (Chemelli et al., 1999), or the alpha 6 subunit of the GABA (A) receptor, expressed only in cerebellar granule cells (Pirker et al., 2000).

In particular, no region-specific gene had been reported in the amygdala, a structure of appeal because it mediates fear responses of psychiatric interest. The amygdala is extensively studied in rodents and primates since it plays a crucial role in the control of sexual, aggressive and fear behaviors, usually referred to as 'emotional' behaviors (Davis, 1992; Gallagher and Chiba, 1996; LeDoux, 1995). Because many human mental disorders, including anxiety, phobia, post-traumatic stress syndrome and panic attack, involve malfunctions in the brain's ability to control fear, studies of the neural basis of this emotion may help us understand and eventually treat these disturbances (LeDoux, 1994).

The amygdala is composed of over a dozen subnuclei with specific functions (Davis, 1992; Pitkänen et al., 1997). A simplified organizational view of the amygdala is that it receives sensory information through its main input structure, the lateral nucleus, and directs behavior through its main output system, the central nucleus. This subnucleus

projects to autonomic centers in the brain stem and hypothalamus, and is responsible for the stereotypical generation of fear responses. In addition, the amygdala receives projections from multiple other brain regions, like the hippocampus (that mainly projects to the basolateral nucleus), frontal cortex (that mostly connects with the basomedial nucleus) and hypothalamus (that heavily connects with the medial nucleus (Pitkänen et al., 1997; Swanson and Petrovich, 1998).) Thus, this forebrain structure is capable of integrating multiple incoming information and of modulating the animal's behavior according to external stimuli and internal state.

To find amygdala-specific transcripts in the mouse relative to other brain regions, I set out to search for genes differentially expressed in this region. At the same time, the screen could determine whether differences in brain areas reflect absolute differences in gene expression or rather combinatorial expression of transcripts. In other words, is the molecular uniqueness of each brain structure determined by the expression of a few region-specific genes, or are most genes expressed throughout the brain, instead, but at different levels or in distinct combinations in each region? Recently developed microarray technology (reviewed in Supplement, 1999) was well suited to answer these questions. Briefly, DNA microarrays exploit the preferential binding of complementary single-stranded nucleic acid sequences. The sample (in this case, each brain region) is hybridized to an ordered array of "probes," or immobilized DNA molecules of known sequence arranged in a defined matrix. Specific hybridization to each probe on the array can be read and translated into abundance levels of each gene in the sample. Unlike conventional nucleic acid hybridization methods, DNA microarrays can quantitatively monitor expression of thousands of genes simultaneously, which has revolutionized

biological experiments lately. However, neural science has initially lagged behind other fields of biology in the exploitation of microarray technology (Serafini, 1999), perhaps due to the enormous complexity of the subject of study.

I have therefore employed microarray technology coupled with *in situ* hybridization to identify and characterize amygdala-enriched gene products compared to genes expressed in four other brain regions, including the cerebellum, hippocampus, olfactory bulb, and periaqueductal gray. This work and others (Lee et al., 1999; Mirnics et al., 2000; Rampon et al., 2000; Sandberg et al., 2000) were among the first ones to apply microarray technology to the study of the brain (reviewed in Cao and Dulac, 2001). In addition, I have later polished the molecular characterization of the amygdala by identifying genes differentially enriched within its various subnuclei. This was accomplished by comparing gene expression profiles from finely dissected pieces of tissue obtained by laser-capture microdissection (LCM), followed by amplification of RNA contents. Briefly, LCM allows the extraction of specific cell populations from a thin tissue section (Emmert-Buck et al., 1996). The strategy consists on the transfer of cells to a laser-activated specially designed transfer film. Thus, with the aid of a microscope, a low-power laser is aimed at the region of interest and specific cells can be collected on the film, while surrounding material is left intact.

The coupling of LCM-tissue extraction, RNA amplification, microarray analysis and *in situ* hybridization inspection of identified enriched genes proved a powerful combination of techniques to characterize brain regions at an unusually high level of detail. Remarkably, we observed that the majority of amygdala-enriched genes had expression domains, whose limits coincided, at least in part, with subregion boundaries

demarcated by classical histological techniques, such as Nissl staining. This was the case regardless of the homogeneity of the starting material, since we obtained subregion-confined genes when the entire amygdala tissue was extracted in a rather crude manner by hand dissection, as well as when specific amygdala subnuclei were carefully dissected by laser capture. By extension, this suggests that classically defined brain nuclei may be also delineated by domains of gene expression. The results obtained here set the grounds for establishing a "molecular brain atlas", in which each structure is also depicted by its molecular repertoire.

In addition, I have taken a seemingly opposite strategy to study the outstanding cellular diversity observed in the nervous system. I have used the expression of a single gene, *Neurogenin-2* (*Ngn2*), to label all cellular derivatives of particular neural progenitor cells marked by the expression of such gene in the peripheral nervous system. Since neural progenitor cells build the nervous system, their characterization may also shed light on brain composition and presumably synaptic connectivity.

Ngn2 is a proneural bHLH transcription factor expressed early in the ontogeny of some neural crest cells, as well as in a variety of CNS neural precursors, e.g. in cortex. It is necessary for the development of some sensory neurons (Ma et al., 1999) but dispensable for the generation of autonomic neurons and glia (Fode et al., 1998). The genetic requirement of *Ngn2* for sensory but not autonomic neurogenesis raised the possibility that intrinsically restricted sensory precursor cells existed in neural crest cells, which give rise to the neurons and glia of the peripheral nervous system. Previous lineage-tracing studies *in vivo* (Fraser and Bronner-Fraser, 1991) have indicated the presence of neural crest precursors with restricted fates. However, it was not possible to

distinguish whether the manifestation of restricted fates reflected true cellular commitment or rather stochastic variations in the behavior of a more homogeneous progenitor population. Later *in vitro* studies (Henion and Weston, 1997) favored the idea that the neural crest is a heterogeneous cellular population containing progenitor cells that generate neurons or glia. Nevertheless, the neuronal subtype(s) generated by restricted progenitors was not examined.

I have used a genetic strategy to label specific *Ngn2*-expressing progenitor cells and their derivatives during nervous system formation in transgenic mice. To do this, a temporally inducible Cre/loxP recombination system (Danielian et al., 1998; Sauer, 1993) was employed. This permits the visualization of the fates adopted by derivatives of progenitor cells expressing *Ngn2* during a well-defined time window. This strategy, although not necessarily equivalent since this is not a clonal cell marking approach, poses many advantages over classical lineage-tracing experiments performed by injection of a vital dye (reviewed in Bronner-Fraser et al., 1991) or a replication-incompetent retrovirus (Frank and Sanes, 1991). First, extensive experimental manipulation is not required as in the case of lineage-dye injection or viral infection. Second, the integrity and survival of labeled cells are not challenged by expression of the reporter transgene. In addition, higher number or labeled cells can be typically obtained. Analyses of transgenic mice designed to observe the fate of *Ngn2*$^{+}$ expressing cells suggested that these cells are strongly biased towards the generation of sensory rather than autonomic derivatives when compared to the behavior of the bulk neural crest cell population. Moreover, they showed that despite this neuronal subtype bias, *Ngn2*$^{+}$-progenitors generated neurons and glia with equal probabilities.

This finding runs counter to classical views of neural cell lineage segregation, since it has been usually proposed that neuronal and glial diversification occurs prior to neuronal subtype specification (Anderson, 2001; Gage, 2000; Temple and Qian, 1996).

Progenitor cells are responsible for generating the various neurons and glia in appropriate locations, ratios and times for the right development of the nervous system. Understanding the logic of their development thus provides another perspective towards uncovering the secrets of nervous system formation and function.

Overview of the Thesis

Described in Chapter 1 is the overall strategy employed to search for genes differentially expressed in the brain. I describe the preparation of the samples, an overview of DNA microarray technology, the custom algorithms developed to analyze the data and the posterior validation by *in situ* hybridization techniques. I also present here the unexpected finding that most amygdala-enriched genes identified respect intra-amygdaloid expression boundaries corresponding to cytoarchitectonically-defined subnuclei. This prompted the further study of gene expression within these specific areas. Therefore, I later refined the search to find genes differentially expressed within amygdala subnuclei. Chapter 2 explains the use of a similar screen to the one described in the first chapter but starting with finely dissected material from distinct amygdala subnuclei. To this end, I applied a relatively novel technique based on Laser Capture Microdissection (LCM) to obtain the samples that were then submitted to microarray hybridization. More amygdala-enriched genes with restricted subnuclear distribution are presented in this chapter, together with a methodological discussion that addresses the benefits of LCM use and the lessons learnt from these screens.

There has been an explosion in the number of papers using DNA microarray technology recently. A large number of manuscripts have also appeared these last years describing new methods for the quantitative analysis of these data. I briefly describe some of these other methods in Appendix 1 and compare the results of some of them (SOM and Bullfrog) to those obtained with my algorithm in the analysis of my datasets.

Diversification in peripheral neural cells is the focus of Chapter 3. Results obtained with genetic lineage tracing experiments indicate that neuron-subtype specification precedes neuron-glial fate determination. A brief description of the genetic strategy employed is also presented in this chapter.

I present some conclusions and observations in the Concluding Remarks section. I discuss the major advancements obtained by the use of these novel techniques and, in particular, their relevance to study the brain. A number of potentially interesting future directions for further investigation has stemmed from this research and I discuss some of them here as well.

1 AMYGDALA-ENRICHED GENES IDENTIFIED BY MICROARRAY TECHNOLOGY ARE RESTRICTED TO SPECIFIC AMYGDALOID SUBNUCLEI

1.1 Introduction

Functional imaging and lesion studies have uncovered a modular functional organization of the brain (Kandel et al., 2000). It is also well established that the brain is subdivided into cytoarchitectonically and physiologically distinct regions. However, little is known about the regional differences in gene expression that may regulate brain function. The speculation of the existence of differentially expressed genes which may confer such specificity embarked me on the search of genes differentially expressed in the adult mouse brain. To maximize this search, I have employed commercially available microarrays that simultaneously monitor expression of thousands of genes and then concentrated on the study of amygdala-specific genes to illustrate the validity of the screen.

Microarray technology represents a potentially powerful new approach to identifying genes specifically expressed in different cell or tissue types (Brown and Botstein, 1999; Lipshutz et al., 1999). The application of microarray technology to the brain, however, poses problems of interpretation not encountered in more homogeneous cell populations, because of its complex anatomical organization and extreme cellular

heterogeneity. This anatomical complexity necessitates that microarray analysis be integrated with systematic *in situ* hybridization studies in order to resolve the cellular distribution of identified transcripts.

Here I report the application of such an integrated analysis to the identification of genes expressed in the amygdala, a brain region implicated in emotional behaviors (Davis, 1992; LeDoux, 1995). *In situ* hybridization has revealed that the majority of genes identified as amygdala-specific on microarrays exhibit intra-amygdaloid expression boundaries corresponding to cytoarchitectonically defined subnuclei. These results support the idea that brain subdivisions detectable by classical neuroanatomical methods reflect underlying differences in gene expression, and demonstrate that systematic identification of molecular markers for such subregions is a feasible near-term goal.

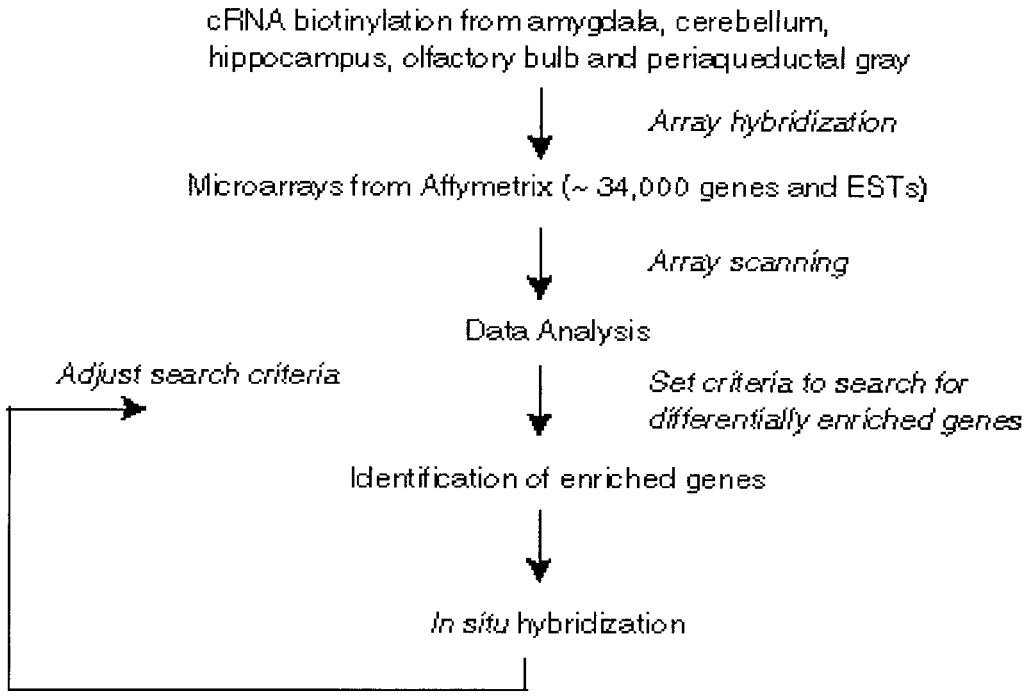
Parts of this chapter have been already reported (Zirlinger et al., 2001).

We thank Raymond Mongeau for dissecting the PAG, Alan Smith and Jingwei Xiao (Stanford/HHMI) for performing microarray hybridizations, Geoffrey Meissner, Cindy Hsu and Sean Pintchovski for help with *in situ* hybridization, Gabriel Miller for help with Figure 2 and Gabriele Mosconi for managerial assistance. We also acknowledge Mark Zylka and Barbara Wold for helpful discussions.

1.2 Methods

1.2.1 Experimental design

A schematic diagram of the strategy used to identify region-enriched genes is shown:



Briefly, five brain regions were chosen for analysis: amygdala, cerebellum, hippocampus, olfactory bulb, and periaqueductal gray (PAG). Biotinylated cRNA probes were prepared for each region and hybridized to Affymetrix microarrays. A custom made program was written to analyze the data which allowed to set different criteria to define what constituted "differentially enriched" genes. *In situ* hybridization was performed on selected enriched genes and based on these results, refinements in the criteria to search for enriched genes were introduced.

1.2.2 Probe preparation

Brains from CD-1, 3 week-old male mice were dissected and kept cold under RNAse-free conditions over the course of the preparation. For isolation of the amydala

and PAG, 34 mice were used. Thick sections (500-600 μm) were sliced with a vibratome and the structures were dissected from these sections under a dissecting scope, following delineations from the mouse brain atlas (Franklin and Watson, 1997). Dissected areas span approximately from -1.06 to -2.18 mm and from -2.92 to -4.24 mm with respect to Bregma, for amygdala and PAG, respectively. Hippocampi, olfactory bulbs, and cerebella were dissected in their entirety from 17 brains. Briefly, total RNA was extracted using Trizol (Gibco), and polyA$^+$ RNA was purified by poly-dT affinity chromatography (Oligotex direct mRNA kit, Qiagen). At least 5 μg of polyA$^+$ RNA was extracted from each brain region. First-strand cDNA was synthesized with a T7-polydT primer (Genset Corp.) using the Superscript Choice System (Gibco). Following second-strand synthesis, approximately 20 μg of biotinylated cRNA hybridization probe was generated from each sample using the MEGA Script T7 kit (Ambion). Because the purpose of the microarray analysis was to identify candidate genes for *in situ* hybridization analysis, rather than to provide accurate measurements of individual transcript abundance, a single set of microarrays was hybridized with each probe.

1.2.3 Overview of Affymetrix microarray technology

Oligonucleotide solid-phase arrays (Fodor et al., 1991; Lockhart et al., 1996) (also known as GeneChips), comprising 34,325 murine genes and ESTs, were purchased from Affymetrix (One set consisted of five subarrays= Mu 11kA, 11kB, 19kA, 19kB and 19kC chips). Each gene or EST is represented on the GeneChips by approximately 20 independent (non-overlapping) "probe cell" sequences, each 25 nucleotides in length. Each perfect match probe cell is located above a control probe cell containing a single-

base mismatch. Biotinylated cRNA from each sample (brain region) is then hybridized to the chips. A fluorescent-tagged streptavidin is allowed to bind to the arrays and the fluorescent signal, which is proportional to RNA abundance, is read with a scanning confocal microscope. A score termed the "average difference" ($\bar{\Delta}$) is assigned to each gene, calculated as the average signal from the 20 perfect-match probe cells minus the corresponding mismatch probe cells. Average difference values are thus indicators of RNA abundance and were used for data analysis (see below). The perfect match and mismatch probe cells pairing strategy, together with the multiple representation of a gene, facilitate data analysis by subtracting nonspecific hybridization and background signals.

The reason for using these arrays was that their reliability and reproducibility in other cell types had been previously shown (Wodicka et al., 1997). However, only known genes or ESTs are represented on the arrays, which precluded the analysis of previously uncharacterized genes.

1.2.4 Data analysis

Prior to analysis, the data were normalized to correct for small differences in the amounts of each cRNA probe applied to the microarrays. Normalization factors were calculated (Affymetrix software) by comparing the mean fluorescent intensity of each array with respect to the corresponding amygdala array. Briefly, the output of each array was multiplied by a factor (the normalization factor) to make its average intensity equivalent to the average intensity of the baseline (amygdala) array. The average intensity of an array is calculated by averaging all the average difference values of every probe set on the array, excluding the highest 2% and lowest 2% of the values. On

average, the mean $\bar{\Delta}$ value for each amygdala array was 1160. Normalized average difference values were exported and analyzed with custom software (available at http://www.its.caltech.edu/~mariela/gene_screen.html), written in Matlab (The MathWorks, Natick, MA), and with GENECLUSTER, which implements self organizing maps (SOMs) (Tamayo et al., 1999).

Two criteria were applied to identify genes enriched in each of the five brain regions: (1) the average difference ($\bar{\Delta}$) value for the gene in that region; and (2) the ratio (fold-difference) of its $\bar{\Delta}$ value in the reference region to that in each of the other four regions. A given gene g_i, with an average difference value in the amygdala of $\bar{\Delta}_{g_i}^{amyg}$, was considered to be enriched in the amygdala relative to the other areas examined if it satisfied the following constraints for these two criteria:

i) $\bar{\Delta}_{g_i}^{amyg} > minimum$

ii) $\bar{\Delta}_{g_i}^{amyg} / \bar{\Delta}_{g_i}^{other} > threshold$ or $\bar{\Delta}_{g_i}^{amyg} / \bar{\Delta}_{g_i}^{other} < 0$ for all four other regions (cerebellum, hippocampus, olfactory bulb, and PAG).

To identify genes that were enriched in one region with respect to any three of the other regions, constraint ii) had to be met for any three (but not four) of the remaining four samples. The *minimum* $\bar{\Delta}$ values and *threshold* ratios were iteratively adjusted and optimized based on the results of preliminary *in situ* hybridization experiments (see below).

1.2.5 *In situ* hybridization

Three to four week-old male and female CD-1 mice were used. Clones were purchased from Research Genetics when available, or templates for probes were synthesized by PCR using specific primers and cDNA from mouse brain. For some genes, sense probes were also synthesized to control for non-specific hybridization. Digoxigenin-labeled RNA probes were made and hybridization was performed essentially as previously described (Henrique et al., 1995), with some modifications. Briefly, fresh frozen, 20 μm thick coronal sections were cut with a cryostat. Sections were dried and fixed in 4% paraformaldehyde, washed in PBS and subjected to acetylation using 0.25% acetic anhydride in 1M Triethanolamine-HCl pH 8.0. Slides were prehybridized for 1-3 hr, and hybridized overnight at 70°C, using a probe concentration of 0.5 - 1 μg/ml. Sections were washed twice in 0.2X SSC at 70°C for 30 min, incubated with anti-digoxigenin alkaline phosphatase-conjugated Fab fragments (Roche) at a 1:2000 dilution in 0.1 M maleic acid buffer, pH 7.5, with 0.2% Tween-20, 20% sheep serum and 2% blocking reagent (Roche). Staining was developed for 4-16 hr with NBT and BCIP (Roche) in alkaline phosphatase buffer to yield a purple product. Slides were fixed in 4% formaldehyde and mounted with glycerol. Adjacent sections were Nissl-stained for comparison. Images were collected with a Zeiss Axioskop, or an Olympus IMT2 microscope attached to a CCD camera and Neurolucida software (Microbrightfield, Colchester, VT), using 35 mm film or electronically acquired composite images, respectively. To aid in the visualization of brain regions, Nissl staining was done on adjacent sections.

1.3 Results

To evaluate the distribution of mRNA species within a given brain region, we plotted a histogram of the average difference values for all genes from each region (Fig. 1A). Only the plot for one subset of chips (11kA) is shown here, but all other chips had similar distributions. The histogram describes a unimodal distribution, wherein the bulk of the mRNA species is of relatively low abundance. Only a minority of the mRNA species were highly expressed. These results are in agreement with previous estimates of mRNA abundance and complexity in other tissues (Lewin, 2000; Zhang et al., 1997).

For an initial comparison of gene expression levels between the different brain regions, I first plotted all possible pairwise comparisons of the average difference values for all genes represented on the 11kA array (Fig. 1B). Although the hippocampus appeared to have more off-diagonal points of greater abundance relative to the other four regions, the bulk of the transcripts lay close to a line of slope=1. The lack of obvious clusters of genes underscored the need for algorithms to systematically identify differentially expressed genes.

To facilitate the parallel analysis of relative gene expression levels between all five brain regions, we developed a custom algorithm (see Data Analysis section). I then systematically varied different parameters in this algorithm to maximize the search for region-enriched genes. For example, I searched for genes whose average difference values were at least 3.5, 5, or 6 times higher in any given reference region compared to the remaining four. Based on *in situ* hybridization experiments with genes identified in early iterations of this search, it was concluded that a *threshold* ratio of 3.5 was optimal. Using higher threshold ratios (e.g., 5- or 6-fold) failed to identify many genes found by

the lower-stringency search, whose differential expression could be validated by *in situ* hybridization (for example, probe 4, Fig. 2D). Conversely, lowering the threshold below 3.5 was more likely to identify genes whose expression would not be region-specific. To filter out genes satisfying the ratio criteria, but whose absolute expression levels were likely to be too low to be detectable by our *in situ* hybridization procedure, we empirically arrived at a minimum average difference value. For genes enriched in the amygdala sample, this lower limit corresponded to one-tenth of the mean average difference value for all genes on a given array. On the 11kA array, for example, this minimum $\bar{\Delta}$ value was 110.4, a figure approximately five-fold above the noise level for this array (22.5).

To compare the performance of our search algorithm to that of an independent method, we carried out a clustering analysis using the Genecluster program (Tamayo et al., 1999), that implements self-organizing maps (SOMs). All genes identified by our method were also included in SOM-derived clusters corresponding to single region-enriched genes. However, the total number of genes in each of these SOM clusters was about 6 to 10 times larger than the number of genes identified using our algorithm, even with the use of stringent filters for SOMs. Logically, these additional genes were rejected by our algorithm because they either fell below $\bar{\Delta}_{min}$, or because their –fold difference relative to the other four regions was too low. Most of these additional genes, however, were not considered good candidates for *in situ* hybridization because they either were not expressed well above background levels, or had a low fold-difference relative to the other regions.

1.3.1 Analytical characterization of differentially expressed genes

We found that only 455 of the 34,325 genes and ESTs analyzed (1.3%) fulfilled our selection criteria for enrichment in any one of the five brain regions, relative to the other four (Table 1). Of these, 33 genes were enriched in the amygdala. On average, 0.3% of the sampled genes were highly enriched in any one of the five brain regions (Table 1). We also computed the number of genes that were "present" (i.e. had significant expression above background levels) in all five regions, as well as those that had no detectable expression. 9,604 genes (28% of the genes on the array) were expressed in all areas examined, whereas 15,303 (45%) were present in none. Thus, of the 19,022 genes with detectable expression in one or more regions, half were present in all regions. Among the present genes, only 2.4 % were differentially expressed in one region (455/19,022). A complete table with all 455 genes or ESTs and their corresponding $\bar{\Delta}$ values is shown in Table 3, which can also be found in http://www.its.caltech.edu/~mariela/gene_screen.html.

To investigate whether certain classes of genes were preferentially represented among these sequences, we classified all annotated differentially expressed genes based on their structure or function. Of the 455 sequences, 117 (26%) were annotated genes. In four cases, annotation was accomplished by using 5' RACE to clone the coding region. The genes were classified among 21 different functional categories, following the Gene Ontology (GO) Consortium guidelines (Ashburner et al., 2000). The categories that were the most highly represented (contained >7% of the 117 genes) comprised signaling molecules (26%, n=30), DNA binding molecules (17%, n=20), enzymes (15 %, n=18), and structural proteins (9%, n=10). Some examples of these are shown in Table 2. See

Table 4 for a full list of the functional categories and the percentage of genes in each class for each of the five brain regions analyzed.

Several of these genes had previously been reported to be enriched in their given areas. These included vasopressin (Caffe and Vanleeuwen, 1983; Price et al., 1987) and arp-1 (Dasilva et al., 1995) in the amygdala, P400 (Nakanishi et al., 1991) and NeuroD (Lee et al., 1995; Schwab et al., 1998) in the cerebellum, and tyrosine hydroxylase in the olfactory bulb (Hokfelt et al., 1984). Of 13 genes identified as cerebellum-enriched in a recent Genechip study (Sandberg et al., 2000), we independently identified 6. Of the remaining 7 genes, 2 were rejected by our algorithm because they had $\bar{\Delta}$ values < 0, and 5 were rejected because they were also expressed at substantial levels in the olfactory bulb, a structure not analyzed in the earlier study. However, as that study included other brain regions, such as the cortex, not examined in our experiments, the two data sets are complementary.

1.3.2 Validation of GeneChip results by in situ hybridization

It was essential to validate the results of the microarray analysis by an independent method. We employed *in situ* hybridization rather than biochemical assays such as RNase protection, because the complex anatomical organization of the brain necessitates a method with high spatial resolution. Thirty-five genes were analyzed by *in situ* hybridization. Of these, about 60% were expressed in a manner consistent with the results of the microarray analysis. Twenty percent did not show any signal, 13% hybridized everywhere and 7% were inconsistent with the microarray results (i.e., hybridized more strongly to regions that were predicted to have lower abundance). Since

it was impractical to optimize probe design and hybridization parameters for each gene, it is possible that the actual false negative and false positive rate is lower than we observed.

To determine the extent to which our algorithm conditions could be further relaxed, we performed *in situ* hybridization experiments for four "best candidate" genes identified by Genecluster that marginally failed to meet our selection criteria. Three of these did not show any signal, but one was indeed expressed in the amygdala (probe 41, Fig. 2C). However, this gene was identified by Genecluster only with the use of a very lax filter that included many other genes that fell well below our selection criteria.

Strikingly, although our selection criteria required only a 3.5-fold difference in the level of expression in one region compared to the others, in many cases this seemingly modest quantitative difference on the arrays translated into an apparent qualitative difference when examined by *in situ* hybridization. Thus, the expression of many amygdala-enriched genes was simply not detected by *in situ* hybridization in the other regions examined in the Microarray analysis. This may reflect the fact that many of the genes had fairly low average difference values in the amygdala, so that a 3.5–fold lower level of expression in one of the other regions might be below the detection limit of the non-isotopic *in situ* method. As might be expected, most of the amygdala-enriched genes proved to be expressed in at least one other brain area not tested in the Microarray experiment, such as the cortex (Fig. 3C).

The absolute $\overline{\Delta}$ values obtained from the Microarrays do not distinguish whether a given gene is expressed at high levels in a small subpopulation of cells, or at lower levels in a larger population. Among the genes that we examined, one-fourth (25%) showed strong expression in relatively small, scattered cell populations, while the

majority (75%) were expressed more broadly. Because the pieces of tissue we dissected for RNA isolation were relatively large and heterogeneous, it is likely that our analysis was biased against genes expressed at lower levels in small subpopulations of cells.

1.3.3 Amygdala-enriched genes respect subnuclear boundaries

The amygdala is a complex structure that can be anatomically subdivided into at least 13 distinct regions (Pitkänen et al., 1997), such as the lateral, basolateral, medial, and central nuclei (Fig. 2A). This structural organization raises two questions: 1) Do the boundaries of amygdaloid nuclei reflect boundaries of gene expression domains?; 2) Do gene expression patterns reveal features of amygdaloid organization not visible by classical neuroanatomical techniques? To address these questions, we examined in detail the *in situ* hybridization pattern of 12 genes predicted by the microarray analysis to be enriched in the amygdala.

Surprisingly, the majority (75%) of these genes exhibited restricted, contiguous domains of expression, whose boundaries at least partly coincided with those of amygdaloid subnuclei (Fig. 2A). (The remaining genes were expressed in scattered populations of cells). Within this larger group of genes, about 50% completely respected nuclear boundaries (Fig. 2B, C; Fig. 3A, C, E). The other half respected nuclear boundaries along part of their length, but in places extended beyond these boundaries into a well-defined territory not coincident with any described amygdaloid subdivisions (e.g., Fig. 3D, dotted vs. dashed lines).

All of the amygdala-enriched genes we examined could be parsed into three groups, according to the distinct ontogenetic origins of the subnuclei in which they are

expressed. One group of genes (5 genes; 42%) was expressed in the lateral, basolateral, and cortical nuclei (Fig. 2A, blue), which are cortical-like structures embryologically derived from the pallium. Probes 29 (activin receptor type II; Fig. 2B) and 75 (unconventional type myosin; Fig. 3A) are characteristic of this group. The second group (5 genes; 42%) was formed by genes expressed in the central and medial nuclei (Fig. 2A, yellow), which have sub-pallial (striatal or pallidal) origin. Probe 41 (laminin ß3, Fig. 2C) is characteristic of this group. The third group (16%) consisted of two genes (the transcription factor arp-1 (Fig. 2D) and Ccte chaperonin epsilon subunit (Fig. 3F) with widespread expression throughout the amygdala, including pallial and subpallial nuclei. Thus, the majority (84%) of the genes were expressed in either of two subsets of amygdaloid subnuclei related by a common developmental origin.

Genes in the first and second groups also shared some other features of their expression. For example, several of the genes in the first group (e.g., probes 75 and 50, Figs. 3A and 3C) were also expressed to varying extents in the neocortex, consistent with the pallial origin of the amygdaloid regions in which this group is expressed. Conversely, a number of the genes in the second group (e.g., probe 28 and probe 41, Figs. 3D and 2C) also labeled the hypothalamus, a pattern also noticed by Sutcliffe and colleagues when characterizing expression patterns of hypothalamic genes (Gautvik et al., 1996). Interestingly, all genes in the first group were expressed in contiguous cell populations. This observation may reflect the fact that the lateral and basolateral complexes are relatively homogeneous with respect to both cell type and neurotransmitter content (McDonald, 1992; Price et al., 1987). By contrast, 80% of the genes in the second (striatal) group, such as the neuropeptide vasopressin, were expressed in scattered

subpopulations of cells. This observation is consistent with the fact that amygdaloid neuropeptides are generally expressed in scattered cell populations (Roberts et al., 1982), and also that the centromedial aspect of the amygdala is the most neuropeptide-rich region in the brain outside the hypothalamus (Roberts, 1992). Other genes in this subgroup included the Lim homeodomain transcription factors Lhx-6 (Fig. 3B) and Lhx–7. It is possible that these factors are involved in the regulation of amygdaloid neuropeptide gene expression.

1.4 Discussion

The modular functional organization of the mammalian brain is likely to reflect, at least in part, its anatomical parcellation into distinct substructures. We have used microarray analysis in conjunction with *in situ* hybridization to identify molecular markers of this anatomical regionalization. Using commercially available microarrays, we identified in each of five selected brain regions, on average, 91 genes that were highly enriched. This estimate is very close to that arrived at in a previous study employing subtractive hybridization (Gautvik et al., 1996), which estimated the number of transcripts highly enriched in the hypothalamus to be on the order of one hundred. Our figure constitutes 0.3% of the approximately 34,000 genes interrogated, and 0.5% of all genes expressed in at least one of the five areas (91/19,022). Similar values were recently reported by Sandberg et al., who analyzed the expression of about 13,000 genes and ESTs in a different subset of brain regions than we examined (Sandberg et al., 2000). These values may, however, be an underestimate because genes expressed at low levels in small subsets of cells may have been systematically excluded by both analyses.

Nevertheless, the gross molecular homogeneity of the brain was apparent. It is of course also likely that many other unknown region-specific genes exist, which were not interrogated by the Affymetrix GeneChips. Other microarray methods that do not rely on previous knowledge of sequences may prove useful in identifying these.

Among the identified differentially expressed genes with known function, 67% fell into 4 of 21 functional categories, comprising signaling molecules, transcription factors, enzymes or structural proteins. However, the majority (72%) of the differentially expressed genes were unannotated ESTs, making it difficult to draw firm conclusions about categorical representation.

1.4.1 Analytic considerations

For simply identifying region-specific or highly enriched genes, our custom algorithm proved more efficient than SOMs cluster analysis (Tamayo et al., 1999). That is because our program permits the explicit specification of multiple criteria for "enriched" genes. In contrast, Genecluster identifies collections of genes that share similar features. Therefore, no constraint about the ratio of expression in one brain region relative to all the others can be independently set. However, SOM analysis is designed for gene-profiling studies, where the comparison of expression patterns among a large collection of samples is sought (Tamayo et al., 1999). This is fundamentally different from positively selecting highly enriched genes that fulfill specific ratio criteria.

1.4.2 Validation of microarray data

A recent study (Sandberg et al., 2000) also employed Affymetrix GeneChips to characterize region-specific gene expression in the brain, but did not validate the microarray results by *in situ* hybridization. Our results suggest that *in situ* hybridization is essential to confirm GeneChip data. Of the 35 genes we tested, 80% yielded detectable *in situ* hybridization signals. Of these, about 25% exhibited patterns apparently inconsistent with the Microarray data. Thus, 60% of the 35 genes examined were validated by *in situ* hybridization. Of the 14 cases of inconsistency, most (65%) reflected probes that hybridized everywhere. These cases may simply represent suboptimal probe design rather than any inherent inaccuracy of the GeneChip method. The remaining cases, however, constituted probes that gave strong *in situ* signals in regions predicted to be weak or negative by the microarrays. It is possible that replicate microarray experiments with independently prepared samples and chips would have lowered the number of false-positives. However, considering that at least 17 mice were used to prepare cRNA probes from each brain region, it is unlikely that the discrepancies we observed are due to inconsistent tissue dissection, or to biological differences between the animals used to prepare microarray probes and those used for *in situ* hybridization.

Even for those genes whose *in situ* pattern was consistent with the predictions of the microarray data, *in situ* hybridization was also essential to identify sites of expression not included among the original five samples. This is important, as it is technically impossible to analyze every brain region or nucleus in a given microarray experiment. Our *in situ* analysis also revealed how extrapolating mRNA abundance levels based on $\bar{\Delta}$ values from the Microarrays is not necessarily informative, since this value reflects

both the abundance of a given transcript within expressing cells, as well as the proportion of cells expressing the transcript in a given brain region. We have found examples of genes with low $\bar{\Delta}$ values that were expressed at very high levels in a few cells, and conversely, genes expressed broadly at more modest levels, that yielded high $\bar{\Delta}$ values.

1.4.3 Towards a molecular anatomy of the amygdala

The amygdala, a brain region implicated in emotional learning (Davis, 1992; LeDoux, 1995), lies at the interface between the cortex and sub-cortical structures such as the striatum and hypothalamus, and therefore is well-positioned to integrate computational and neuromodulatory functions. Accordingly, the amygdala is structurally and ontogenetically heterogeneous, consisting of over a dozen subnuclei (Pitkänen et al., 1997; Price et al., 1987; Puelles et al., 1999). The central and medial nuclei are subpallial (striatal or pallidal) derivatives, whereas the cortical-like subnuclei (nucleus of the lateral olfactory tract, lateral, basolateral, cortical, and claustral amygdala) are pallial (cortical) derivatives. This division is based on the expression of homeobox transcription factors whose expression domains delineate pallio-subpallial boundaries in the embryonic telencephalon (Puelles et al., 1999). Such a general division fits well with the proposal of Swanson and Petrovich (Swanson and Petrovich, 1998), who have divided the amygdala into cortical- and striatal-like structures based on anatomical features and neurotransmitter expression. The majority of the amygdala-enriched genes we examined obeyed this basic dichotomous subdivision.

We made no special effort to microdissect distinct subnuclei in preparing the microarray hybridization probe; rather, a relatively crude dissection of the entire

amygdala was used. Thus, it is striking that 75% of the amygdala-enriched genes that we examined by *in situ* hybridization (n=12) exhibited expression boundaries at least partly coinciding with those of one or more subnuclei (Fig. 4v). A priori, this need not have been the case. At least four other kinds of expression patterns could have been obtained (Fig. 4): (i) uniform expression throughout the amygdala; (ii) contiguous subdomains bearing no relationship to classically defined subnuclei; (iii) scattered expression in cells contained within specific subnuclei; and (iv) scattered expression not respecting subnuclear boundaries. It is striking that no genes fell into either of the first two categories. These data suggest not only that the boundaries of amygdaloid subnuclei reflect gene expression boundaries, but moreover that the majority of amygdala-enriched genes may respect such boundaries. The genes we have identified should, therefore, provide useful markers for amygdaloid subnuclei, some of which can be difficult to visualize by Nissl staining on thin histological sections.

Our data also indicate, however, that not all gene expression boundaries correspond precisely to boundaries of amygdaloid subnuclei. For example, we observed three genes that had a similar, well-defined expression domain that included the medial amygdala, but which extended into a limited subregion of the adjacent basomedial amygdala (Fig. 3D, dotted line). This suggests that the medial and basomedial amygdaloid cells expressing these genes share previously unrecognized molecular similarities, and may form a novel subdivision not detected by classical anatomical methods. Thus, gene expression domains do not simply validate classically defined anatomical units, but may also reveal organizational features not easily visualized by existing staining techniques.

At present, the rate-limiting step in the analysis of microarray data derived from the brain is its validation by *in situ* hybridization. When efficient, large-scale, high-throughput automated *in situ* hybridization procedures for adult brain sections become available, it should be possible to exploit Microarray data to generate a "molecular brain atlas," in which each structure is also delineated by its molecular repertoire. The results presented here demonstrate that such a long-term goal is, in principle, feasible. The genes identified by such an exercise, moreover, are not simply markers, but also will provide tools to genetically dissect the roles of such brain substructures in specific behaviors.

Table 1-1: Genes that are at least 3.5- or 5-fold enriched in each of the five areas.

	Amy	Cb	Hpc	OB	PAG	Total	Avg
3.5-fold relative to all 4 other areas	**33 (0.1)**	**159 (0.5)**	**89 (0.3)**	**101 (0.3)**	**73 (0.2)**	**455 (1.3)**	**91 (0.3)**
5-fold relative to all 4 other areas	**21 (0.1)**	**86 (0.3)**	**57 (0.2)**	**68 (0.2)**	**44 (0.1)**	**276 (0.8)**	**55 (0.16)**
5-fold relative to any 3 other areas	**65 (0.2)**	**164 (0.5)**	**105 (0.3)**	**127 (0.4)**	**95 (0.3)**	**556 (1.6)**	**111 (0.3)**

Genes that are at least 3.5 or 5-fold enriched in each of the five areas are shown. Percent of total genes interrogated (34,325) are in parentheses. Enriched genes with respect to all 4 other regions or to any other 3 regions are indicated. (Amy= amygdala, Cb= cerebellum, Hpc= hippocampus, OB= olfactory bulb and PAG= periaqueductal gray).

Table 1-2: Examples of genes enriched at least 3.5-fold in each region.

Functional category	Amy	Cb	Hpc	OB	PAG
Signaling	vasopressin (M88354)	cerebellum P400 protein (X15373)	ND	B219/ OB receptor (ET61693)	angiotensinogen precursor (Msa.7127.0)
DNA-binding	arp-1 (X76653)	Neuro-D (U28068)	Friend of GATA-1 (FOG) (AF006492)	Dlx-1 (U51000)	Gata-2 (AB000096)
structural	unconventional type myosin (TC37197) (*)	pro-alpha-2 (I) collagen (Msa.2220.0)	dynactin (Msa.12975.0)	pro-collagen type V alpha-2 (Msa.544.0)	ND
enzyme or ligand binding	ND	parvalbumin (X67141)	neuropsin (D30785)	tyrosine hydroxylase (M69200)	angiotensin-converting enzyme (Msa.24687.0)
EST	TC35462 [activin receptor type II (*)]	TC33451	TC36417	TC20543	TC36249

Some examples of genes enriched at least 3.5-fold in each region. Gene names and (Affymetrix probe set names) are presented. (*): Gene identity was determined by 5' RACE. ND= not detected among the 117 genes that were annotated. Abbreviations are as in Table 1.

Table 1-3: List of enriched genes.

Affymetrix probe sets[1] enriched at least 3.5-fold in each brain region relative to all others. Average difference values (avg diff) in the 5 brain regions are presented (Amy= amygdala; Cb= cerebellum; Hpc= hippocampus; OB= olfactory bulb; PAG= periaqueductal gray.) Bold numbers indicate the region where the gene is enriched.

enriched in	Probe Set	avg diff Amy	avg diff Cb	avg diff Hpc	avg diff OB	avg diff PAG
Amy	D49658_at	**114.0**	-19.1	-87.7	27.0	-17.6
Amy	M88354_s_at	**4161.0**	-1406.4	-790.7	-845.3	-748.9
Amy	u26460_s_at	**362.0**	-559.3	43.9	-543.6	-106.7
Amy	ET62460_i_at	**2943.0**	641.7	633.3	266.2	713.4
Amy	Msa.27328.0_	**177.0**	17.5	2.6	-75.6	23.1
Amy	Msa.38650.0_	**1026.0**	-418.6	-421.7	-276.1	-411.6
Amy	Msa.38727.0_	**351.0**	-70.2	-147.2	-36.2	15.7
Amy	U96386_s_at	**2485.0**	-166.7	25.0	-1.1	48.9
Amy	x76653_s_at	**551.0**	94.0	23.6	150.1	-13.8
Amy	TC15611_at	**815.0**	220.8	175.0	117.4	208.0
Amy	TC16820_at	**169.0**	-85.6	-226.6	-88.6	-109.1
Amy	TC17395_at	**917.0**	-7.4	83.1	149.7	61.2
Amy	TC17396_at	**86.0**	-1.1	-0.8	8.4	15.3
Amy	TC19868_at	**81.0**	20.1	-90.3	-28.7	4.1
Amy	TC26406_r_at	**746.0**	104.4	172.6	107.4	83.9
Amy	TC27128_at	**579.0**	-136.6	-110.2	-150.6	-444.8
Amy	TC29518_s_at	**3801.0**	499.9	479.3	1044.8	854.9
Amy	TC29994_at	**3039.0**	-203.6	751.1	729.6	796.2
Amy	TC30095_g_at	**1369.0**	101.8	121.2	323.4	362.9
Amy	TC30886_at	**1605.0**	63.1	0.0	245.2	282.2
Amy	TC31535_at	**840.0**	-1458.3	-1199.2	-23.3	-1.0
Amy	TC35282_at	**2359.0**	-6.9	578.1	129.6	498.8
Amy	TC35411_at	**1784.0**	-8.3	29.1	246.5	-224.2
Amy	TC35462_at	**2249.0**	-409.9	-534.5	173.8	-16.8
Amy	TC35509_at	**3393.0**	191.2	221.4	151.7	965.3
Amy	TC36086_at	**370.0**	-370.0	-455.8	-2251.9	-1167.9

[1] A description of Affymetrix Probe sets (i.e. gene name and Genbank Accession) is shown in Table 3 online at http://www.its.caltech.edu/~mariela/gene_screen.html
TIGR cluster probe sets are named based on their cluster identifiers. For e.g. probe set TC11111_at represents TIGR cluster TC11111.

Amy	TC36144_at	**3240.0**	663.0	856.3	719.0	777.0
Amy	TC37197_at	**2824.0**	-3642.4	-2970.8	-1725.6	-467.7
Amy	TC37997_g_at	**1421.0**	-925.7	-980.1	-769.6	-131.9
Amy	TC38200_at	**1250.0**	-495.2	148.5	52.1	227.8
Amy	TC38589_g_at	**1059.0**	-90.8	167.5	196.0	254.2
Amy	TC39813_g_at	**406.0**	-731.8	78.6	-518.3	-52.8
Amy	TC41875_at	**530.0**	-405.8	-598.5	-903.9	-243.4
Cb	AA009039_at	366.0	**6939.3**	1341.6	895.3	811.1
Cb	aa035912_at	398.0	**5983.2**	955.1	448.6	843.2
Cb	aa035912_g_	957.0	**3802.4**	1031.9	679.5	704.4
Cb	aa111276_s_	-161.0	**383.4**	-215.2	-85.9	-124.3
Cb	AA177965_at	78.0	**290.7**	-57.6	7.0	42.5
Cb	aa185911_s_	-3.0	**137.7**	-60.3	-16.0	-65.3
Cb	aa199418_s_	-88.0	**270.6**	-90.4	-123.9	-58.0
Cb	aa217487_s_	61.0	**1046.0**	23.3	121.9	183.4
Cb	aa266033_s_	184.0	**827.0**	95.9	201.8	124.3
Cb	AA266172_at	421.0	**2165.5**	359.0	406.7	236.2
Cb	aa266791_s_	674.0	**2376.8**	483.7	235.8	374.0
Cb	aa289572_s_	36.0	**1092.8**	43.9	173.9	108.8
Cb	aa472865_s_	656.0	**3264.1**	520.7	738.4	594.6
Cb	aa473309_f_at	781.0	**4446.8**	361.8	864.3	771.7
Cb	aa529056_g_	592.0	**3764.1**	681.1	962.2	1062.8
Cb	aa562600_s_	-51.0	**235.2**	-146.6	-22.0	-532.4
Cb	AA589492_R	15.0	**182.6**	4.1	31.0	4.1
Cb	AA617336_R	394.0	**1435.1**	371.4	393.7	364.6
Cb	aa624821_at	84.0	**500.0**	91.8	137.9	52.8
Cb	AA688923_at	-1351.0	**694.1**	-1282.7	-667.5	-573.9
Cb	AA688944_at	-343.0	**308.8**	-375.5	-254.8	-377.1
Cb	AB000777_g_	-516.0	**1242.0**	-900.3	-326.7	-401.9
Cb	af016697_g_at	1390.0	**7592.3**	1097.7	1412.9	1635.7
Cb	AF031816_at	85.0	**1483.8**	372.7	212.8	178.2
Cb	c76668_RC_a	66.0	**255.3**	17.8	51.0	54.9
Cb	C77771_RC_a	-13.0	**369.1**	-34.3	100.9	-75.6
Cb	c78385_RC_a	13.0	**492.4**	30.1	103.9	110.8
Cb	c78995_RC_a	-524.0	**592.8**	57.6	-327.7	98.4
Cb	C80068_RC_a	-166.0	**301.2**	15.1	15.0	-122.2
Cb	D19038_RC_a	-641.0	**231.4**	20.6	-493.6	-753.1
Cb	d31898_s_at	1637.0	**5973.6**	945.6	762.4	909.5
Cb	D83262_at	337.0	**1786.9**	171.3	132.9	439.2
Cb	D84391_f_at	1099.0	**5649.5**	1330.6	1447.8	1251.3
Cb	L02241_s_at	128.0	**596.6**	-16.4	118.9	119.1
Cb	L10426_s_at	172.0	**1585.2**	216.5	281.8	69.4

Cb	L12147_s_at	22.0	**677.9**	-79.5	-7.0	74.6
Cb	l12705_s_at	351.0	**4090.1**	376.9	254.8	557.3
Cb	l28177_s_at	-253.0	**222.8**	-102.8	39.0	28.0
Cb	l35029_s_at	164.0	**2566.1**	420.7	376.7	78.7
Cb	m21531_s_at	2944.0	**11351.0**	1312.8	450.6	1562.1
Cb	M21532_s_at	182.0	**15783.0**	-78.1	224.8	167.8
Cb	M90388_s_at	21.0	**425.5**	48.0	41.0	-5.2
Cb	r74641_s_at	34.0	**4633.2**	-39.7	-50.0	1228.6
Cb	U19860_s_at	683.0	**2557.5**	485.1	607.5	336.7
Cb	U28068_s_at	-17.0	**7939.3**	243.9	95.9	6.2
Cb	u58112_s_at	-38.0	**165.4**	4.1	-26.0	-91.2
Cb	Msa.11910.0_	115.0	**1064.0**	-56.5	88.7	183.7
Cb	Msa.1375.0_g	170.0	**1168.0**	287.7	0.0	89.5
Cb	Msa.1461.0_a	121.0	**1630.5**	113.0	27.4	117.2
Cb	Msa.18074.0_	1477.0	**9112.4**	2493.7	2162.6	1799.6
Cb	Msa.2220.0_f	-3.0	**906.1**	-49.9	247.6	79.4
Cb	Msa.24313.0_	-265.0	**3480.3**	660.9	589.4	180.0
Cb	Msa.2440.0_i	-35.0	**757.0**	13.1	47.1	139.4
Cb	Msa.24555.0_	-32.0	**213.1**	-240.4	-86.5	53.5
Cb	Msa.24665.0_	-237.0	**129.1**	-15.8	-391.1	-226.1
Cb	Msa.2619.0_s	1731.0	**32818.0**	2417.5	739.5	3855.7
Cb	Msa.28697.0_	-163.0	**263.2**	-177.4	25.2	-186.4
Cb	Msa.326.0_g_	-69.0	**784.5**	-160.3	179.7	-27.7
Cb	Msa.3299.0_s	-118.0	**1998.9**	462.5	-4.4	80.3
Cb	Msa.419.0_s_	-185.0	**3761.0**	392.8	169.8	-110.7
Cb	Msa.43183.0_	1345.0	**6317.6**	1646.2	1474.6	1541.2
Cb	Msa.549.0_s_	-64.0	**206.8**	14.5	15.3	49.8
Cb	Msa.728.0_at	117.0	**557.7**	-73.6	46.0	103.4
Cb	S74567_g_at	-114.0	**1125.4**	-107.7	290.3	-157.8
Cb	w41032_s_at	1667.0	**8552.2**	1913.0	1438.4	1845.7
Cb	w82359_g_at	146.0	**1663.1**	2.6	155.6	-29.5
Cb	X15373-2_s_a	2128.0	**23082.0**	1824.9	403.2	530.7
Cb	x16490_s_at	-23.0	**210.6**	-25.0	18.6	-19.4
Cb	X51986_s_at	467.0	**24775.0**	290.4	212.5	316.5
Cb	x60304_at	1432.0	**6756.3**	1384.8	573.0	703.2
Cb	X61397_s_at	220.0	**19185.0**	-76.2	-21.9	438.4
Cb	X61448_s_at	951.0	**19231.0**	921.0	3040.1	2143.8
Cb	x63963_s_at	297.0	**2423.8**	147.2	650.8	250.1
Cb	X65128_s_at	333.0	**2119.2**	415.2	525.9	452.2
Cb	x67141_at	778.0	**23721.0**	1616.0	186.2	1942.6
Cb	x67141_g_at	-308.0	**5493.0**	751.5	340.7	81.2
Cb	X78197_s_at	255.0	**3271.0**	101.2	292.5	498.4

Cb	y00305_s_at	293.0	**13892.0**	3417.3	479.9	833.4
Cb	Y11666_s_at	84.0	**1259.5**	270.7	230.1	198.4
Cb	TC14307_s_at	-582.0	**2632.7**	262.1	-117.4	346.7
Cb	TC15800_at	604.0	**6981.1**	1719.3	565.3	1242.1
Cb	TC16082_at	657.0	**3443.0**	909.7	956.9	628.2
Cb	TC16546_at	25.0	**287.4**	29.0	-63.5	-26.5
Cb	TC16732_at	406.0	**2157.3**	496.0	446.7	447.7
Cb	TC17153_at	-71.0	**101.4**	9.7	-142.5	-210.1
Cb	TC17287_at	-13.0	**154.2**	36.3	-14.4	-126.5
Cb	TC17839_at	-251.0	**502.9**	-165.3	-64.7	-147.9
Cb	TC18167_at	69.0	**1856.2**	26.6	82.6	158.1
Cb	TC18873_at	-411.0	**121.5**	-120.2	-231.1	-426.3
Cb	TC21031_at	1142.0	**5816.9**	1608.8	1013.1	197.8
Cb	TC21707_at	-134.0	**1194.9**	260.5	220.4	193.8
Cb	TC22115_at	-55.0	**1208.6**	0.8	4.8	-79.5
Cb	TC23559_at	364.0	**2801.7**	471.0	449.1	372.2
Cb	TC23741_at	65.0	**738.5**	-91.1	22.8	-43.8
Cb	TC23801_at	-2.0	**946.6**	-21.8	167.7	231.5
Cb	TC23801_g_at	397.0	**5743.0**	350.0	804.8	734.2
Cb	TC24676_at	43.0	**170.1**	-146.8	46.7	-123.4
Cb	TC24816_s_at	-124.0	**586.3**	-102.4	-75.4	-67.3
Cb	TC24926_at	600.0	**9482.8**	411.3	1422.7	1007.5
Cb	TC25240_at	-109.0	**165.9**	-59.7	-8.4	-139.7
Cb	TC25270_at	-56.0	**316.9**	-92.7	-104.2	-131.6
Cb	TC26229_at	-397.0	**12669.0**	-409.5	-344.4	1797.9
Cb	TC27828_at	-333.0	**331.1**	-207.5	-49.0	-224.5
Cb	TC28671_at	607.0	**2365.3**	602.4	324.5	396.5
Cb	TC29334_f_at	771.0	**17642.0**	578.5	1557.3	949.3
Cb	TC29445_at	191.0	**2912.8**	102.8	625.7	483.6
Cb	TC30591_at	-70.0	**636.4**	121.2	85.2	146.9
Cb	TC30958_at	2536.0	**13267.0**	1719.0	2902.2	3014.7
Cb	TC31108_at	529.0	**3969.2**	1122.1	1128.9	628.3
Cb	TC31518_at	16.0	**727.9**	99.2	137.8	66.1
Cb	TC31783_at	359.0	**3756.6**	363.6	85.2	451.1
Cb	TC31991_at	-11.0	**753.6**	82.6	42.0	91.3
Cb	TC32216_s_at	953.0	**6455.6**	1685.9	1092.7	439.5
Cb	TC32340_at	-428.0	**5895.2**	3.7	1306.3	-613.6
Cb	TC32450_at	-6.0	**345.3**	11.0	96.9	68.2
Cb	TC32531_at	-331.0	**1284.4**	-91.8	181.0	-606.3
Cb	TC32703_at	-488.0	**394.2**	-745.6	-638.6	-380.8
Cb	TC32804_s_at	86.0	**698.3**	183.7	145.9	-1.0
Cb	TC32850_g_at	-1061.0	**167.5**	-1309.4	-87.6	-119.6

Cb	TC33451_at	41.0	**1222.6**	-49.6	192.6	-89.2
Cb	TC33667_at	2199.0	**10253.0**	1465.5	1950.8	1440.2
Cb	TC33746_at	397.0	**13043.0**	216.7	309.4	371.3
Cb	TC33767_at	314.0	**1556.2**	405.9	316.4	-115.4
Cb	TC34215_at	96.0	**425.1**	-51.4	57.2	16.8
Cb	TC34607_s_at	-64.0	**882.5**	156.1	155.3	-74.5
Cb	TC34901_at	-318.0	**512.7**	-391.2	123.8	-395.5
Cb	TC35087_at	-140.0	**209.1**	-206.8	-110.6	13.2
Cb	TC35235_at	456.0	**6859.7**	311.6	1237.3	-1283.1
Cb	TC36041_at	106.0	**945.0**	101.9	37.9	1.2
Cb	TC36118_g_at	558.0	**4400.3**	817.0	376.1	489.2
Cb	TC36594_at	57.0	**540.6**	61.2	129.6	10.8
Cb	TC36605_at	1220.0	**5238.0**	1057.3	1431.7	1385.0
Cb	TC36788_at	333.0	**2004.1**	230.1	-254.4	299.8
Cb	TC37170_at	587.0	**3101.8**	125.2	858.1	855.0
Cb	TC37531_at	651.0	**5691.9**	1233.5	1106.2	496.4
Cb	TC37804_at	-243.0	**239.3**	67.0	-369.8	-16.8
Cb	TC37834_s_at	81.0	**917.5**	170.4	110.6	161.9
Cb	TC38320_at	27.0	**542.0**	-53.9	118.5	-57.6
Cb	TC38584_at	1408.0	**7781.3**	977.2	829.6	1506.1
Cb	TC39605_at	879.0	**5017.9**	1280.1	538.9	1317.8
Cb	TC39631_at	65.0	**343.9**	88.8	-143.8	54.0
Cb	TC40036_at	25.0	**2115.6**	-55.3	-159.6	525.2
Cb	TC40154_at	187.0	**2291.6**	33.5	33.2	-10.8
Cb	TC40249_at	63.0	**1423.7**	355.3	47.4	55.2
Cb	TC40339_at	1068.0	**7166.5**	639.3	1466.5	1465.3
Cb	TC40661_at	431.0	**2790.9**	-468.9	-1261.0	729.1
Cb	TC40833_at	187.0	**1492.4**	-214.1	-55.3	-488.0
Cb	TC40877_f_at	446.0	**2086.7**	477.7	414.0	406.5
Cb	TC41060_at	422.0	**9912.0**	234.5	1158.3	509.6
Cb	TC41163_at	856.0	**4624.5**	888.3	711.1	705.1
Cb	TC41216_at	-382.0	**2122.4**	196.6	-523.1	202.7
Cb	TC41320_at	478.0	**1742.8**	339.3	-45.8	491.6
Cb	TC41324_at	139.0	**734.5**	-116.5	-516.7	100.7
Cb	TC41351_at	-24.0	**195.3**	14.6	-1.6	16.8
Cb	TC41712_at	16.0	**240.7**	58.3	-1199.4	-74.3
Cb	TC41829_at	-154.0	**206.3**	-211.2	-112.2	-182.3
Cb	TC42055_at	878.0	**5313.7**	1338.3	1267.4	1509.7
Cb	TC42156_at	-82.0	**1983.5**	23.3	-2293.0	-721.9
Cb	TC42157_at	34.0	**580.5**	-128.2	-6223.1	-1393.4
Hpc	aa000469_g_	1561.0	1819.4	**6809.3**	1161.1	1414.0
Hpc	AA023445_at	-109.0	-209.4	**601.6**	44.0	-120.2

Hpc	aa217006_s_	-220.0	-615.7	**374.1**	-484.6	-153.3
Hpc	aa288448_at	247.0	-263.9	**922.3**	-179.9	-176.1
Hpc	aa408337_RC	202.0	182.6	**1011.3**	145.9	104.6
Hpc	aa408837_RC	-1.0	-48.8	**223.4**	-12.0	-12.4
Hpc	aa422456_s_	42.0	38.2	**231.6**	32.0	48.7
Hpc	aa590472_at	261.0	315.5	**1770.5**	125.9	148.1
Hpc	aa683731_s_	1561.0	570.8	**6205.0**	318.7	314.9
Hpc	AF006492_at	152.0	312.6	**1214.1**	111.9	138.8
Hpc	AF006492_g_	77.0	155.8	**1771.9**	66.9	-20.7
Hpc	AF026072_at	-129.0	-241.9	**238.4**	-42.0	-355.3
Hpc	AFFX-Gapdh	-155.0	-295.4	**215.2**	-198.8	-132.6
Hpc	D30785_s_at	63.0	-99.4	**678.3**	60.0	-97.4
Hpc	T25564_RC_a	-139.0	128.1	**634.5**	140.9	-204.1
Hpc	u29086_s_at	175.0	22.9	**822.2**	-42.0	-1.0
Hpc	U65986_s_at	-75.0	-21.0	**270.0**	28.0	-68.4
Hpc	AFFX-Gapdh	1.0	522.6	**1989.2**	-12.1	-90.4
Hpc	AFFX-Gapdh	-885.0	-337.1	**555.8**	-590.5	-616.5
Hpc	ET61808_i_at	-82.0	-174.2	**160.3**	-39.4	-169.8
Hpc	ET62782_at	-74.0	-84.0	**474.3**	-434.9	-119.1
Hpc	Msa.12575.0_	-52.0	-270.7	**747.6**	-209.3	52.6
Hpc	Msa.12975.0_	-222.0	-853.5	**306.1**	17.5	-481.7
Hpc	Msa.18389.0_	83.0	-294.5	**1211.4**	-353.9	-306.4
Hpc	Msa.19442.0_	-276.0	-457.4	**486.1**	-31.8	-226.1
Hpc	Msa.21961.0_	-236.0	-15.0	**120.9**	5.5	-71.1
Hpc	Msa.41663.0_	-166.0	-124.1	**287.7**	60.3	-96.0
Hpc	Msa.6704.0_s	-210.0	-166.7	**367.9**	-188.4	-106.1
Hpc	Msa.6742.0_f	156.0	401.0	**2426.7**	463.4	16.6
Hpc	X61800_s_at	-323.0	76.4	**913.1**	-37.2	-407.9
Hpc	Z50192_s_at	-232.0	-190.5	**174.7**	32.9	-532.5
Hpc	AFFX-Gapdh	-164.0	54.9	**784.7**	-212.0	-305.9
Hpc	TC14336_at	-51.0	-74.0	**385.5**	-103.0	-1.0
Hpc	TC14608_at	-114.0	493.4	**2320.9**	33.5	142.8
Hpc	TC14829_s_at	1784.0	366.6	**7511.9**	2125.7	1831.5
Hpc	TC15061_at	85.0	-49.7	**2616.1**	497.0	-28.6
Hpc	TC15245_g_at	222.0	380.3	**1420.1**	182.0	183.6
Hpc	TC16424_g_at	436.0	171.2	**3059.6**	168.9	162.1
Hpc	TC16425_s_at	427.0	26.4	**2428.2**	52.7	-18.4
Hpc	TC17256_at	-173.0	-157.4	**141.1**	-209.6	-28.6
Hpc	TC17444_i_at	-196.0	-249.3	**240.3**	-180.8	-194.8
Hpc	TC17850_at	-209.0	-40.1	**162.9**	-354.5	-345.7
Hpc	TC17980_g_at	151.0	-278.9	**1328.2**	-142.5	-25.5
Hpc	TC18982_g_at	155.0	-2.1	**2387.8**	180.8	-182.5

Hpc	TC19176_at	-138.0	183.8	**656.4**	74.2	79.5
Hpc	TC19193_at	-99.0	-66.6	**363.7**	-81.4	3.1
Hpc	TC19710_g_at	247.0	225.0	**1446.7**	293.4	183.6
Hpc	TC19871_at	-120.0	-324.3	**312.1**	-310.2	27.5
Hpc	TC20006_at	-286.0	-805.0	**367.7**	-293.4	-512.9
Hpc	TC20089_s_at	-198.0	37.0	**236.3**	-161.7	-305.9
Hpc	TC20098_f_at	611.0	435.3	**3242.7**	614.4	758.7
Hpc	TC20098_i_at	74.0	182.8	**1252.4**	148.5	283.5
Hpc	TC20339_at	-236.0	-175.4	**465.3**	0.0	-244.7
Hpc	TC20454_at	58.0	27.5	**271.8**	-13.2	40.8
Hpc	TC20506_at	27.0	47.5	**181.5**	35.9	12.2
Hpc	TC20678_at	-283.0	109.9	**433.9**	116.2	-144.8
Hpc	TC20709_at	-32.0	84.5	**366.9**	55.1	98.9
Hpc	TC20738_at	333.0	232.4	**2180.6**	344.9	550.7
Hpc	TC21082_at	-79.0	-8.5	**202.4**	-15.6	-32.6
Hpc	TC21313_at	-696.0	356.0	**1677.4**	-239.5	-707.7
Hpc	TC22106_at	-79.0	5.3	**537.1**	92.2	138.7
Hpc	TC22286_at	-50.0	-197.6	**146.8**	-37.1	-124.4
Hpc	TC22963_at	524.0	339.1	**1858.8**	101.8	361.0
Hpc	TC23320_f_at	-360.0	-396.2	**690.3**	-419.2	-225.4
Hpc	TC23337_at	-112.0	-195.5	**178.2**	-86.2	-237.6
Hpc	TC23506_at	42.0	153.2	**754.0**	79.0	31.6
Hpc	TC23617_f_at	-154.0	18.0	**261.3**	-401.2	1.0
Hpc	TC23836_at	-28.0	45.4	**294.4**	69.5	-73.4
Hpc	TC24276_at	-71.0	61.3	**214.5**	2.4	-40.8
Hpc	TC24724_s_at	239.0	-668.7	**1239.5**	80.2	-309.0
Hpc	TC24742_at	1208.0	557.8	**6299.8**	658.7	455.8
Hpc	TC24876_at	-19.0	-5.3	**116.9**	-27.5	-76.5
Hpc	TC25056_at	-110.0	3.2	**249.2**	-14.4	-124.4
Hpc	TC25500_at	215.0	412.0	**1824.1**	304.2	235.6
Hpc	TC25822_at	105.0	100.5	**569.3**	-45.5	27.3
Hpc	TC26202_s_at	260.0	-347.8	**1368.2**	1.2	231.8
Hpc	TC27162_at	318.0	559.1	**2227.7**	-19.8	353.5
Hpc	TC28297_at	28.0	59.3	**214.9**	26.9	34.6
Hpc	TC32654_at	47.0	41.2	**181.8**	-28.0	30.4
Hpc	TC33633_at	-44.0	-20.6	**121.2**	-54.9	5.2
Hpc	AFFX-Gapdh	-955.0	332.9	**1208.7**	-322.4	-885.0
Hpc	TC35302_at	-138.0	-359.0	**1052.9**	-662.1	61.2
Hpc	TC36417_i_at	-2.0	-4.1	**3326.1**	-50.6	45.6
Hpc	TC36790_s_at	92.0	386.5	**2520.8**	165.9	333.4
Hpc	TC37471_at	-583.0	-1008.3	**231.6**	-510.4	-1409.0
Hpc	TC37672_f_at	-150.0	-1557.1	**831.5**	-94.8	-769.8

Hpc	TC40680_g_at	-96.0	-242.1	**273.8**	53.7	-244.6
Hpc	TC41370_s_at	129.0	-121.1	**969.9**	170.7	-160.7
Hpc	TC41807_at	58.0	-143.1	**525.7**	-107.5	-103.1
OB	aa002925_s_	244.0	-14.3	293.3	**1126.1**	97.4
OB	aa015322_s_	-33.0	-83.2	61.7	**527.6**	137.8
OB	aa028446_s_	151.0	436.0	196.0	**1944.4**	-139.8
OB	aa174489_at	18.0	721.9	300.1	**3508.2**	128.5
OB	AA691224_at	46.0	-28.7	-9.6	**204.8**	-64.2
OB	C79906_RC_a	236.0	323.2	249.4	**1525.8**	122.2
OB	C79906_RC_g	447.0	637.7	483.7	**3669.0**	276.6
OB	C80836_RC_a	-27.0	-28.7	-8.2	**139.9**	-539.7
OB	k01700-2_s_a	-21.0	-47.8	-64.4	**302.8**	34.2
OB	M32745_s_at	-75.0	26.8	-108.3	**212.8**	57.0
OB	M69200_s_at	621.0	370.0	513.9	**6243.9**	1503.1
OB	m70642_s_at	175.0	202.7	130.2	**1928.4**	74.6
OB	u03283_s_at	-6.0	88.9	28.8	**392.7**	5.2
OB	u37465_s_at	493.0	236.2	405.6	**1816.5**	455.8
OB	u51000_s_at	458.0	-312.6	270.0	**8852.8**	-230.0
OB	u57343_s_at	242.0	59.3	-80.9	**1575.7**	53.9
OB	u67840_s_at	364.0	70.8	112.4	**1573.7**	31.1
OB	u68058_s_at	38.0	74.6	276.8	**2172.2**	-179.2
OB	u87456_s_at	-6.0	-63.1	-63.0	**466.6**	-13.5
OB	U88566_s_at	-175.0	-8.6	-34.3	**158.9**	-184.4
OB	ET61677_at	-476.0	-114.1	-211.5	**783.3**	-145.8
OB	ET61693_at	-25.0	-16.3	-19.7	**185.2**	-21.2
OB	Msa.11014.0_	59.0	67.7	-106.4	**270.6**	34.1
OB	Msa.2414.0_a	8.0	-50.1	-86.7	**168.7**	27.7
OB	Msa.25279.0_	-102.0	62.7	61.8	**478.8**	55.4
OB	Msa.27849.0_	-185.0	-676.8	14.5	**267.3**	-162.4
OB	Msa.3708.0_s	-16.0	-1599.1	-1781.6	**679.2**	172.6
OB	Msa.3810.0_s	1076.0	26.3	166.9	**5069.1**	522.3
OB	Msa.4422.0_s	93.0	307.1	-48.6	**9235.4**	196.6
OB	Msa.444.0_at	137.0	-115.3	-461.2	**809.6**	-1.8
OB	Msa.4623.0_s	-282.0	-734.4	-869.8	**383.4**	-658.0
OB	Msa.544.0_s_	9.0	60.2	2.6	**508.3**	6.5
OB	U96700_s_at	21.0	-134.1	-84.1	**208.2**	-74.8
OB	w84196_s_at	172.0	185.5	-57.8	**1229.2**	77.5
OB	TC14785_f_at	-282.0	188.1	-150.0	**670.6**	63.2
OB	TC15770_g_at	-305.0	-365.5	-854.0	**142.5**	-377.3
OB	TC15936_g_at	-24.0	-45.4	-39.5	**228.7**	48.9
OB	TC16343_at	-110.0	-70.8	-58.9	**104.2**	-135.6
OB	TC17513_at	-22.0	10.6	-72.6	**221.6**	-48.9

OB	TC18381_at	39.0	7.4	-28.2	**263.5**	-36.7
OB	TC18855_f_at	-87.0	-91.9	-141.9	**796.4**	-59.1
OB	TC18964_f_at	-11.0	-47.5	-155.6	**319.8**	53.0
OB	TC19736_at	-515.0	-913.8	-791.9	**617.9**	-801.5
OB	TC20174_at	-883.0	-1170.6	-1375.8	**3470.5**	-238.6
OB	TC20214_at	26.0	-24.3	-155.6	**130.5**	22.4
OB	TC20543_s_at	36.0	15.8	0.8	**843.1**	79.5
OB	TC20573_at	-22.0	-61.3	-63.7	**238.3**	-50.0
OB	TC20950_at	-737.0	-1186.4	-441.1	**2535.2**	-1368.5
OB	TC21085_at	-41.0	-8.5	-67.7	**196.4**	-106.1
OB	TC22268_at	202.0	-19.0	127.4	**835.9**	-142.8
OB	TC22661_at	5.0	-12.7	-57.3	**206.0**	4.1
OB	TC25512_s_at	-258.0	-94.0	-212.1	**556.9**	-240.7
OB	TC26138_at	-86.0	-228.0	-253.4	**216.0**	-26.2
OB	TC26564_at	-3.0	-520.5	-380.2	**172.8**	36.7
OB	TC26753_g_at	-162.0	-1043.5	-1493.1	**213.6**	-289.5
OB	TC26829_i_at	26.0	-10.3	-95.5	**308.2**	-45.1
OB	TC27099_s_at	2104.0	1681.2	2042.2	**8152.1**	1646.8
OB	TC27790_at	82.0	-179.1	-141.4	**471.6**	82.9
OB	TC28487_at	-124.0	-137.9	-62.4	**374.7**	-115.4
OB	TC28608_at	639.0	95.3	444.4	**2785.5**	547.6
OB	TC29521_at	-515.0	-779.4	-202.0	**855.7**	-628.3
OB	TC29675_at	262.0	195.8	-194.7	**990.0**	219.2
OB	TC30868_f_at	29.0	-105.6	-113.9	**230.0**	-93.4
OB	TC31261_at	658.0	240.9	319.6	**3278.1**	514.0
OB	TC31261_g_at	112.0	136.6	51.4	**1097.4**	143.7
OB	TC31401_at	110.0	-103.1	62.4	**1159.2**	-19.9
OB	TC31604_at	62.0	146.9	-468.3	**680.6**	-25.2
OB	TC31996_at	-213.0	64.4	-323.2	**410.9**	25.2
OB	TC32133_at	-52.0	18.0	-49.6	**232.3**	-8.4
OB	TC32211_g_at	-699.0	-182.9	-1050.5	**239.3**	-174.1
OB	TC32255_g_at	13.0	46.4	14.7	**358.4**	39.9
OB	TC32944_at	-296.0	-297.6	-183.7	**166.9**	14.7
OB	TC33215_at	37.0	37.4	95.5	**342.1**	89.2
OB	TC33295_at	-8.0	-235.8	-657.5	**605.9**	-67.1
OB	TC33678_s_at	-71.0	-333.7	-576.7	**166.9**	-332.5
OB	TC33964_at	335.0	897.9	247.9	**3487.1**	291.6
OB	TC34316_at	-308.0	-186.8	-213.0	**693.5**	-16.8
OB	TC34405_g_at	-328.0	-27.1	-598.7	**248.7**	-270.6
OB	TC34509_at	17.0	346.6	-145.1	**1901.7**	-38.8
OB	TC34509_g_at	357.0	927.6	139.6	**3314.3**	578.0
OB	TC34531_f_at	-34.0	12.9	-117.5	**240.5**	-2.1

OB	TC34963_at	-327.0	-206.3	-107.8	**2008.5**	140.3
OB	TC35065_at	616.0	396.2	410.7	**2920.3**	581.6
OB	TC35158_at	-334.0	-89.4	-502.4	**591.0**	-1432.9
OB	TC36785_at	663.0	218.7	640.8	**6534.4**	816.6
OB	TC36785_g_at	2295.0	590.1	1306.3	**18514.0**	2767.6
OB	TC37090_at	-716.0	-326.0	-1000.5	**184.9**	-418.5
OB	TC37997_at	78.0	-308.1	-65.5	**293.9**	-136.7
OB	TC38418_at	-31.0	45.4	-65.5	**186.5**	32.4
OB	TC38669_at	-31.0	-2.8	-29.1	**335.0**	2.4
OB	TC39131_g_at	66.0	82.5	-301.5	**1174.1**	-394.5
OB	TC40403_at	155.0	-269.6	139.8	**971.9**	239.8
OB	TC40734_at	-349.0	328.8	148.5	**1717.7**	-230.2
OB	TC40769_at	58.0	148.6	109.2	**628.9**	125.9
OB	TC40908_at	29.0	266.9	-45.1	**1088.8**	-27.6
OB	TC41850_at	-979.0	-1630.0	-1885.9	**1226.3**	-1594.8
OB	TC41873_at	90.0	152.7	294.2	**1651.4**	170.3
OB	TC41873_g_at	378.0	349.4	447.1	**5369.7**	341.8
OB	TC41923_s_at	-783.0	-1072.9	-2273.3	**1733.5**	-780.6
OB	TC41945_at	544.0	551.6	588.3	**2691.2**	659.5
OB	TC42120_at	327.0	255.9	-64.1	**3207.9**	612.8
PAG	AA245183_at	-200.0	46.8	-126.1	-218.8	**556.3**
PAG	aa386903_at	-385.0	-362.4	-135.7	-117.9	**119.1**
PAG	aa734486_g_	-378.0	-306.9	-420.7	-331.7	**397.8**
PAG	AB000096_g_	-53.0	67.9	91.8	-88.9	**368.8**
PAG	af013604_s_at	130.0	293.5	156.2	152.9	**4232.6**
PAG	c79089_RC_s	435.0	1972.4	250.8	263.8	**10187.0**
PAG	d00754_s_at	56.0	-187.4	217.9	-52.0	**909.5**
PAG	D00812_s_at	206.0	243.8	264.5	176.9	**929.2**
PAG	D11091_s_at	412.0	449.4	422.1	437.7	**1641.9**
PAG	D16580_s_at	9.0	308.8	-182.3	290.8	**1203.7**
PAG	D18061_RC_a	-6.0	-34.4	-8.2	21.0	**239.3**
PAG	d89571_at	-355.0	44.0	-1208.7	-280.8	**562.5**
PAG	l07037_s_at	114.0	-238.1	-134.3	-346.7	**801.8**
PAG	M31690_f_at	189.0	-52.6	4.1	221.8	**1272.1**
PAG	ET62843_f_at	63.0	-72.7	-758.1	-158.9	**371.0**
PAG	Msa.15757.0_	-196.0	-55.1	93.3	-116.1	**349.8**
PAG	Msa.17804.0_	-110.0	82.7	-80.1	29.6	**3271.6**
PAG	Msa.2021.0_a	-254.0	-350.9	294.3	-280.5	**1229.3**
PAG	Msa.23858.0_	-133.0	61.4	2.6	-96.4	**263.9**
PAG	Msa.2451.0_s	-42.0	448.7	-512.4	-123.8	**1599.3**
PAG	Msa.24687.0_	-155.0	-446.2	-286.4	-410.8	**650.6**
PAG	Msa.26324.0_	-64.0	86.5	-459.9	-204.9	**594.3**

PAG	Msa.33045.0_	-7.0	-66.4	-1.3	-87.6	**669.1**
PAG	Msa.6114.0_s	426.0	184.2	168.2	282.7	**1629.8**
PAG	Msa.6768.0_s	-4.0	-125.3	0.0	96.4	**641.4**
PAG	Msa.7127.0_s	2065.0	452.4	-1744.8	-87.6	**12361.0**
PAG	Msa.8112.0_s	198.0	-66.4	-138.0	77.8	**859.2**
PAG	Msa.8622.0_s	487.0	22.6	-337.7	61.4	**3094.4**
PAG	W08454_s_at	39.0	107.8	-467.7	-99.7	**419.9**
PAG	W67100_s_at	0.0	84.0	105.1	11.0	**464.2**
PAG	TC14206_at	230.0	280.0	129.0	495.8	**1966.1**
PAG	TC16865_at	621.0	498.7	77.4	355.7	**2775.8**
PAG	TC17400_at	-497.0	-328.6	-688.7	-518.5	**257.0**
PAG	TC18118_g_at	-144.0	-220.8	-190.3	-148.5	**227.4**
PAG	TC19890_g_at	-45.0	-210.2	-1.6	15.6	**179.5**
PAG	TC19903_at	25.0	-63.4	3.2	-170.1	**398.7**
PAG	TC21111_g_at	-56.0	-160.6	-715.3	-288.6	**360.0**
PAG	TC22581_s_at	-15.0	137.3	21.0	-22.8	**633.3**
PAG	TC23551_g_at	5.0	1.1	15.3	18.0	**93.8**
PAG	TC24421_at	-29.0	-10.6	-2.4	13.2	**97.9**
PAG	TC24428_s_at	5.0	-120.4	27.4	21.6	**147.9**
PAG	TC26053_at	271.0	-150.7	-1089.0	269.7	**2427.3**
PAG	TC26201_s_at	1046.0	149.4	989.9	527.7	**4742.3**
PAG	TC27169_g_at	101.0	-435.4	36.7	71.2	**403.8**
PAG	TC29206_at	139.0	-168.8	-156.1	-220.6	**692.3**
PAG	TC29314_at	-41.0	12.9	-473.8	-235.8	**413.3**
PAG	TC30851_at	3527.0	3255.5	1173.5	1571.3	**13556.0**
PAG	TC32701_g_at	260.0	838.7	828.3	-188.0	**5092.6**
PAG	TC33864_g_at	1061.0	405.8	820.9	767.0	**3784.6**
PAG	TC34956_at	166.0	145.8	630.6	505.7	**6532.8**
PAG	TC34957_at	547.0	433.3	225.7	872.3	**4735.3**
PAG	TC35074_at	49.0	-74.3	-16.0	85.3	**356.1**
PAG	TC35078_s_at	424.0	319.1	355.3	499.4	**1823.9**
PAG	TC35230_at	57.0	70.2	71.4	-74.3	**274.6**
PAG	TC35254_at	-163.0	11.0	-93.2	-131.2	**309.4**
PAG	TC35286_at	-134.0	-101.8	-196.6	60.1	**464.1**
PAG	TC35359_at	74.0	181.6	69.9	39.5	**845.4**
PAG	TC35493_at	-120.0	-166.4	46.6	-42.7	**729.1**
PAG	TC35564_at	-1180.0	-354.9	-1668.9	-1517.0	**967.7**
PAG	TC35765_at	305.0	184.3	-94.7	104.3	**1480.9**
PAG	TC36245_at	-134.0	-167.8	-69.9	-28.4	**239.8**
PAG	TC36249_at	122.0	-5.5	43.7	28.4	**678.7**
PAG	TC36868_s_at	-371.0	-66.0	-253.4	37.9	**292.6**
PAG	TC36897_s_at	51.0	46.8	-179.1	-617.9	**191.9**

PAG	TC37333_at	-139.0	2.8	5.8	-167.5	**178.7**
PAG	TC37746_s_at	-250.0	12.4	53.9	-132.7	**278.2**
PAG	TC37944_at	-370.0	-162.3	-538.8	-169.1	**232.6**
PAG	TC38491_at	18.0	59.1	-174.8	-28.4	**256.6**
PAG	TC39975_at	165.0	-371.4	-257.8	33.2	**1008.5**
PAG	TC40689_at	-133.0	286.1	-135.4	-12.6	**2362.3**
PAG	TC40745_at	48.0	-386.5	-565.0	-926.0	**376.5**
PAG	TC40881_at	304.0	-231.1	-371.4	-349.2	**1268.7**
PAG	TC41736_at	-154.0	-99.0	-297.1	-2024.3	**1855.0**

Table 1-4: Categories of genes.

Of the 455 total genes and ESTs differentially enriched, 117 genes were annotated and classified based on molecular function. Guidelines from the Gene Ontology (GO) Consortium were followed for classification. (A): the number of genes enriched at least 3.5-fold and the number of annotated genes among those in each region is presented. The percentage of annotated genes relative to enriched genes is also shown. (B): number of total genes identified and percentage in each region are shown for each category. Highlighted categories are the most represented. Abbreviations are as in table 3.

(A)

	Amy	Cb	Hpc	OB	PAG	total
enriched genes	33	159	89	101	73	455
# annotated genes	8	47	17	24	21	117
%annotated genes / enriched	24	30	19	24	29	26

(B)

molecular function	total # genes	% genes	% in Amy	% in Cb	% in Hpc	% in OB	% in PAG
motor protein	1	1	13	0	0	0	0
ribonuclear	2	2	0	4	0	0	0
nuclear	1	1	0	2	0	0	0
signaling	30	26	25	30	0	29	33
transporter	3	3	0	4	0	0	5
ligand binding or carrier	8	7	0	9	6	8	5
DNA binding	20	17	50	17	12	17	10
structural	10	9	0	6	18	17	0
enzyme	18	15	0	15	18	8	29
enzyme inhibitor	4	3	0	4	0	4	5
membrane protein	6	5	13	4	0	4	10
ubiquitin	2	2	0	0	6	4	0
immunity, defense protein	1	1	0	0	6	0	0

zinc finger protein	2	2	0	0	12	0	0
cell adhesion	1	1	0	0	6	0	0
chaperone	1	1	0	0	6	0	0
translation	5	4	0	2	6	8	5
cell cycle regulator	1	1	0	2	0	0	0
transcription factor binding	1	1	0	0	6	0	0

1.4.4 Figure legends

Figure 1-1: Distribution of normalized average difference values in each sample.

(A) Distribution of normalized average difference ($\bar{\Delta}$) values in each sample. Only the 11ka array is plotted here, but all other chips had similar distributions. For clarity, only values below 7000 are shown. The dotted line indicates the mean value (1104). Bin size is 200. (B) Plots of average difference values for all possible two-way comparisons. The first row has the amygdala (A) on the x-axis. The second row are plots with the cerebellum on the x-axis, the third and fourth rows have the hippocampus and olfactory bulb, respectively, on the x-axis. The y-axes are the cerebellum (C), hippocampus (H), olfactory bulb (O) and PAG (P), from the first to fourth columns, respectively. For clarity, average difference values were divided by 1000 and only values below 30,000 are shown. The dashed line indicates a slope=1.

Figure 1-2: *In situ* hybridization of amygdala-enriched genes.

(A) Nissl staining of a coronal section (left side of the brain). To the right, a schematic representation of various amygdala subnuclei is shown. Cortical-like nuclei (lateral, basolateral and cortical) are shown in blue. Striatal-like subdivisions (central and medial) are in yellow, and the basomedial region is in orange (BMP= basomedial, posterior. BMA= basomedial, anterior). (B-D) Low magnification pictures of the left hemibrain. Amygdala-details are shown in the magnified area (boxes). To the right,

computer-aided schematics of staining in the amygdaloid region. Note that the nuclear boundaries vary slightly depending on the axial level. Color boundaries of subnuclei follow the diagram from A. (B) **Probe 29** (activin receptor type II, TIGR identifier TC35462). Intense labeling in the lateral, basomedial and cortical amygdala is apparent (black arrows). Note that the medial nucleus is devoid of staining (white arrow). No signal was detected in the cerebellum or PAG. Very few cells were stained in the olfactory bulb (not shown). A sense probe (not shown) labeled the hippocampus and piriform cortex (arrowheads) in the same way as the antisense probe, so the signal in these regions may be mainly due to non-specific hybridization. (C) **Probe 41** (laminin ß3, Genbank accession U43298). Signal is visible in the medial amygdala (black arrow) and ventromedial hypothalamus (white arrow). No staining was detected in cerebellum, hippocampus, olfactory bulb, and PAG (not shown). (D) **Probe 4** (arp-1, Genbank accession X76653). Strong signal is detected in the lateral and basolateral complexes (black arrow). Note also weaker signal in the medial amygdala (white arrow). The reticular thalamic nucleus also showed clear hybridization (arrowhead). No staining was detected in the other four regions examined on microarrays (not shown).

Figure 1-3: Expression of amygdala-enriched genes in different amygdaloid subnuclei.

(A) **Probe 75** (unconventional type myosin, TIGR identifier TC37197). Note the sharp discontinuity in expression levels between the lateral (arrow), and basolateral (arrowhead) nuclei. Staining was also observed in cortical layers 2/3 (white arrow). No staining was detected in the cerebellum, olfactory bulb or PAG (not shown). (B) **Probe 45-6** (Lhx6, Genbank accession AB031040). Lhx 6 hybridized to many scattered cells in

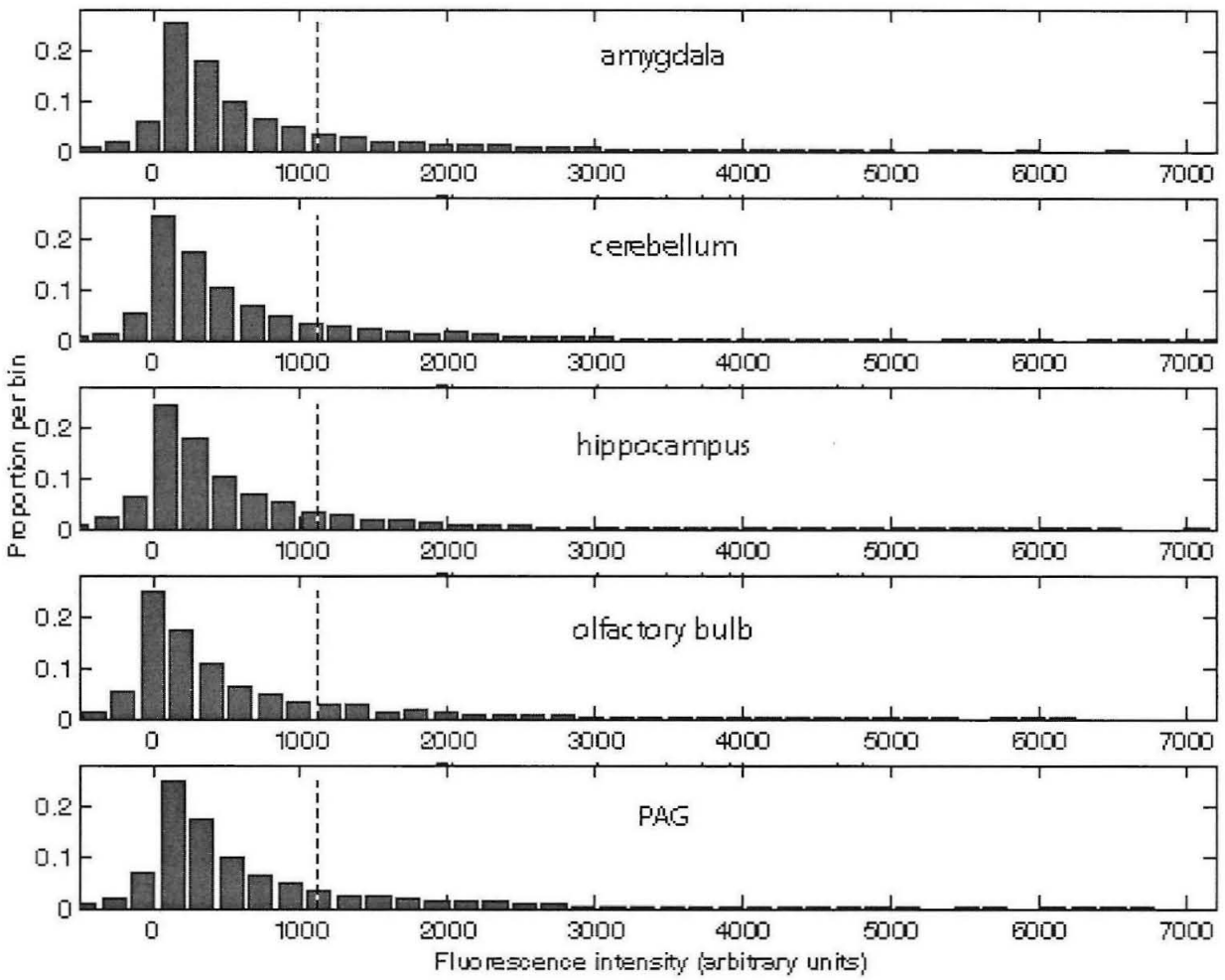
the forebrain, and was particularly concentrated in the dorsal aspect of the medial amygdala (arrow); the cerebellum was unlabeled (not shown). Lhx 6 was not represented on the microarray, but was analyzed due to its coexpression with Lhx7 (Grigoriou et al., 1998; Zhao et al., 1999), which was also enriched in the amygdala (not shown). (C) **Probe 50** (neuronal pentraxin receptor, TIGR identifier TC18750). The expression domain matches the boundaries of the lateral and basolateral amygdala (arrow). Staining was also observed throughout cortex (arrowhead) and in hippocampus (not shown). No signal was detected in the cerebellum or PAG. The olfactory bulb had weak staining (not shown). (D) **Probe 28** (plasma glutathione peroxidase, TIGR identifier TC31122). Intense labeling is apparent in the medial amygdala (arrow), hypothalamus and PAG (not shown). Note also signal in a contiguous subregion of the basomedial amygdala (dotted line). Two other genes also showed expression in this same region (not shown). (E). **Probe 68** (CSF-induced cysteine protease, TIGR identifier TC 30215). Hybridization in the basomedial amygdala (arrow) was detectable. Staining was also observed in the hippocampus, but was absent in the remaining regions of study (not shown). (F) **Probe 20** (Ccte chaperonin epsilon subunit, TIGR identifier TC30886). Signal was detected in the medial amygdala (arrow) and in the lateral, basolateral and basomedial complexes (not shown). No staining was detected in the other 4 regions of study (not shown).

Figure 1-4: Possible gene expression patterns in the amygdala.

Possible gene expression patterns in the amygdala and the percentage of amygdala-enriched genes examined that exhibited such patterns. (i) Contiguous, pan-amygdaloid expression. (ii) Contiguous expression in subdomains whose boundaries bear

no relationship to those of classically defined amygdaloid subnuclei. (iii) Expression in scattered cells contained within specific subnuclei. (iv) Expression in scattered cells not respecting subnuclear boundaries. (v) Contiguous expression in subdomains whose boundaries match, at least in part, those of amygdaloid subnuclei. The majority of genes examined (75%) exhibited pattern (v).

A


Distribution of average difference values
amygdala
cerebellum
hippocampus
olfactory bulb
PAG
Proportion per bin
Fluorescence intensity (arbitrary units)

B

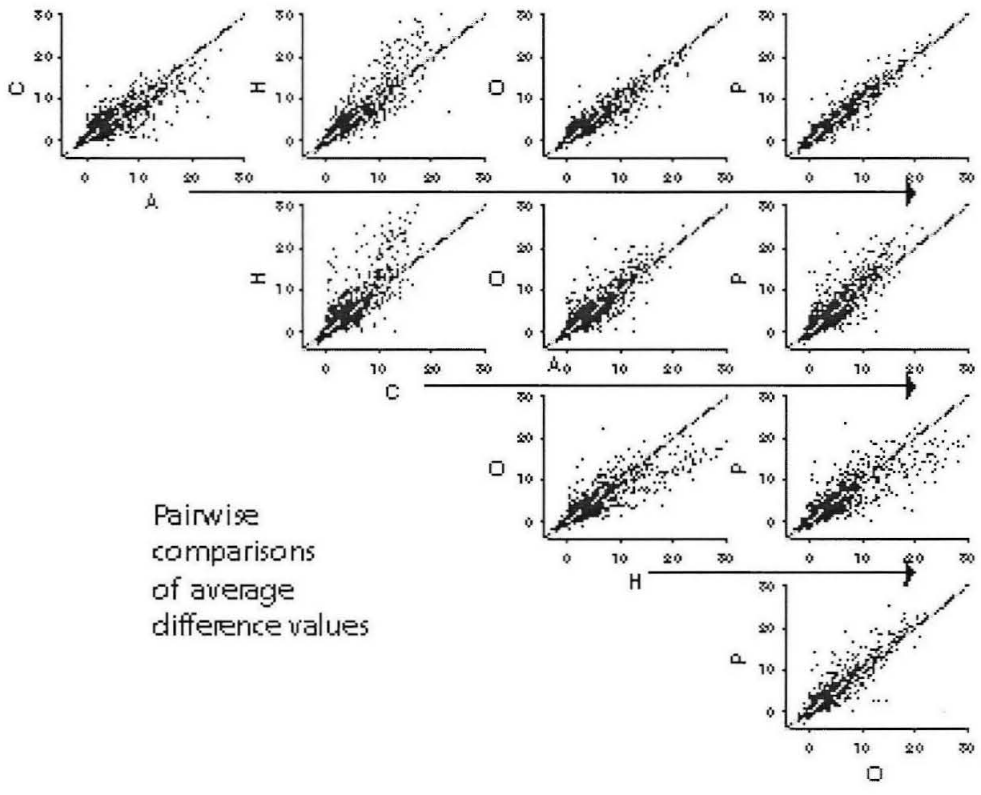

Pairwise comparisons of average difference values

Fig. 1

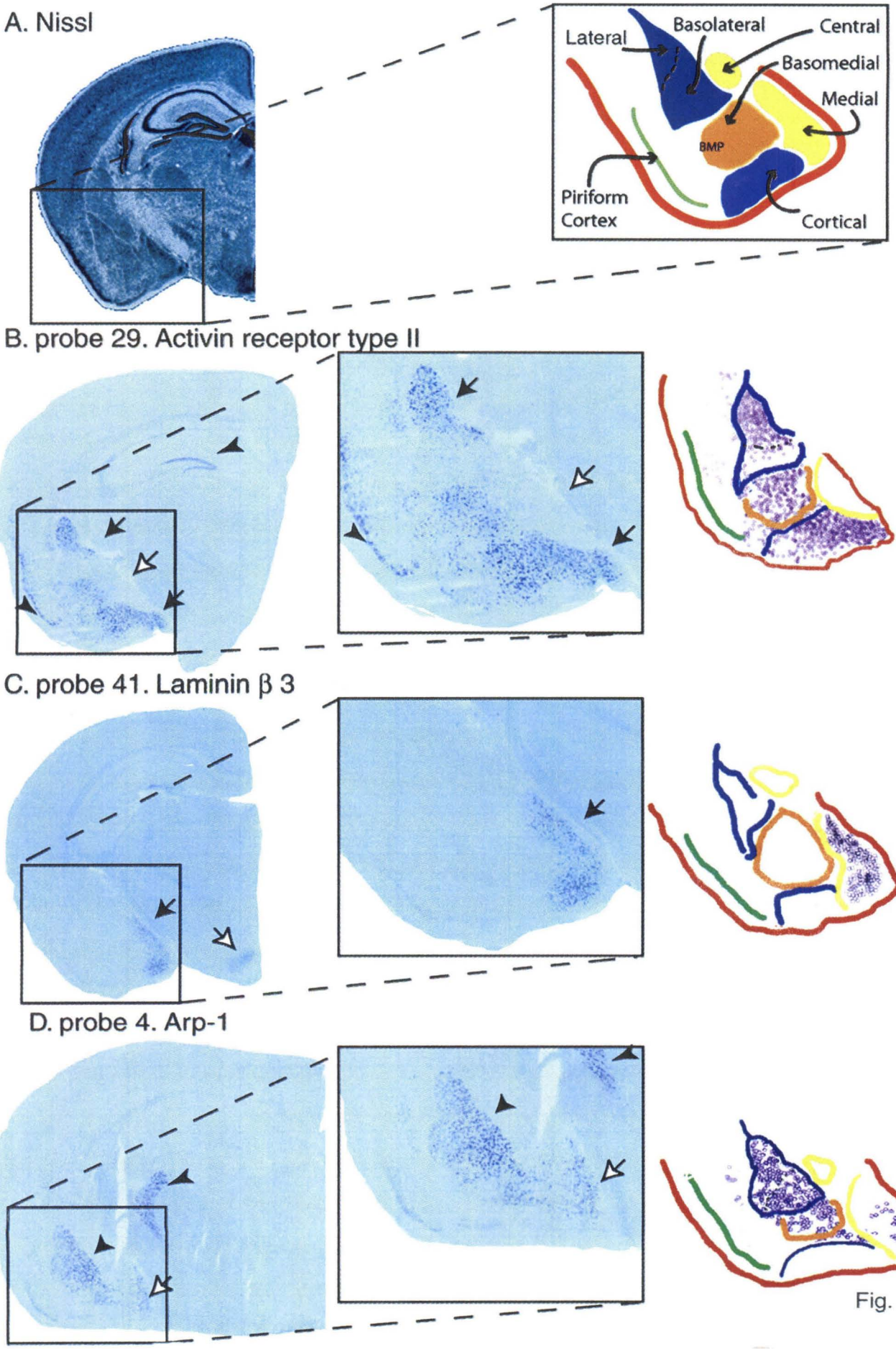
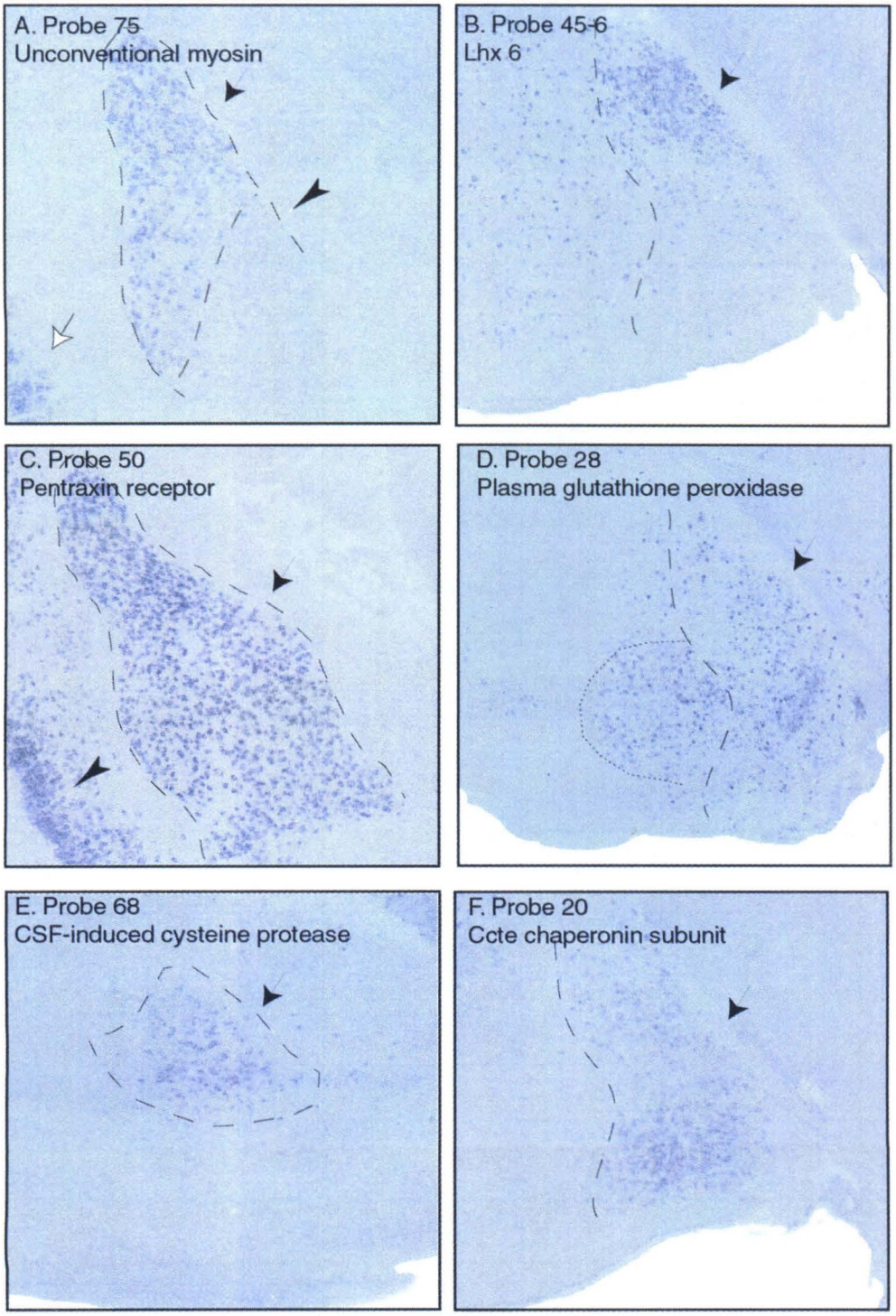
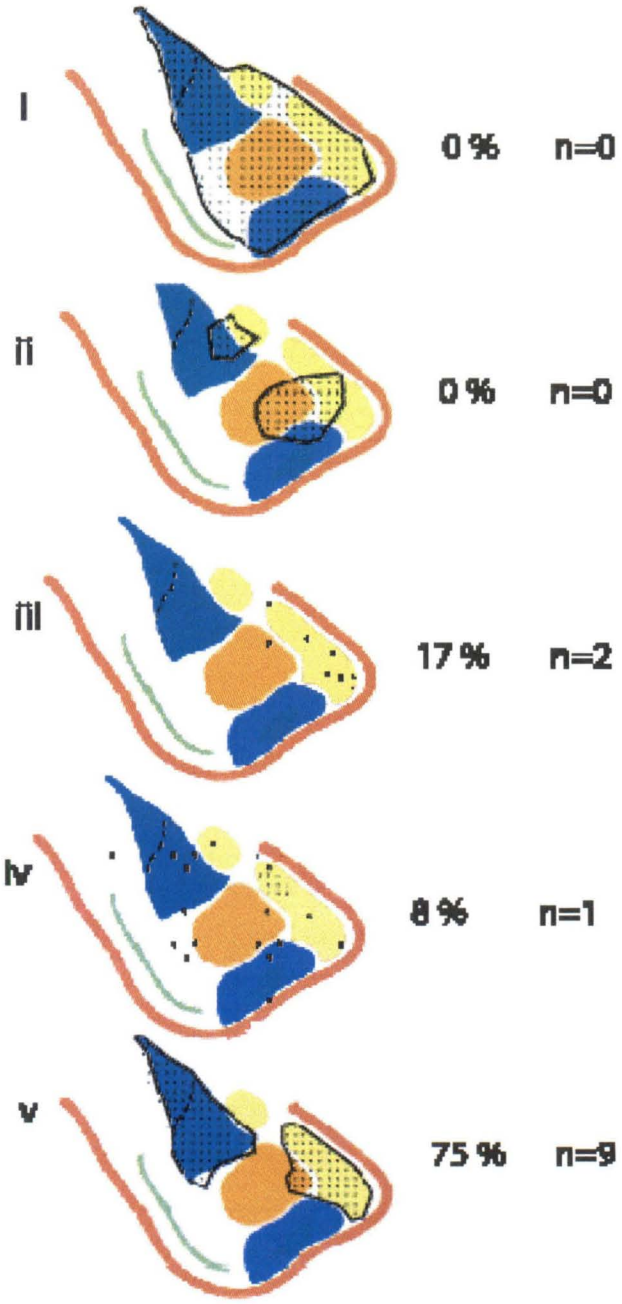

A. Nissl
Lateral
Basolateral
Central
Basomedial
Medial
BMP
Piriform
Cortex
Cortical
B. probe 29. Activin receptor type II
C. probe 41. Laminin β 3
D. probe 4. Arp-1

Fig. 2



Chapter 1- Fig. 3



Chapter 1- Fig. 4

2 LASER-CAPTURE MICRODISSECTION COMBINED WITH MICROARRAY TECHNOLOGY PERMIT THE IDENTIFICATION OF GENES DIFFERENTIALLY EXPRESSED WITHIN DISTINCT AMYGDALA SUBNUCLEI

2.1 Introduction

The previous chapter described the successful characterization of amygdala-enriched gene products using microarray technology followed by *in situ* hybridization. We found that only 91 genes were differentially enriched in each brain region. Considering that 19,022 genes were present in the brain regions examined based on microarray readings, only 0.5% of the expressed genes were thus differentially expressed in each region (91/19,022). It is noteworthy that very similar figures were obtained by an independent study (Sandberg et al., 2000). This relatively high homogeneity among brain regions was somewhat unexpected for two reasons. First, neurons are very heterogeneous, with potentially hundreds of different neuronal cell types in the cerebral cortex alone (Serafini, 1999). Second, early studies done prior to the advent of large-scale microarray technology had suggested that a large fraction of the genome was expressed in the brain (Milner and Sutcliffe, 1983). Thus why have we found relatively few differentially enriched transcripts? Two main reasons come to mind. On the one hand, the use of high-density oligonucleotide arrays by definition limits the screen to genes and

ESTs of previously known sequence. Therefore, if indeed highly specific brain transcripts exist, they may not have been cloned yet and thus represented on the arrays. On the other hand, highly specific, non-abundant transcripts may be lost due to a "dilution problem." Because of the cellular diversity in the brain, important expression differences occurring in a subpopulation of cells that make up a small fraction of the total population may simply be diluted out and undetected. We decided that even though the cost of the commercial arrays was high, their convenience and proved reliability merited their use. Thus, to try to solve the dilution problem we decided to start with more homogeneous tissue, obtained by laser-capture microdissection and search for genes differentially expressed within amygdala subnuclei. Laser capture microdissection (LCM) has been developed to extract pure cell populations from specific regions of tissue sections, under direct microscopic visualization (Simone et al., 1998). A transfer film is applied on the surface of the tissue section placed on a standard glass slide, and is activated by a low-power laser beam. When activated, the film focally adheres to the cells of interest and thus permits their collection, while surrounding tissue that has not been submitted to the laser shot is left intact. In principle, the low energy laser used does not alter cellular contents, which can then be reliably collected for nucleic acid or protein extraction (Emmert-Buck et al., 1996).

The use of finer dissections introduced a different problem, though: RNA had to be amplified prior to probe synthesis given the low number of cells isolated by LCM. I relied on linear, T_7 RNA polymerase-based amplification methods, which had been used in the past (reviewed in Kacharmina et al., 1999), and also in combination with LCM-extracted dorsal root ganglion RNA (Luo et al., 1999). Consequently, I compared gene

expression profiles among the central, lateral, and medial amygdala subnuclei and contrasted those with the profiles obtained in the previous screen, where a rather crude hand dissection of the whole amygdala was performed. The results presented in this chapter show that such a combination of technologies, namely LCM-extraction of RNA from distinct brain subnuclei, followed by linear RNA amplification and microarray analysis provides a useful tool to further characterize the molecular constituents of small neuronal subpopulations.

2.2 Brief overview of anatomy and function of different amygdala subnuclei

As mentioned in Chapter 1, the amygdala is thought to control emotional behaviors, such as fear, anxiety, and emotional learning (reviewed in Rogan and LeDoux, 1996). A variety of lesion studies done in rodents or monkeys indicate that the amygdala is necessary for the evaluation of fearful stimuli. For example, rats with damage to the amygdala will approach a natural predator, such as an anesthetized cat (Blanchard and Blanchard, 1972). What is now called the Kluver-Bucy syndrome was recognized a long time ago in monkeys with lesions in the amygdala (Kluver and Bucy, 1939). These monkeys were generally placid and performed tasks normally perceived as threatening. In addition, human patients with damage in the amygdala fail to recognize fearful visual or auditory stimuli (Adolphs et al., 1994; Scott et al., 1997), and also show general hypoemotionality. Moreover, electrical stimulation of the amygdala in rodents can elicit typical autonomic manifestations of fear and cessation of ongoing activities (freezing), a stereotypical response often measured in behavioral paradigms (reviewed in Clark, 1995).

Since it is viewed that some psychiatric conditions such as generalized anxiety or post-traumatic stress disorders are maladaptations of normal fear responses (Rosen and Schulkin, 1998), the role of the amygdala in fear conditioning has attracted special attention (reviewed in Flint, 1997; LeDoux, 1995). Briefly, this associative learning phenomenon consists of the generation of fear by a harmless stimulus. It is achieved by the repetitive co-presentation of a neutral stimulus (such as a tone) and an aversive stimulus (e.g. electric shock). After a number of trials, animals (including humans) learn to associate both stimuli and show aversive responses to the single presentation of the previously neutral stimulus.

Anatomically, the amygdala is a complex forebrain structure composed of over a dozen subnuclei (Pitkänen et al., 1997), such as the central, lateral, basomedial, and medial subnuclei.

Ontogenetically, the amygdala has mixed embryological origins (Puelles et al., 1999), as evidenced by different homoeobox gene expression domains in the developing prospective amygdala and by different neurotransmitter expression in the adult (Esclapez et al., 1993). Accordingly, its various subnuclei have different functions. The lateral nucleus is the site of sensory input convergence (Maren and Fanselow, 1996; Pitkänen et al., 1997). It presumably integrates different inputs before relaying information to other subnuclei. Just as the lateral nucleus is the main input system of the amygdala, the central nucleus is the output. In general, there is a unidirectional informational flow, from lateral structures (such as the lateral and basolateral nuclei), to more medial ones (including the medial, basomedial, and central nuclei) (Pitkänen et al., 1997; Swanson and Petrovich, 1998). Interestingly, there are heavy internuclear projections from the lateral nucleus to

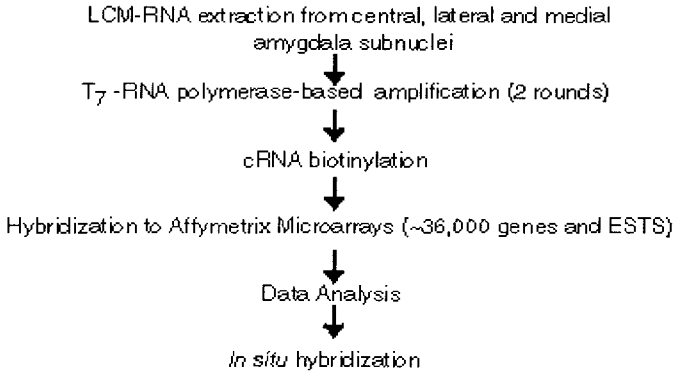
the basolateral, basomedial and medial nuclei, whereas there are fewer reciprocal connections from these subnuclei to the lateral one. Most subnuclei in turn heavily project to the central nucleus, which has sparse axonal projections to any other amygdaloid nucleus (Pitkänen et al., 1997). The central nucleus projects to hypothalamic and brainstem regions involved in the reactions to fear (Davis, 1992), that produce perspiration, increased blood pressure, and bradycardia. The posterior cortical and medial amygdaloid nuclei are interesting because they receive input from the accessory olfactory bulb (Swanson and Petrovich, 1998). The medial nucleus heavily projects to the medial hypothalamus, in regions known to mediate specific feeding, reproductive and defensive behaviors (Canteras et al., 1994). In turn, the medial subnucleus receives input from ventromedial hypothalamic neurons. Thus, this subnucleus integrates pheromonal information and may indirectly regulate fundamental innate behaviors. In addition, the medial amygdala is sexually dimorphic, as the volume is about 20% larger in male than in female rats. Sex differences in synaptic organization in this nucleus have also been reported (Roberts, 1992). The accessory olfactory bulb, which directly projects to the medial nucleus, is also larger in males, thus sex differences in medial amygdala volume may be partly due to unequal synaptic input to this region. Moreover, the posterior medial amygdala harbor many estrogen and androgen receptor-containing neurons, thus its function is probably modulated by gonadal steroids (Roberts, 1992).

Because their roles are better defined than other amygdala nuclei and their anatomical demarcations are more clearly visualized with Nissl staining, I decided to conduct RNA-expression studies from the lateral, central and medial subnuclei.

2.3 Methods

2.3.1 Experimental design

We compared gene expression among three amygdala subnuclei: central, lateral and medial. The overall strategy was similar to the one described previously, except that the tissue was dissected by laser-capture (LCM):



Briefly, subnuclei were dissected by LCM, RNA was amplified prior to probe synthesis, and biotinylated cRNA probe was hybridized to Affymetrix U74v2 microarrays, comprising about 36,000 mouse genes and EST clusters

(http://www.affymetrix.com/products/murineU74_content.html). The Murine U74v2 sets are composed of three subarrays (U74v2A, U74v2B, U74v2C). Three biotinylated probe replicates were synthesized for each subnucleus, by independent cDNA synthesis from a single RNA LCM-extraction.

2.3.2 Sample preparation

The following protocol involving RNA amplification prior to probe synthesis was adapted from (Luo et al., 1999; Mahadevappa and Warrington, 1999; Ohyama et al., 2000). Please refer to the Affymetrix manual for a detailed protocol for standard probe synthesis.

Histology: Brains from two C57/Bl6 mice sacrificed by cervical dislocation were used. 15-um thick sections were sliced with a cryostat and stored at –80°C immediately in a dry container. The day of the microdissection, sections were first fixed in 100% Ethanol for 3min and then hydrated through a series of 95%, 75%, and 50% ethanol solutions for 20 sec each. They were rinsed in water for 20 sec and stained with Nissl (0.5% Cresyl violet + 0.1 M Sodium acetate buffer, made with Molecular Biology Grade Biowhittaker water) for 40 sec. Consequently, slides were briefly rinsed in water, and then dehydrated in the graded ethanol solutions series above mentioned. Finally, sections were immersed in 2 washes of xylenes, for 20 sec and 1 minute, respectively. Slides were allowed to air-dry for 5-10 min, and then were placed in a dry container with Drierite desiccant.

Laser-capture Microdissection (LCM): An Arcturus LCM instrument was used for dissection. A large beam (of approximate 80 mW in Power and a duration of 8.5ms) was directed towards the subnuclei of interest and cellular contents were collected on large caps (Arcturus Cap Sure TF-100). Typically, around 2000 pulses were shot for each piece of tissue targeted with such laser beam.

RNA extraction: Solutions from the Qiagen Rneasy total RNA isolation kit were used (cat. 74104). About 6-8 "pieces of subnuclei" from 15 μm thick sections were

collected on the same cap. 40 µl of RLT buffer were added to each cap inverted on an eppendorf tube. Contents were spun down at 4000 rpm for 30 sec. in a microcentrifuge. The tubes were left at –80°C until the microdissection was finished. Finally, all cellular contents from the same subnucleus were pooled in a final volume of 500 µl of RLT buffer. About 30 "pieces of microdissected subnuclei" were used for each probe preparation for microarray hybridization.

RNA purification and DNase treatment: Instructions from the Qiagen Rneasy total RNA isolation kit were followed. A DNase treatment step was included prior to elution from the columns (Qiagen cat 79254). RNA was eluted in 50µl water and then concentrated using Microcon-100 columns (Fisher cat. 42413) into 10µl (Spun at 3000 rpm for 12 min in the microcentrifuge).

First cDNA synthesis and quality control: 1µl of 20µM T7-$(dT)_{24}$ primer (Genset Corp.) was mixed with 10 µl RNA and heated 10 min at 70°C and 5 min at 42°C.

Then the following reagents from the Superscript Choice System kit (Gibco 18090-019) were added: 1µl RNAse inhibitor, 4 µl transcription 5X buffer, 2 µl DTT, 1 µl of 10 mM dNTP, µl Superscript II RT. The reaction was incubated for 1 hr at 42°C. PCR controls were done with specific housekeeping gene primers such as actin, GAPdH, hprt and alpha-tubulin with 0.5 µl of cDNA, using 40 cycles. Only samples that gave good PCR products were used for subsequent syntheses.

Second strand synthesis: The following was mixed with the first strand cDNA product: 96 µl RNase-free water, 30 µl of 5X second strand buffer, 3 µl of 10mM dNTP, 4 µl DNA polymerase I, 1 µl RNAse H, 1 µl DNA ligase, and incubated at 16°C for 2 hr.

2 µl of T4 DNA polymerase were added for 10 min at 16°C. The reaction was stopped with 10 µl of 0.5M RNAse-free EDTA. DNA was then extracted with 150 µl Phenol: Chloroform: Isoamylacetate (25:24:1) using PLG light columns (Fisher). The DNA was washed 3 times with a final volume of 500 µl and concentrated using the same Microcon-100 columns (Spun at 3000 rpm for 12 min in the microcentrifuge).

T7-polymerase RNA amplification (first round): 12 µl of ds cDNA were used for RNA amplification. They were mixed with the following from the Ampliscribe T7 transcription kit (Epicentre Technologies AS2607): 2 µl of 10X transcription buffer, 0.5 µl each of 100 mM of NTP, 2 µl DTT, 2 µl T7 polymerase. Incubation was carried out at 37°C for 14 hr. 180 µl of water was added and RNA was extracted with Phenol: Chloroform: Isoamylacetate (25:24:1). RNA was washed 3 times and concentrated with Microcon-100 columns as described above and collected in 9 µl of water.

Second cDNA synthesis: for first strand, in a PCR tube 9 µl RNA and 2µl random hexamers (50 ng/µl) were mixed, heated 10 min at 70°C and then cooled at 4°C for 5 min in a PCR machine. The following reagents were added: 1 µl RNAse-inhibitor, 4 µl 5X transcription buffer, 2µl DTT, 1 µl (10 mM) dNTP, 1µl Superscript II RT and incubation was carried out at 37°C for 1 hr.

For second strand synthesis, 1 µl RNAse H was first added for 20 min and then the reaction was stopped by heating at 95°C for 2 min in a PCR machine. After addition of 1 µl of 60µM T7-$(dT)_{24}$, the mix was first heated at 70°C for 5 min and then at 42°C for 10 min in a PCR machine. The following were added: 90 µl RNase-free water, 30 µl 5X second strand buffer, 3 µl (10 mM) dNTP, 4 µl DNA polymerase I, and 1 µl RNAse H and incubated at 16°C for 2 hr. 2 µl T4 DNA polymerase were added and incubated 10

min at 16°C. The reaction was stopped with 10 µl of 0.5 M depc-treated EDTA. DNA was then extracted with 150 µl Phenol: Chloroform: Isoamylacetate (25:24:1) using PLG light columns (Fisher). The DNA was washed 3 times with a final volume of 500 µl and concentrated using the same Microcon-100 columns (Spun at 3000 rpm for 12 min in the microcentrifuge). The DNA was collected in 9 µl.

T7-polymerase RNA amplification (second round): 9 µl of ds cDNA were used for RNA amplification. They were mixed with the following from the Ampliscribe T7 transcription kit (Epicentre Technologies AS2607): 2 µl of 10X transcription buffer, 1.3 µl each of 100 mM of NTP, 2 µl DTT, 2 µl T7 polymerase. Incubation was carried out at 37°C for 14 hr. 180 µl of water was added and RNA was extracted with Phenol: Chloroform: Isoamylacetate (25:24:1). RNA was washed 3 times and concentrated with Microcon-100 columns as described above and collected in 9 µl of water.

cDNA synthesis (third round): This was carried out exactly as in the second round. After extraction, DNA was washed and collected into 22 µl water.

Biotinylation (third round of RNA synthesis): The following from the ENZO RNA transcript labeling kit (cat 42655) were mixed: 4 µl vial 1 (HY reaction buffer), 4 µl vial 2 (biotin NTPs), 4 µl vial 3 (DTT), 4 µl vial 4 (RNAse inhibitor), and 2 µl vial 5 (T7 polymerase). Incubation was at 37°C for 4–5 hr. (mixing contents every 30-45 min)

RNA clean-up: The Qiagen Rneasy total RNA isolation kit was used (cat. 74104). First, 1.5 µl of RNA was saved for running a gel. 62 µl of RNAse-free water and 350 µl of RLT buffer were added and mixed. 250 µl ethanol were added and the solution was applied to the affinity column. The elution was applied to the same affinity column again, followed by two washes with RPE buffer. RNA was eluted twice in 50 µl of

RNAse-free water each, (so that the final volume was 100µl). For RNA precipitation, 50 µl of 7.5 M NH4Ac, 250 µl cold EtOH and 1 µl glycogen were added, incubated at –20°C 1 hr, spun 30 min at 4°C, and washed twice with 500 µl of 80% cold Ethanol. Clean biotinylated RNA was resuspended in 15 µl RNAse-free water. Absorbance at 260/ 280 nm was checked with 0.5 µl. An electrophoretic gel was also run with 0.5 µl RNA to check that a good smear of RNA synthesis was obtained. (RNA was stored at -80°C at this point.)

RNA Fragmentation: (For this, the initial concentration should be > 0.6 µg/µl and the final concentration in the fragmentation reaction should be > 0.5 µg/µl.) 3.5 µl 5X fragmentation buffer (200 mM Tris-acetate, pH 8.1, 500 mM KOAc, 150 mM MgOAc) were added and incubated at 94°C for 35 min.

Finally, 0.5 µl were run on a gel to confirm that fragmentation of RNA occurred. (Fragmented RNA was stored at –20°C).

2.3.3 Data analysis

Custom Software: Data were first normalized by the Affymetrix software, so that the average intensity of each array was equal to an arbitrary target intensity. In this case, the target intensity was 2,500, a value typically set by the HHMI facility at Stanford, where these arrays have been scanned. (The average intensity of an array is calculated by averaging all the average difference values of every probe set on the array, excluding the highest 2% and lowest 2% of the values.) Normalized average difference values were exported and analyzed with custom software, written in Matlab (The MathWorks, Natick, MA), which is available at http://www.its.caltech.edu/~mariela/gene_screen.html. The

software was based on the one described in Chapter 1, except that it also contemplated the existence of biotinylated cRNA probe triplicates for each subnucleus, called C1, C2, C3 (for central nucleus replicates 1, 2, and 3); L1, L2, and L3 (for lateral nucleus); M1, M2, and M3 (for medial nucleus).

Three criteria were applied to identify genes enriched in each replicate of the three amygdala subnuclei: (1) the average difference ($\bar{\Delta}$) value for the gene in that subregion; (2) Significance of labeled cRNA hybridization to the probe cells on the microarrays; and (3) the ratio (fold-difference) of its $\bar{\Delta}$ value in the reference subregion to that in each of the other six subregions (three replicates for each of the other two subnuclei being compared to). A given gene g_i, with an average difference value in central nucleus-replicate 1 $\bar{\Delta}_{g_i}^{C1}$, was considered to be enriched in this replicate relative to the other subnuclei examined if it satisfied the following constraints for these three criteria:

i) $\bar{\Delta}_{g_i}^{C1} > minimum$

ii) pos – neg > 6 if pairs used>10 or

pos – neg > 3 if pairs used<10

iii) $\bar{\Delta}_{g_i}^{C1} / \bar{\Delta}_{g_i}^{other} > 2.5$ or $\bar{\Delta}_{g_i}^{C1} / \bar{\Delta}_{g_i}^{other} < 0$ for all six other samples (L1, L2, L3, M1, M2 and M3).

In this case, the *minimum* value was set = 200 (for reference, the arrays were normalized to a target intensity of 2500 and the noise levels were between 20 and 30).

Constraint ii) is a good indication of whether the hybridization of the synthesized cRNA probe to the arrays is significant or rather due to nonspecific binding, and is analogous to the *Present* call in the Affymetrix software[2].

Visual Inspection: The hybridization parameters to the microarrays of each enriched gene identified with the custom software were visually inspected to further narrow down the list of candidate genes for *in situ* hybridization follow-up experiments. Genes that had average difference values that varied considerably across replicates from the same subnucleus were not considered good candidates. Similarly, I did not expect to observe reliable differences in gene expression by *in situ* hybridization for genes with relatively high average difference values in all replicates whose difference (pos-neg) was about the same for all subnuclei.

2.3.4 *In situ* hybridization

The protocol for the *In Situ* hybridization was described in Chapter 1.

2.4 Results

2.4.1 Analytical characterization of differentially expressed genes

Before analyzing the data in detail, I wanted to assess whether the quality of cRNA probe preparation was good in spite of the extra manipulation (i.e., LCM

[2] When the intensity of the perfect match probe cell is significantly greater than that of the corresponding mismatch probe cell, the probe pair is termed positive (pos). When the intensity of the mismatch probe cell is significantly greater than that of the corresponding perfect match probe cell, the probe pair is termed negative (neg). For a more detailed definition of the pos and neg parameters, see the Affymetrix user manual. These two parameters are evaluated by the Affymetrix software for *Present* or *Absent* calls. In this software, 3 conditions are required for a *Present* call: a large (pos/ neg) ratio, a high proportion of (pos/ pairs used) and a high ratio of hybridization between perfect match probe cells / mismatch probe cells . Since the first 2 conditions are not independent, I decided to express the constraint regarding the pos and neg parameters in a single clause in ii). Condition i) in my software deals with hybridization intensity.

extraction, RNA amplification) compared to the previous screen, where tissue was dissected intact and amplification was not required. I first calculated the distribution of the average difference values for all genes from each replicate (Fig. 1). The histograms resembled those obtained with probe prepared from intact tissue, like the ones plotted in Fig. 1A, Chapter 1. In addition, I checked whether the genes that had been identified as amygdala-enriched in the previous screen had consistent average difference values in the current study. Of the 33 genes previously identified, about 20 were represented on the current microarrays [3]. Of these, 10 that had high average difference values also had high average difference values in the current screen. Notably, the differential expression within the amygdala of some genes previously characterized by *in situ* hybridization was apparent in the new dataset. To name a few, Lhx-7 had mean average difference values of 1261, 63, and 174 in the central, lateral, and medial amygdala, respectively, and was only detected in the central nucleus by *in situ* hybridization (not shown). Similarly, Arp-1 had average difference values of 1164 in the central, 2700 in the lateral and 3094 in the medial subnuclei and was expressed in the lateral and medial nuclei (Chapter 1, Fig. 2D).

These preliminary analyses indicated that cRNA probe preparation was not awfully biased and that at least part of the data reproduced the findings of the previous screen. I then analytically estimated the extent of differentially expressed genes within the amygdala subnuclei revealed by this screen.

The number of differentially enriched genes that fulfilled the selection criteria described in the Data Analysis section is presented in Table 1. There were 358 genes

[3] The previous screen was done with older microarray sets (Mu 11K and Mu 19K) no longer available from Affymetrix. For the screen described in this chapter, the new generation of arrays, also from Affymetrix (U74v2), was used. The latest set includes a more comprehensive coverage of the mouse genome. However, there are several probe sets (from the Mu19K subarrays, corresponding to TIGR clusters sequences), that are not represented in the U74v2 arrays.

differentially enriched. This constitutes about 1% of the ~36,000 genes and ESTs examined by the microarrays. On the other hand, I was curious to know the percentage of genes that were expressed in all subnuclei. The number and percentage of genes that were deemed *Present* by the Affymetrix software and had average difference values larger than 200 in all triplicates of all subnuclei is shown in Table 2, for the U74v2A subarray. For comparison, the numbers of *Present* genes with average difference >200 in each subnucleus are also shown. Roughly, 32% of interrogated genes were expressed in each subnucleus, and 25% of genes were detected in all of them.

To narrow down the list of candidate genes with the highest relative differences of expression within the amygdala, a visual inspection of various parameters was carried out (see Data Analysis section). For example, 104 different genes of the U74v2A array were differentially expressed (Table 1, the genes appearing in more than one replicate were counted only once). However, only 15 genes remained in consideration after visual inspection. Of those, 5 were further discarded because they were expressed in many other tissues or brain regions based on other microarray experiments[4]. The remaining top 7 candidates were then selected for *in situ* hybridization. 5 of these gave *in situ* hybridization patterns consistent with the microarray readings. I was unable to synthesize *in situ* probe for the other 2 candidates.

[4] Three other datasets were utilized for this purpose: the one described in Chapter 1(Zirlinger et al., 2001) of 5 brain regions, the one described in (Sandberg et al., 2000) of 6 brain regions, and another dataset (available at http://expression.gnf.org) including expression data of several brain regions and other organs.

2.4.2 Validation of Genechip results by *in situ* hybridization

Selected candidates were chosen for *in situ* hybridization analysis (see Data Analysis section for selection criteria). Probes for 18 different genes or ESTs (from all three subarrays) were synthesized. Of these, 4 (22%) gave no signal. Of the remaining 14 probes, 11 hybridized in complete agreement with the microarray readings (11/14= 79%). 3 probes gave a hybridization signal with moderate agreement with microarray readings (3/14= 21%). Similarly to the previous study, the rate of false positives may be actually lower, since in most cases the probe for *in situ* was not optimized, or simply, the abundance level of the transcripts may be too low for the detection limit of the nonisotopic method used. The overall rate of agreement between microarray readings and *in situ* hybridization is remarkably close to the one obtained in the previous screen, which was about 60% in both cases.

Figure 1 shows examples of genes enriched in the central subnucleus. Probe 873, neuromedin N/ neurotensin (Fig. 2A) is primarily expressed in the central nucleus (arrow), and in scattered cells in the medial nucleus (arrowhead). The lateral nucleus (white arrow) is devoid of signal. Probe Bh1, neuromedin B precursor (Fig. 2B) is also highly enriched in the central nucleus (arrow), while the lateral (white arrow) and the medial (arrowhead) nuclei show undetectable expression levels. Neuromedin B was also highly expressed in the olfactory bulb. Both neuropeptides neuromedin B precursor and neuromedin N were in addition expressed in the hypothalamus. One EST, probe 153/138, was also enriched in the central nucleus (Fig. 2C, arrow). Examples of genes enriched in the lateral nucleus are shown in Fig. 3. Probe 117, kinesin-like protein is more abundantly expressed in the lateral nucleus (arrow), than in the central or medial nuclei

(Fig. 3A). A similar pattern of expression was observed with probe 126, an EST (Fig. 3B). Note cortical expression of probes 117 and 126 (black arrowheads). Gastrin releasing peptide, probe 440 (Fig. 3C), was also highly enriched in the lateral nucleus (arrow) and in the basomedial nucleus (black arrowhead). Figure 4 depicts genes enriched in the medial subnucleus (arrows), such as prolactin receptor (Fig.4A) and thyrotropin releasing hormone receptor (Fig. 4B). Gastrin releasing peptide, thyrotropin releasing hormone receptor, and prolactin receptor were also expressed in the hypothalamus (not shown). Counterstaining was performed on the same sections with Neurotrace, a fluorescent nuclear staining in order to delineate subnuclei (Panels D, E and F, dashed lines). In addition, Nissl staining was performed on contiguous sections (example shown in Fig. 4C).

2.5 Discussion

2.5.1 Methodological considerations

Doing replicate cRNA biotinylated probes for each sample proved crucial. The overall rate of agreement between *in situ* hybridization and microarray results was 60% (11 genes out of 18 tested), and would have been lower had we omitted replicate analysis. This rate of agreement was the same as the one obtained in the screen described in the earlier chapter where a single sample for each region was used. This suggests that replicate analysis from LCM-extracted material coupled to RNA amplification ensures results comparable to microarray analysis from hand dissected brain regions, at least as validated by *in situ* hybridization. It should be noted once again that the rate of false

negatives might indeed be lower than observed if more sensitive methods of detection, such as radioactive *in situ* hybridization or other biochemical assays, were used. Inspection of Table 1 illustrates another advantage of analyzing replicate samples. Medial region replicate 2 (M2) showed an unusually large number of enriched genes when compared to the other replicates. The reason for this is unclear, but may be due to partial degradation of the cRNA biotinylated probe[5]. However, most genes did not appear to be enriched in the other two medial amygdala replicates, and were thus not seriously considered as good candidates for *in situ* hybridization follow-up experiments. Moreover, the majority (73% or 8/11) of differentially enriched genes validated by *in situ* hybridization were enriched in at least two replicates. The remaining genes were also relatively enriched in all corresponding triplicates of a particular subnucleus but fulfilled the enrichment criteria set in the algorithm in just one[6].

It is interesting to comment on the abundance levels of the enriched genes identified. Among the 115 enriched genes from subarray U74v2A (Table 1), their mean average difference value was 13,471. Excluding the 36 enriched genes from M2, which showed unusually high values, the mean average difference value was 2,431 (n=79). For comparison, the array mean intensity was normalized to 2,500. Thus, enriched genes identified by this screen seemed to be expressed at relatively high levels. This is surprising, since higher complexities in RNA populations are found in low abundance

[5] It is also apparent from Fig. 1 that average difference values in M2 were higher. Note the larger mean average difference value for this sample (dashed line). I suspect of RNA degradation because the Medial region was the last one to be dissected by LCM and because the proportion of *Present* genes was lower in this sample (~37%) compared to the others (~45%). If this is true, the implication is then that shorter cRNA molecules bind with higher affinity to the microarrays.

[6] In the remaining 2 replicates, the fold-difference was not larger than 2.5, the ratio-threshold set in my custom program to identify enriched genes, and were thus not selected by the algorithm in such two cases.

transcripts (Lewin, 2000). Together, this has several implications. First, RNA amplification of LCM-extracted tissue may be only efficient for highly expressed genes, and thus genes expressed at lower levels, which are more likely to be differentially expressed, are lost in the process of LCM extraction or amplification. Alternatively, lower abundance genes may have been successfully isolated by LCM and amplified, but might have shown variable hybridization to the microarrays, and thus non-reproducible readings in different replicates of the same region. Finally, as already stated in the Introduction, many yet unidentified highly specific genes may not be represented on the microarrays.

It is then valid to ask, is it worth the effort to start with LCM-dissected tissue? I think the answer is yes, because it partially overcomes the dilution problem referred to in the Introduction. I shall exemplify this claim with the behavior of genes characterized by *in situ* hybridization in the LCM-screen, which also showed higher levels of expression in the amygdala compared to the other four brain regions previously studied (not shown). Among those 11 enriched genes, 7 were represented on the older generation of arrays used for the previous screen. In all cases, their fold difference in amygdala expression relative to the other four brain regions was too low. (The mean ratio of average difference values in the amygdala relative to the other four regions was about 1.6 in these 7 cases.) Moreover, 4 of those 7 genes were considered to be *Absent* in the amygdala by the previous microarray readings. To summarize, most amygdaloid subnucleus-enriched genes currently identified were not recognized as amygdala-enriched in the previous screen because their levels of expression appeared too low, most likely due to expression averaging in the large piece of dissected amygdaloid tissue.

In agreement with our previous study (Zirlinger et al., 2001), these results also suggest that *in situ* hybridization is essential to confirm microarray data. About 80% of genes that showed *in situ* hybridization signal were consistent with microarray readings; the rest showed moderate agreement. In addition, *in situ* hybridization also revealed other sites of expression, such as the hypothalamus or cortex, not included in the microarray samples. The high spatial resolution of *in situ* hybridization allowed for the direct observation of gene expression in specific subnuclei. The genes identified with this screen were expressed in contiguous cell populations and most showed boundaries of expression contained in defined amygdaloid subnuclei. This validates the hypothesis proposed in our earlier study (Zirlinger et al., 2001) that most amygdala-enriched genes may respect subnuclear boundaries of expression.

2.5.2 When is a differentially expressed gene a good candidate for *in situ* hybridization?

Some conclusions can be drawn from inspection of the various parameters provided by the Affymetrix software. Again, I will illustrate this point exemplifying with the genes validated by *in situ* hybridization in both screens. The parameters most useful for this analysis are the *number of positives and pairs used, the absolute call (Present / Absent) and the differential call (Increase / Decrease)*. Briefly, the first two parameters indicate the extent of significant hybridization of the cRNA labeled probe to the probe cells on the array and indirectly, whether significant hybridization was observed throughout the whole length of the gene. The differential call indicates whether there was a difference in hybridization across the different samples examined. It is scored based on

the comparison of cRNA hybridization to each probe cell (of the 16 or 20) representing a gene on the array across different samples[7].

Table 3 details the parameters observed for a set of 21 differentially enriched genes whose expression patterns were confirmed by *in situ* hybridization. Among the 13 genes enriched in the amygdala relative to four other brain regions, the ratio of positive probe cells/ pairs used was 60%. This initially suggested that in order for a gene to be detected by *in situ* hybridization, the number of positive cells should be high. I have tried to incorporate this lesson from my first screen, in the criteria to search for differentially enriched genes in my latest screen (see condition ii) in custom algorithm.) Among the 8 genes identified in this screen, such ratio was 66%. Similarly, in 92% of the cases of the first screen the Affymetrix software indicated an absolute call of *Present,* while all genes identified in the second screen were also deemed *Present* by that software. Again, condition ii) in my latest custom algorithm incorporated this idea. (See Data Analysis section for a discussion on this.)

Finally, differential calls in principle provide an indication of whether genes are differentially expressed. However, I have found in my limited examples that these are not always necessarily true. In other words, I have identified genes (~ 40% of the ones analyzed) indeed differentially expressed as detected by *in situ* hybridization that, nevertheless, showed a call of *No change* in expression across different brain regions (Table 3). Conversely, I have observed no change in expression by *in situ* hybridization in genes that had a differential call of *increase* or *decrease* (not shown). Thus, if a candidate differentially expressed gene has different average difference values in distinct

[7] The reader is referred to the Affymetrix user guide for a detailed description of these parameters.

brain regions, I would not automatically consider it a bad candidate if a differential call of *No change* is attributed by the Affymetrix software.

I have also observed that for candidate genes with relatively high average difference values in all regions, it is useful to examine the *pos* and *neg* values closely. In these cases, good candidates are the genes that have higher *pos* (and lower *neg*) values in the region of interest compared to the other regions. Otherwise, if a gene shows high average difference values and similar *pos* and *neg* values in all regions, it is unlikely that relative differences in expression levels can be appreciated by *in situ* hybridization.

I have considered these notions when visually inspecting the genes identified with the custom program to further narrow down the number of candidate genes for *in situ* hybridization validation experiments.

2.5.3 Identity of selected genes differentially expressed in the amygdala

An exhaustive review of all identified genes enriched in the amygdala would constitute a list of disconnected information, so I shall concentrate on few genes that may directly participate in the modulation of amygdaloid function. Of course this is a subjective list, and a structural gene product may be as crucial to amygdala function as a signaling molecule. However, it is simpler today to conceptualize that differential expression of signaling molecules, such as neuropeptides, better explain underlying differences in brain function. I will also attempt to summarize reported sites of expression of these genes both in the brain and the body. For this, I relied on classical literature searches and on expression data based on microarray studies performed on different brain

regions[8] and organs[9]. There are some difficulties in the interpretation of these several reports. First, in classical publications, there is not always a general consensus on sites of expression of particular genes. The discrepancies may be due to different detection methods employed, or may be simply due to partial reports of sites of expression in published literature. In the case of microarray databases, it is hard to extrapolate average difference values to "real" expression, for the numerous reasons discussed before[10]. Nevertheless, I summarize below the expression data and relevant information of some genes identified by this screen.

Since the centromedial aspect of the amygdala is the second neuropeptide-rich region in the brain (following the hypothalamus) (Roberts et al., 1982), I will focus on what is known about some neuropeptides and their receptors.

Neuromedin N and its receptors: Neuromedin N and its analog Neurotensin are synthesized from a common precursor. They were first purified from bovine hypothalamus and intestines (Vincent et al., 1999). Central administration of neurotensin

[8] A database containing several brain regions, including neocortex, olfactory bulb, hypothalamus, hippocampus, amygdala, cerebellum, midbrain and striatum, developed by Carrolee Barlow and colleagues at the Salk Institute.

[9] A body map including most organs and brain regions available at http://expression.gnf.org

[10] Ideally, one would like to predict expression patterns based on Affymetrix parameters from microarray readings. It is hard, however to do so, and the number of examples I have analyzed thus far is probably not high enough to draw general conclusions. However, in order to predict "real" expression, I would implement the same principles that I have described in the previous section to choose candidate genes for *in situ* hybridization validation. In a few words, expression may be only detected if the number of positive cells is high, particularly if a gene has low average difference values. In addition, differences in expression for a gene with relatively high average difference values everywhere may be only detected if the difference of [pos-neg] is significantly different across different samples. Finally, if a gene has a considerably lower average difference value in one tissue compared to other tissues, its expression may not be detected in the former by *in situ* hybridization, even if it is deemed *Present*, particularly if the number of positive cells is lower than in other tissues. Specifically, I have found that if a gene has the "potential" to have several positive cells in some tissues, its expression may be below the detection limit of *in situ* hybridization (or may not be really present) in a sample having considerably fewer number of positive cells.

induces hypothermia, analgesia and increased dopamine release and turnover (Mazella et al., 1996). In the periphery it produces contraction and relaxation in distal colon and proximal ileum, respectively, and can act as a growth factor in many cell types. The central effects can be induced by both intracerebral administration or by direct peptide injection into the central amygdaloid nucleus (Roberts, 1992). It is expressed by at least two different cell types in the central nucleus (Roberts, 1992) suggesting that there may be further differences among neurotensin-containing cells. There are 3 different types of neurotensin receptors reported to date: two are G-protein coupled receptors (GPCRs), whereas the third one is a single-pass membrane receptor identical to gp95/sortiling (Vincent et al., 1999). The function of this latter receptor is not entirely clear yet, but may be involved in intracellular sorting, or in peptide clearance. The GPCRs have different tissue distribution. NTS1, the high affinity receptor is localized in intestine and in many brain regions, such as the diagonal band of Broca, septum, hypothalamus, substantia nigra, and ventral tegmental area (Vincent et al., 1999). Others also report expression of NTS1 in hippocampus, amygdala and cortex, and more moderately in the thalamus (Pettibone et al., 2002). The low affinity receptor, nts2, is expressed at high levels in cerebellum, hippocampus, piriform cortex and neocortex (Mazella et al., 1996). Interestingly, the two receptors appear to be differentially regulated developmentally as well: NTS1 expression is present from birth, peaks around 7-10 days postnatally, and decreases to adult levels between 30-40 days. Conversely, nts2 expression starts after 2 weeks and reaches maximum levels around 30 days postnatally (Mazella et al., 1996; Vincent et al., 1999).

Neurotensin expression in brain is well reported, however, some microarray experiments (Barlow database) fail to detect neurotensin expression in brain. However, this may be due to improper probe cell design in the microarrays, since expression of this neuropeptide is obvious in the brain (Fig. 2A). Microarrays do detect neurotensin receptor (nts2) expression in basically all brain regions examined, as well as in most body organs, with highest expression in kidney, thyroid and spleen.

Neuromedin B (NMB) and NMB receptor: NMB belongs to the bombesin-like peptides family, originally isolated from amphibian skin. There are two bombesin-like peptides in mammals: NMB, related to amphibian ranatensin, and gastrin-releasing peptide (Merali et al., 1999) (see below). NMB is expressed in lung and gastrointestinal tract, where it is released in response to food ingestion, to inhibit further food intake (Merali et al., 1999). The olfactory bulb has high expression levels based on microarray data (my own), which was validated by *in situ* hybridization. Central effects of NMB action include hypothermia. Published manuscripts report expression of the NMB receptor in olfactory nucleus, cortex, hippocampus, amygdala, hypothalamus and brain stem (Ohki-Hamazaki et al., 1999), whereas microarray data indicate that it is present in neocortex, amygdala, spinal cord and hippocampus and in other organs, most notably in prostate. In contrast to deletion of other bombesin-like peptide receptors (see below), NMB receptor knock-out mice have no obvious behavioral phenotype (Yamada et al., 2002).

Gastrin-releasing peptide (GRP) and GRP receptor: As mentioned above, GRP belongs to the bombesin-like family of peptides and is related to bombesin. It regulates food intake by inducing satiety and mediates gastric smooth muscle contraction (Ohki-Hamazaki et al., 1999). It is heavily expressed in the lung and gastrointestinal tract as well as in the brain. Central infusion of GRP produces hypothermia (Ohki-Hamazaki et al., 1999). It is expressed in the hypothalamus as well as in the lateral amygdaloid complex (Fig. 3C). It is interesting that microdialysis studies suggest that GRP is released at the central nucleus of the amygdala by both food intake and stressor exposure (Merali et al., 1998). This constitutes an example of the information flow in the amygdala: from lateral to medial structures, alluded to before. GRP receptor expression is mostly detected in large intestine and to a lesser degree, in kidney, by microarray measurements. Other microarray studies indicate that it is also expressed in cortex, hippocampus, amygdala, hypothalamus and brain stem. The receptor has been knocked-out and transgenic mice show increased social and locomotor activity (Yamada et al., 2002). Interestingly, there is a third type of bombesin-like peptide receptor, BRS-3, whose endogenous ligand has not yet been identified. BRS-3 deficient mice exhibit hyperphagia, obesity and decreased social behavior (Ohki-Hamazaki et al., 1997; Yamada et al., 2002).

Prolactin and prolactin receptors: Prolactin has been implicated in a variety of cellular functions, ranging from osmoregulation, lactation, reproduction and immunomodulation (Horseman et al., 1997). It is secreted mainly by the pituitary and placenta and can also modulate maternal behavior (Grattan, 2001). The prolactin receptor is a type-1 cytokine receptor encoded by a single multi-exon gene that produces many

isoforms from multiple promoter usages and alternative splicing (Ormandy et al., 1998). This generates products with identical extracellular domains, but different cytoplasmic domains, which affects signal transduction (Schuler et al., 2001). The receptor is expressed in multiple tissues, such as mammary gland, testis, lung, blood cells, liver, ovary and prostate (Horseman et al., 1997), which is also reported by microarray data. It is not clear, however, which isoform(s) are expressed in each case. In brain, cerebellum and hypothalamus are both rich in prolactin receptor based on microarray data, although no cerebellum expression was detected by *in situ* hybridization in my experiments (not shown). RNase protection assays indicate that the short and long isoform are expressed in striatum and substantia nigra, but expression of the short form was not detected by RT-PCR (Pi et al., 2002) by the same authors, maybe because of the lower sensitivity of the latter method. In any case, this illustrates how seemingly contradicting results can be obtained with the use of different detection techniques.

Despite expression in multiple organs, gene deletion by homologous recombination in transgenic mice indicated that prolactin receptor is necessary for mammopoiesis, but not for hematopoiesis, and also caused infertility in females (Horseman et al., 1997).

Retinoic-acid metabolizing cytochrome p450: I would like to also mention the identification of this gene, which has very localized expression patterns in the lateral and central amygdala (Fig. 5A). It was originally identified in my first screen, when comparing gene expression among amygdala and four other brain regions. The original transcript represented on the microarrays was an EST (TIGR cluster identifier

TC15611). To establish its identity I performed rapid amplification of cDNA ends (RACE) and computational "extensions" of this gene based on database information. This suggested that the gene was a cytochrome p450, but the conclusion was not firm since I could not extend the product (either by RACE or computationally) to the coding region and the alignments were not perfect. BLAST of the EST sequence to the genomic sequence database rendered alignment to a non-coding region, about ~2 kb downstream of a predicted cytochrome p450. I then synthesized a probe for *in situ* hybridization with the coding sequence of this latter gene and found that the expression pattern (Fig. 5B) was like that of the EST. This supported the idea that the EST indeed corresponded to a p450 cytochrome. Since then, cloning and expression of a novel retinoic-acid metabolizing cytochrome p450 (MacLean et al., 2001) supported that such was the identity of this gene. Interestingly, differential expression of related retinoic-acid metabolizing enzymes during mouse development was appreciated (Abu-Abed et al., 2002).

The p450 cytochrome genes constitute a superfamily composed of at least 74 gene families (Nelson et al., 1996), but the number is growing continually (Nelson, 1999). At least14 families exist in all mammals (Nelson et al., 1996), which comprise 26 subfamilies. Each subfamily usually represents a cluster of tightly linked genes widely scattered throughout the genome. In addition, many genes undergo alternative splicing and are driven from alternate promoters, which produces considerable transcript variability. Moreover, tissue and temporal expressions of cytochrome p450 family members are tightly regulated (Omura, 1999), which adds to their complex pattern of expression and suggests crucial physiological roles for these genes. Altogether, this raises

the intriguing possibility that specific brain regions possess different combinatorials of p450 cytochrome family members expression.

2.5.4 On the use of LCM and DNA microarray technologies to identify specifically expressed transcripts useful for gene targeting in transgenic mice

The results presented in these two chapters demonstrate that the combination of microarray analysis and *in situ* hybridization provide a useful means to identifying enriched genes in distinct brain regions. These genes could be used to target heterologous gene expression in transgenic mice. The ideal gene for targeting purposes should be very specifically expressed in the brain region of interest and nowhere else in the brain (or body). In addition, it should be expressed in such specific manner throughout development of the nervous system, or only after if it is formed[11]. However, it should be noted that the majority, if not all, of enriched genes identified so far by these screens have restricted domains of expression in more than just one brain region. Nevertheless, they could still be useful for designing transgenic mice if a combination of strategies is used. Binary systems could provide higher specificity. For instance, the promoter of a "semi-specific" gene (i.e., one that has more than one restricted site of expression) could drive expression of a heterologous gene of interest preceded by a transcriptional STOP cassette

[11] Unless, of course, one uses an inducible system.

(flanked by loxP sites). Then, a viral injection of Cre recombinase[12] in the brain region of study would locally activate the transgene (Wang et al., 1996). In addition, a similar strategy could be used if the promoter of a second "semi-specific" gene drives Cre expression instead. In this case, both semi-specific genes should have overlapping expression domains only in the region of interest. Other sites of expression for either gene would not matter, as the coexistence of both parts of the system is required for transgene expression. Examples of pairs of genes that could be used in these binary systems are shown in Table 4, provided they are expressed in the same cells.

[12] For a description of the Cre/loxP recombination system, see next Chapter.

Table 2-1: Number of genes differentially expressed.

Subarray U74v2A

Region/ replicate	1	2	3	Overlap in 2 replicates	Overlap in 3 replicates
C	9	18	6	2	5
L	8	8	10	0	0
M	12	36	8	2	2

Subarray U74v2B

Region/ replicate	1	2	3	Overlap in 2 replicates	Overlap in 3 replicates
C	13	10	8	3	1
L	9	11	2	0	0
M	19	51	23	5	4

Subarray U74v2C

Region/ replicate	1	2	3	Overlap in 2 replicates	Overlap in 3 replicates
C	14	8	11	5	0
L	12	10	5	1	1
M	29	32	13	5	1

All U74v2 subarrays combined

Region/ replicate	1	2	3	Overlap in 2 replicates	Overlap in 3 replicates
C	36	36	25	10	6
L	29	29	17	1	1
M	60	119	44	12	7

The number of genes enriched by at least 2.5-fold in each replicate with respect to the other six samples (3 replicates for each of the two remaining subnuclei) is indicated in the first three columns. The fourth (and fifth) columns show the number of enriched genes that were identified in 2 (or 3) replicates of the same subnucleus. The numbers are shown first separately for each of the U74v2 subarrays and then I present the data after combining all subarrays in the bottom table.

Table 2-2 Number of *Present* and *Absent* genes.

Region	**Present and avg diff > 200**	**Absent or avg diff < 200**
3 subnuclei	3155 (25 %)	2060 (16 %)
Central	4210 (34 %)	3057 (24 %)
Lateral	4161 (33 %)	2938 (24 %)
Medial	3578 (29 %)	2800 (22 %)

The number of genes on subarray U74v2A Present based on the Affymetrix software and with average difference (avg diff) values larger than 200 are indicated in the first column. The number of genes that were either Absent or had average difference values lower than 200 are indicated in the second column. The genes with average difference values larger than 200, which were present in all three subnuclei, are indicated in the first row. Percentages of the total 12,488 genes and ESTs represented on the U74v2A subarray are indicated in parentheses.

Table 2-3 Affymetrix parameters of selected genes.

	Amygdala-enriched genes relative to 4 other brain regions	Amygdala-subnuclei – enriched genes
number of genes analyzed	13	8
positive / pairs used	60 %	66 %
absolute call: *Present*	92 %	100 %
differential call: *Decrease*	62 %	60 %
differential call: *No Change*	38 %	40 %

Inspection of selected parameters provided by the Affymetrix software for selected genes whose expression patterns were confirmed by *in situ* hybridization. Parameters for the screen described in Chapter 1 are presented in the first column. The second column shows the results obtained in the screen described in this chapter.

Table 2-4 Combinations of "semi-specific" genes useful for gene targeting.

Gene 1	Gene 2	Target region
X76653, arp-1 (chapter 1, fig. 2D)	TIGR cluster TC30886, chaperonin ε subunit (chapter 1, fig. 3F)	lateral and medial amygdala
X76653, arp-1 (chapter 1, fig. 2D)	AB041584, unconventional type myosin (chapter 1,fig 3A)	lateral (but not basolateral) amygdala
D10214, prolactin receptor (chapter 2, fig. 4A)	U43298, laminin β3 (chapter 1, fig. 2C)	medial amygdala
TIGR cluster TC15611, retinoic acid metabolizing p450 cytochrome (chapter 2, fig. 5A and 5B)	BE135978, EST (Chapter 2, fig. 2C, probe 153/138.)	central amygdala

Examples of combinations of "semi-specific" genes useful for gene targeting in transgenic mice. Genbank accessions or TIGR identifiers of pairs of genes that show overlapping expression in the target regions are indicated. It remains to be tested, however, whether these genes are expressed in the same cells.

2.5.5 Figure legends

Figure 2-1: Distribution of normalized average difference values.

Distribution of normalized average difference ($\bar{\Delta}$) values in each of the nine samples (C1, C2, C3= each of the central amygdala replicates; L1, L2, L3= each of the lateral amygdala replicates; M1, M2, M3= each of the medial amygdala replicates). Only the U74v2a array is plotted here, but all other chips had similar distributions. For clarity, only values below 8,000 are shown. Dotted lines indicate the mean value. Bin size is 200.

Figure 2-2: Genes enriched in the central nucleus.

Genes enriched in the central nucleus. The central nucleus is indicated with a black arrow, the lateral with a white arrow and the medial with an arrowhead. A: Probe 873: neuromedin N/ neurotensin. B: probe Bh1, neuromedin B precursor. C: Probe 153/ 138, EST. D-F: Subnuclei are delineated with a dash line based on neurotrace counterstaining of sections A-C, respectively.

Figure 2-3: Genes enriched in the lateral nucleus.

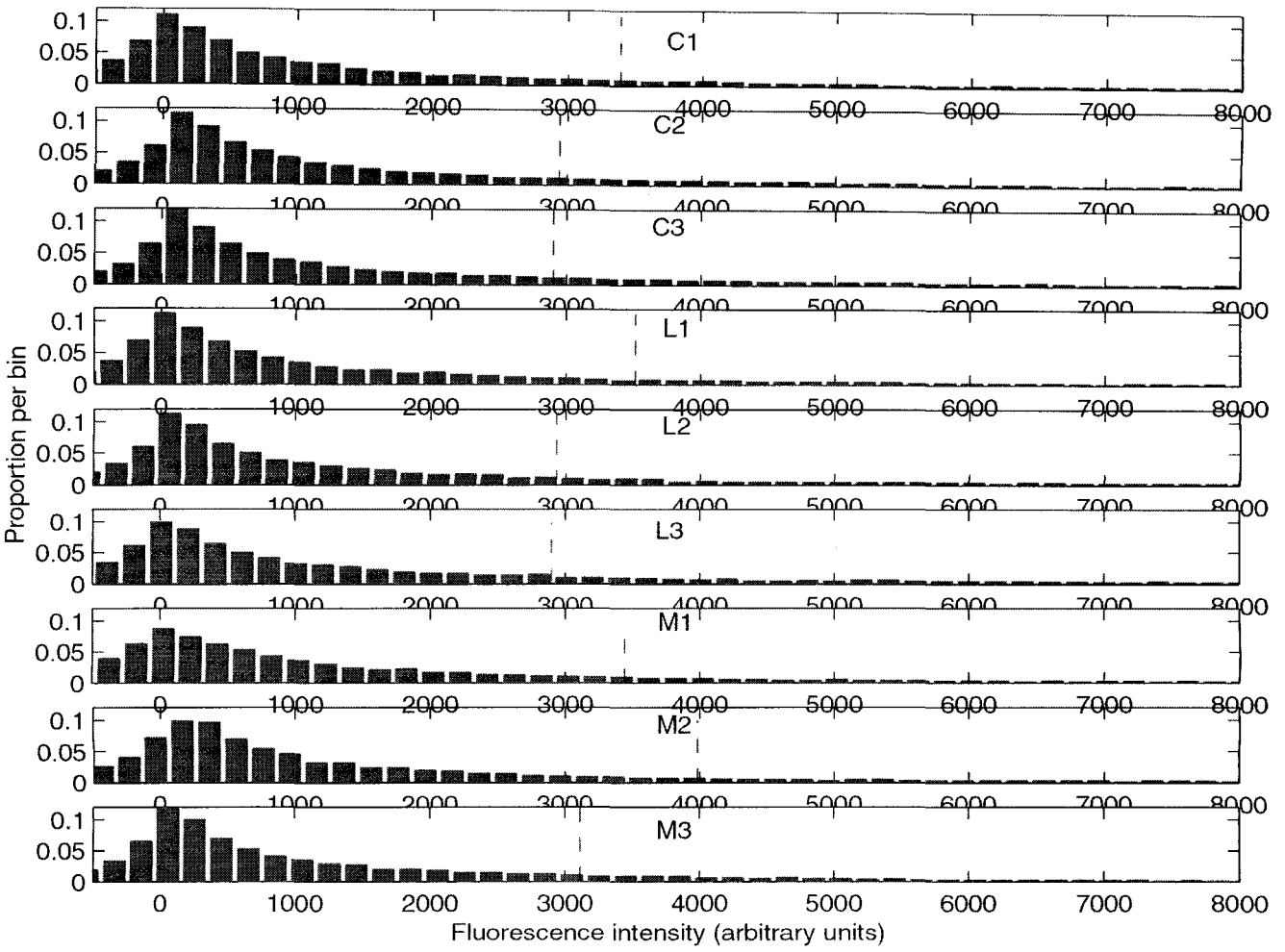
Genes enriched in the lateral nucleus. The lateral nucleus is indicated with a black arrow, the central with a white arrow, and the medial with a white arrowhead. Black arrowhead indicates cortex in A and B, and basomedial amygdaloid nucleus in C. A: Probe 117, kinesin-like protein. B: probe 126, EST. C: Probe 440, gastrin releasing peptide. D-F: Subnuclei are delineated with a dash line based on neurotrace counterstaining of sections A-C, respectively.

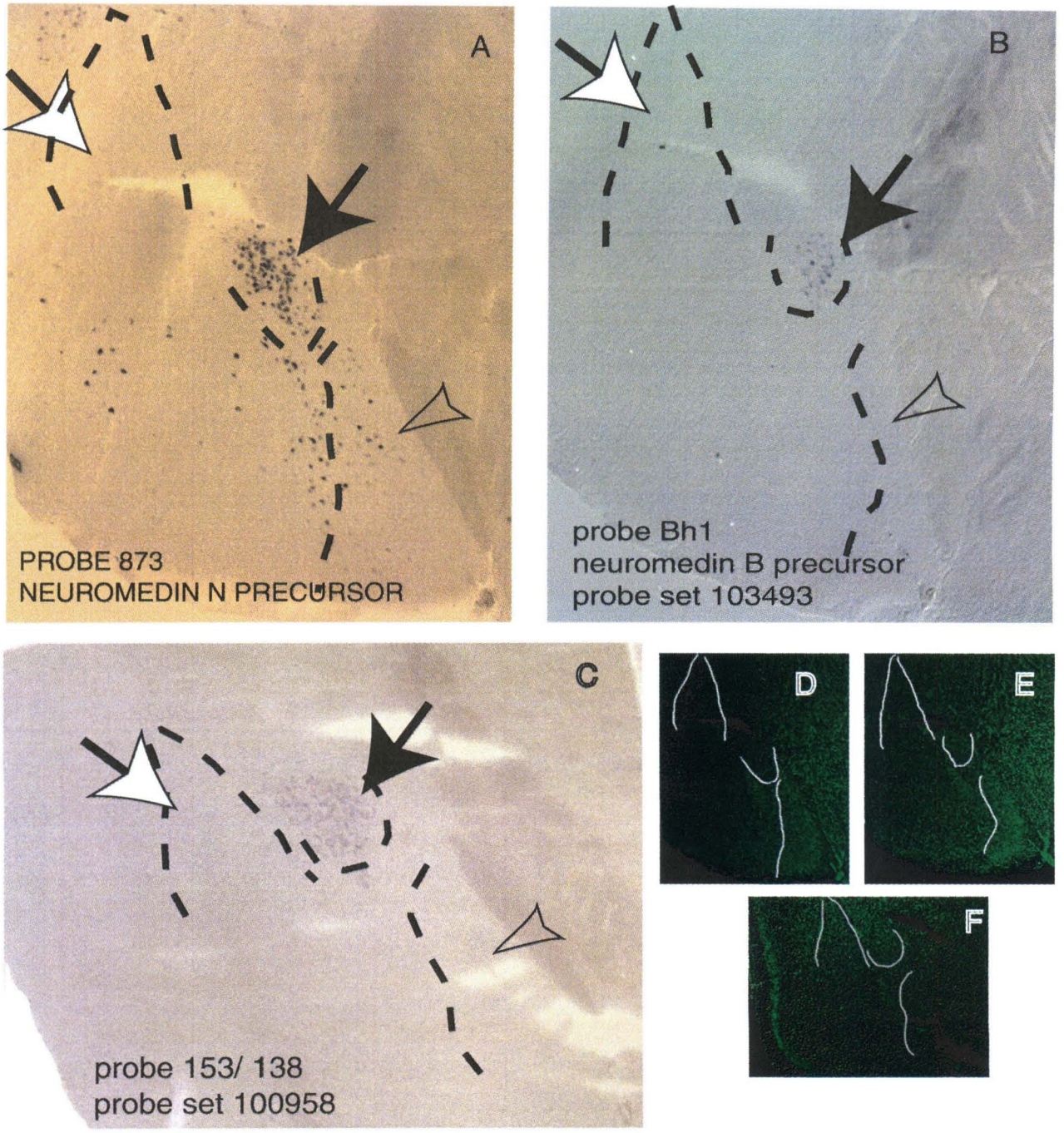
Figure 2-4: Genes enriched in the medial nucleus.

Genes enriched in the medial nucleus. The medial nucleus is indicated with a black arrow, the lateral with a white arrow and the central with an arrowhead. A: Probe 133, prolactin receptor. B: probe c3, thyrotropin releasing hormone receptor. C: Nissl staining of a consecutive section. D-E: Subnuclei are delineated with a dash line based on neurotrace counterstaining of sections A and B, respectively.

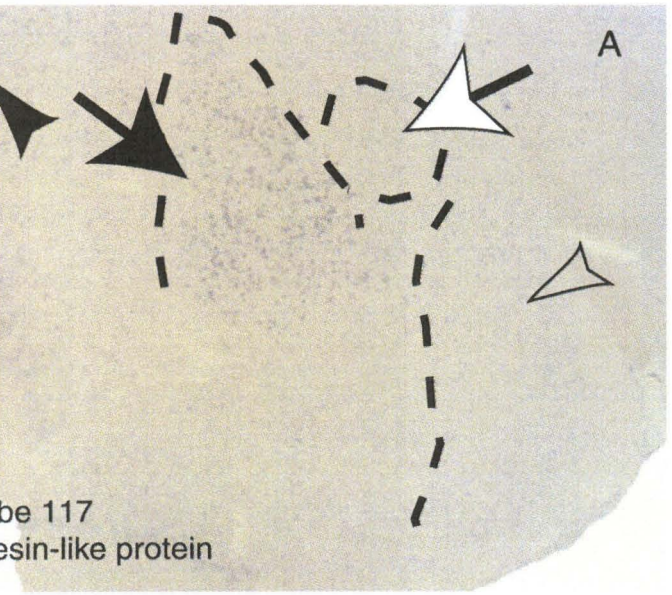
Figure 2-5: p450 cytochrome.

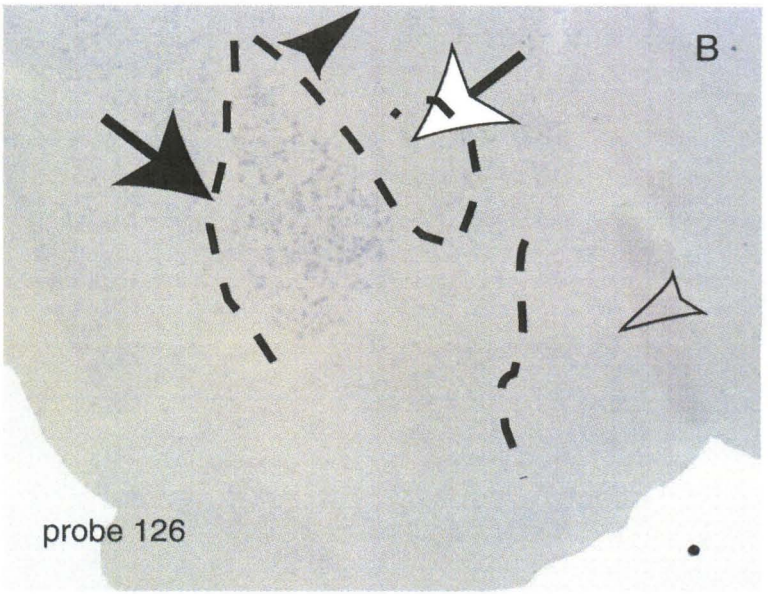
(A) *In situ* hybridization signal in the lateral and central amygdala obtained with probe TC15611, derived from an EST. By 5' RACE, it was determined that it corresponded to retinoic acid metabolizing p450 cytochrome. (B) *In situ* hybridization with retinoic acid metabolizing p450 cytochrome probe shows the same signal as in (A). A-B: black arrows: lateral amygdala nucleus; white arrows: central nucleus. (C) Note extra-amygdaloid expression in subiculum (black arrow) and dentate gyrus (white arrow).

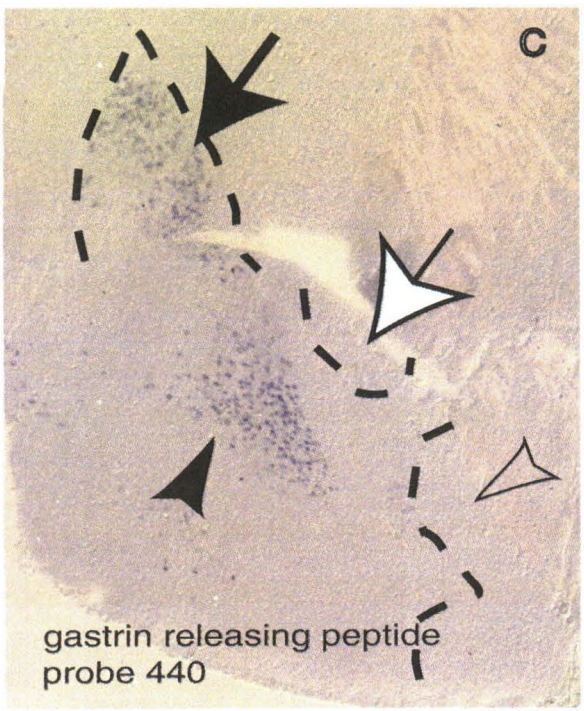

C1
C2
C3
L1
L2
L3
M1
M2
M3
0.1
0.05
0
0
1000
2000
3000
4000
5000
6000
7000
8000
Proportion per bin
Fluorescence intensity (arbitrary units)

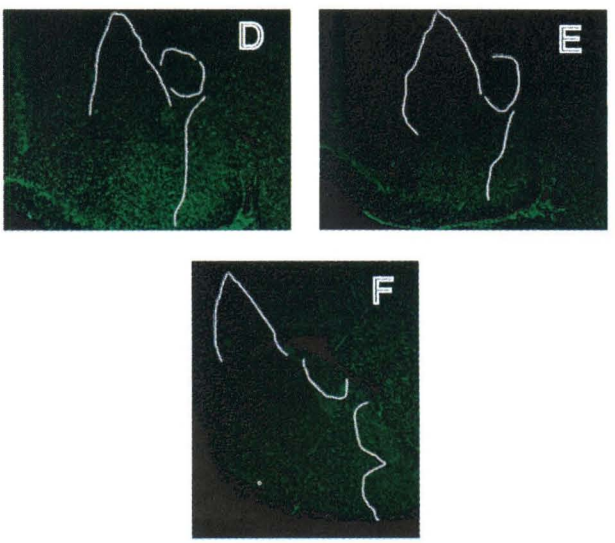


Chapter 2- Fig. 2

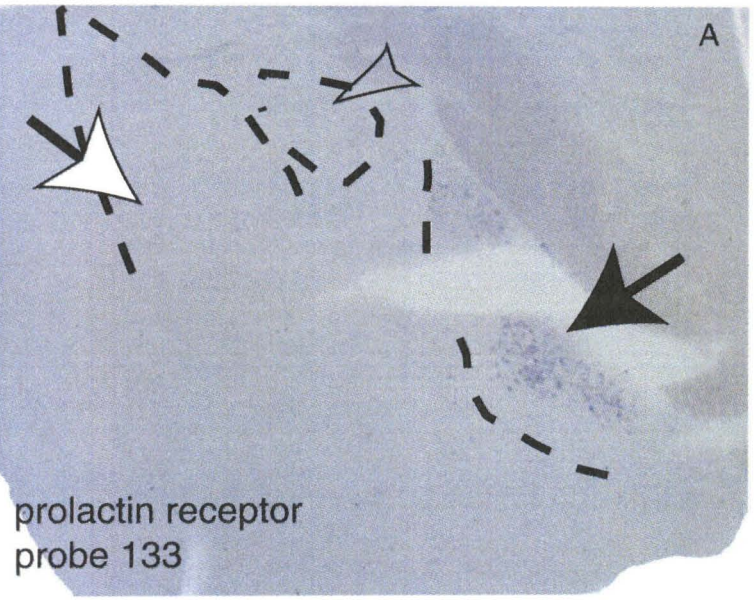


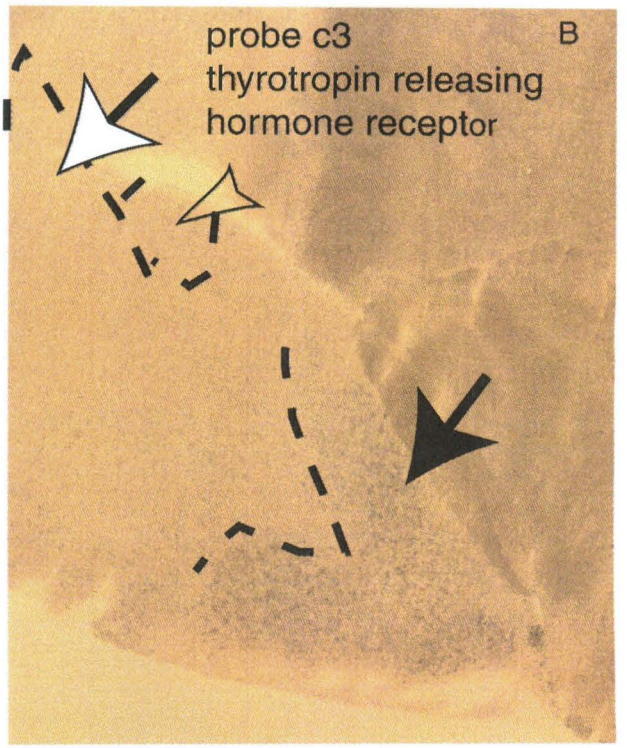


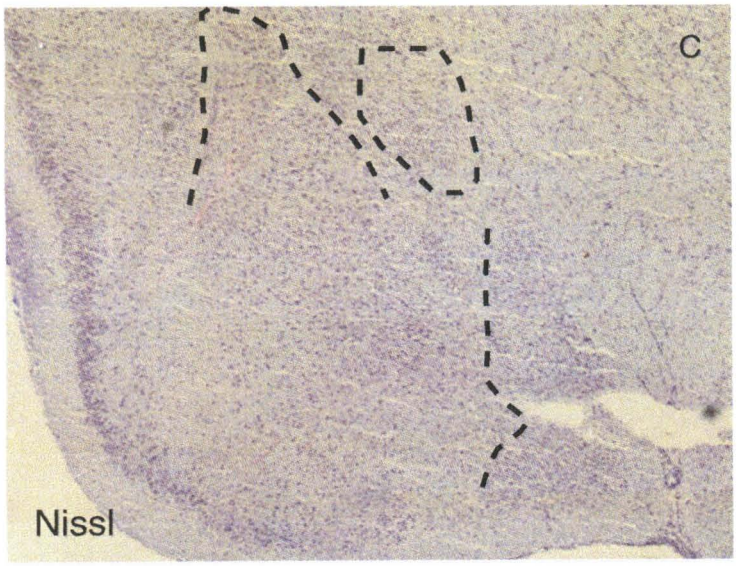


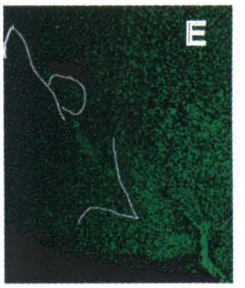


Chapter 2- Fig. 3

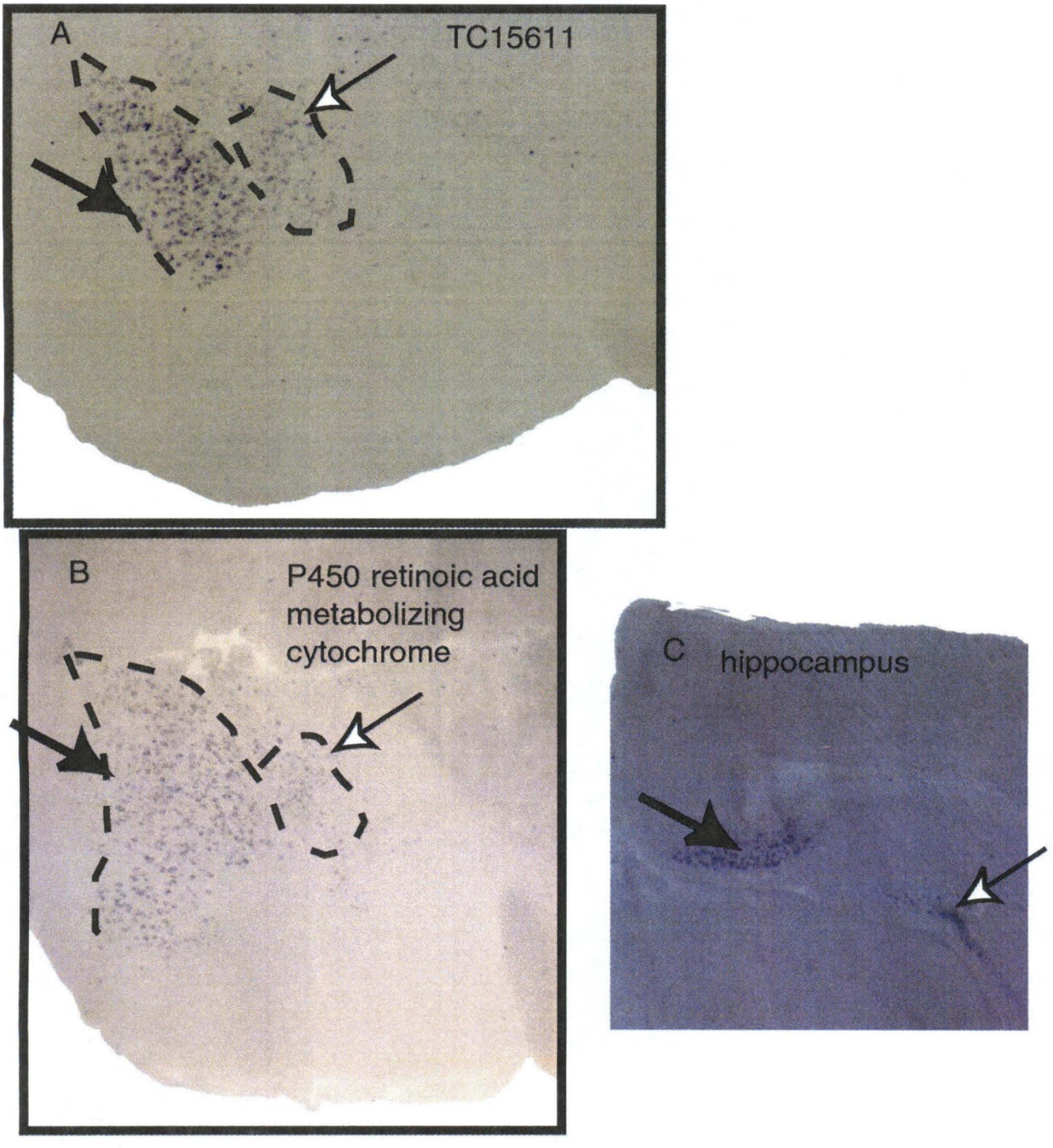








Chapter 2- Fig. 4



Chapter 2- Fig. 5

APPENDIX 1: ANALYSIS OF QUANTITATIVE GENE EXPRESSION DATA

The availability of techniques to simultaneously interrogate levels of expression of thousands of genes in a quantitative way has generated unprecedented amounts of data in biology. However, it is not a simple task to find transcripts that reliably change expression levels under various conditions or in distinct samples among thousands of others that are invariant.

Several new manuscripts describing algorithms, heuristics and procedures to analyze microarray experiments have appeared in the last two or three years (see references below). Furthermore, there is at least one commercial software targeted specifically to analyze gene expression data (GeneSpring, Redwood City, CA) and several other bioinformatics companies that offer both custom software and technical support to study DNA microarray data (for example, Informax Inc., Bethesda, MD). Other clustering algorithms are publicly available to academic researchers (e.g., Tamayo et al., 1999).

The hypothesis, methods and types of questions addressed with these different tools overlap but are not necessarily identical. Here I briefly review some of these methods. (For a short description of these, please refer to

http://www.silicongenetics.com/cgi/SiG.cgi/Products/GeneSpring/tools.smf). I also directly compare the results obtained with our program (see Chapters 1 and 2) and those obtained with another powerful software, Bullfrog, developed by David Lockhart and others, now working at the Salk Institute (San Diego, CA).

Description of other algorithms and heuristics

One of the most common methods used to analyze gene expression data involves the use of clustering algorithms. The general idea behind these algorithms is to separate all the transcripts into clusters that share common expression patterns. The underlying hypothesis is that there are groups of genes with common properties, perhaps participating in a common biochemical pathway, that are co-expressed under certain experimental conditions. This has proven to be very useful in some cases (e.g., (Brown et al., 2000; Eisen et al., 1998; Tamayo et al., 1999), reviewed in (Sherlock, 2000)).

One of the most commonly employed clustering algorithms is based on the nearest neighbor (or k-nearest neighbors) algorithm and hierarchical clustering (Eisen et al., 1998). Unfortunately, one of the drawbacks in the most typical applications of these clustering ideas is that the number of clusters is fixed beforehand. This seems to be a rather artificial situation. If one were to generate data from a single unimodal distribution, these algorithms would still produce several clusters as output. In other words, even in the absence of clusters of genes with common expression, these algorithms "find" plenty of genes belonging to separate clusters. It should be noted that the lack of clusters (or the presence of small clusters) is a perfectly valid biological scenario. It may happen that under particular experimental conditions only a small set of transcripts (out of several thousands) change their expression pattern. A more rigorous and systematic approach to clustering involves the use of Bayesian statistics based on latent variables (see, e.g., Long et al., 2001). Using Expectation-Maximization or Support Vector Machines (Brown et al.,

2000) allows the direct comparison of distinct models based on different numbers of clusters. These techniques are perhaps less widely used now in the DNA microarray analysis community, perhaps due to their increased complexity and computational requirements. However, they permit to simultaneously estimate the cluster boundaries as well as the optimal number of clusters.

Other approaches attempt to directly look for genes that are overexpressed in one sample with respect to all others. It would be interesting to estimate the probability that a given gene shows apparent enrichment by chance. The quantitative way of doing this would require establishing the variability of repeated measurements of a given gene in one sample and comparing this to the difference in expression levels for the same gene in other samples. Unfortunately, the high cost of DNA microarrays still precludes from obtaining a large number of replicates. Even in the best scenario, it is hard to prepare more than two or three replicates. (The importance of replicates has already been discussed in Chapter 2.)

Most of the methods to analyze gene expression datasets are based on studying the changes in the value called "average difference"[13], already described in Chapter 1. This variable is fundamental in estimating changes in levels of gene expression. However, as noted in the previous chapters, very important information can also be obtained by paying attention to the number of positive and negative probe cells for each transcript (recall that there are about 20 different match and mismatch probe cells to interrogate the expression of each transcript and the average difference value is just one parameter derived from microarray scannings). In our case, ignoring these other

[13] Or relative fluorescent hybridization signals in cDNA arrays.

parameters would lead to an increase in the number of reported enriched transcripts, incrementing the number of false positives.

In general, two types of errors can be made in the analysis of DNA microarray data: misses and false positives. The former refers to the probability that an interesting candidate gene is overlooked. The latter refers to the probability that the reported genes do not show a biologically relevant change in expression. Unfortunately, for most of the papers described above, no independent validation of the results by other methods is reported and is therefore difficult to assess the accuracy of the methods. It is not uncommon to see papers reporting only a few candidate genes or describing hundreds or thousands of potentially interesting genes divided into tens of clusters. The biological significance of these results without further validation remains unclear.

I have already discussed the validation by *in situ* hybridization of the results obtained with our algorithmic methods (Chapters 1 and 2). I turn now to the comparison of our results to those obtained by other microarray data analysis methods.

Comparison of our results to those obtained with other algorithms

Genecluster

It was already mentioned in Chapter 1 that our custom algorithm[14] proved more efficient than Genecluster, which implements Self Organizing Maps, in the identification of genes enriched in single brain regions. In our first screen, the number of genes in

[14] Our custom programs are available at http://www.its.caltech.edu/~mariela/gene_screen.html

clusters derived from single region-overexpressed transcripts was 6-10 fold larger than the number of enriched genes identified with our algorithm, depending on the parameters used for clustering. The extra genes belonging to these clusters were mainly expressed at very low levels or did not show a fold change in expression high enough with respect to all the other regions to be considered good candidates for follow-up experiments. Since we were also interested in identifying genes that were enriched in single regions, we did not perform this clustering analysis in our later screen.

Bullfrog

Contrary to clustering algorithms, Bullfrog was specifically developed to analyze microarray data obtained with Affymetrix chips. This provides the opportunity to filter data not only based on average difference values, but also on other parameters calculated by the Affymetrix software, such as *Present Absent* calls, or *Decrease/No change* differential calls[15]. The advantage of this is that the significance of the differential hybridization to the arrays of biotinylated cRNA probes from different samples can be also taken into account. Let's take a hypothetical case: a high average difference value that is mainly due to unusually high hybridization of the labeled cRNA probe to a single perfect match cell among the ~20 representing a particular gene. However, other perfect match probe cells do not show intense hybridization, indicating that the gene may not be really present and that hybridization to a single perfect match probe cell is an artifact. Bullfrog is designed so that this hypothetical gene's high average difference value is not misleading. In other words, the *Present/Absent* calls are considered together with average

[15] Please see Chapter 2 for definition of these various parameters.

difference values to decide if a gene is differentially expressed. In contrast, clustering algorithms would treat this gene as a real, high-abundance transcript.

Bullfrog is user friendly because it allows the experimenter to set different criteria for the identification of enriched genes. In addition, it allows the consideration of replicate experiments so that multiple comparisons can be made between different replicates of distinct samples. Let me illustrate this with the comparison of gene expression between three samples: sample A (with replicates A1 and A2), sample B (with replicates B1 and B2), and sample C (with replicates C1 and C2).

Before feeding the data into Bullfrog, it is required that all possible pair-wise comparison files are generated by the Affymetrix software. For example, to identify genes enriched in A, A1 and A2 are designed as "baseline" files and B1, B2, C1, C2 are the 'experimental' files that must be generated in comparison to the baseline files. In this case, one would need to generate eight Affymetrix comparison files as follows:

experimental file	baseline file
B1	A1
B1	A2
B2	A1
B2	A2
C1	A1
C1	A2
C2	A1
C2	A2

The developers of Bullfrog recommend the following set of criteria to identify genes enriched in A with respect to both B and C:

i) Gene must be Present in the Baseline file in at least 6 out of the 8 comparisons

ii) Gene must have an average difference value >625[16] in at least 6 of the 8 baseline files

iii) Gene must have a Decrease call in at least 6 out of the 8 experimental files

iv) The ratio of [average difference in baseline / average difference in experimental file] must be > 1.8 in at least 6 out of the 8 comparisons.

The versatility of the program can be appreciated, as all underlined values can be independently modified by the user. (It should be noted that 16 analogous comparison files must be generated using B and C as baselines if genes enriched in these samples are also sought.)

To analyze my data from the U74v2A subarray, I changed some parameters so that the criteria implemented resembled the criteria imposed in my custom program (cf. Chapter 2, Data Analysis section). Thus, I modified constraints ii) and iv) as follows:

ii) Gene must have an average difference value >200 in at least 6 of the 8 baseline files

iv) The ratio of average difference in baseline / average difference in experimental file must be > 2.5 in at least 6 out of the 8 comparisons.

Constraints i) and iii) were not modified.

[16] For an array that has been normalized to a target intensity of 2,500.

I first determined the pairs of replicates within each amygdala subnucleus that were the most similar. The pairs that had the least variability were C2 and C3, L2 and L3, and M1 and M2, which differed in between 15-100 genes each. I then used these pairs to generate 24 (8 x 3 samples) comparison files as described above.

The numbers of genes identified with these conditions are listed in Table 1 (in Bullfrog-1 column). The table also shows the number of enriched genes identified with the custom program (cf. Table 1, Chapter 2). I noticed that according to Bullfrog, the medial amygdala had an unusually large number of enriched genes. The reason for this is not clear, but may be due to partial degradation of cRNA biotinylated probe. (Please refer to footnote 5 in Chapter 2 for discussion.)

Further inspection of all M triplicates suggested that M3 was the most different from the 3, but it is probably the best, since it contains the largest number of *Present* genes (47% vs. 37% and 38% in M1 and M2, respectively)[17]. Thus, I reanalyzed the dataset but considering M3 only for the medial nucleus. The analytical criteria and replicates from the central and lateral nuclei were otherwise the same. The results are presented in the column Bullfrog-2 in Table 1. In this case, the number of enriched genes in the medial region resembles that of the other nuclei, whose numbers of enriched genes also vary slightly. The genes identified by Bullfrog-1 and Bullfrog-2 are not all the same. The number of total genes differentially enriched identified by the combination of both is listed in the last column. It can be appreciated already how sensitive the results are with respect to variations in replicate samples. Replicate M2 was particularly "noisy" and thus changed the identification of enriched genes not only in the medial region, but also in the

[17] C2, C3, L2 and L3 had ~45% of *Present* genes each.

samples from the other regions. Results obtained with our custom program also indicated that the M2 sample was different from the rest. Once more, the value of performing replicate analysis when synthesizing cRNA probes with sensitive steps such as LCM-extraction and RNA amplification is underscored.

Even though the order of magnitude of the number of genes differentially enriched was the same for Bullfrog or our custom analysis, their identities were not necessarily the same. 25%, 42%, and 33% of enriched genes were identified by both programs in the central, lateral, and medial amygdala, respectively. More significant perhaps is the efficiency with which Bullfrog identified the genes that were confirmed to be differentially enriched by *in situ* hybridization. Of these, 60% were identified by Bullfrog.

It is important to keep in mind that both programs have their pros and cons. Bullfrog has the potential to be more stringent, since many more parameters can be independently modified. At the same time, it may be more sensitive to duplicate variation. However, in some cases, allowing for some variation between replicates may facilitate the identification of extra enriched genes. In addition, the need to generate all Affymetrix pairwise comparison files is not very practical. Nevertheless, I believe both programs may prove even more helpful if some modifications to the Affymetrix software are done when reading signal hybridization from probes derived from LCM extraction and RNA amplification steps. These may consist, for example, on re-specifying the criteria for determining positive and negative probe cells, which in turn, affect the *Present/Absent* and *Decrease/No change* calls. Several iterations of re-specification of

such criteria followed by independent experimental confirmation may be required to fine-tune the Affymetrix output for better data analysis in such cases.

Table Ap-0-1: Comparison with Bullfrog.

region	custom	Bullfrog-1	Bullfrog-2	Bullfrog-1 + -2
Central	26	5	20	20
Lateral	26	10	4	12
Medial	52	82	36	111
total	104	97	60	143

Numbers of differentially enriched genes identified with our custom software and with Bullfrog among the 12,488 genes represented on the U74v2A subarray. Replicates M1 and M2 were used in Bullfrog-1, whereas M3 was used in Bullfrog-2.The last column shows the number of total genes that were identified in both cases.

3 NEURON SUBTYPE SPECIFICATION PRECEDES NEURON-GLIAL FATE DETERMINATION

3.1 Introduction

Prominent cellular diversity is a key attribute of the nervous system. Neural progenitor cells are responsible for generating such diversity. They give rise to various neurons and glia in different locations, thus playing a role in the complex intercellular relationships characteristic of the nervous system. In addition to the generation of cellular diversity, the normal development of the nervous system requires that different types of cell develop at specific times and in appropriate ratios. To understand how neural progenitor cells develop, it is important to characterize these cells and determine the cell types that they give rise to. I have analyzed some aspects of neural cell lineage diversification in a cellular subpopulation of the developing peripheral nervous system, which is amenable to such studies. Neural crest cells are a transient cellular population, induced at the dorsolateral edge of the neural plate, from where they delaminate and migrate along specific routes to many destinations in the vertebrate embryo. These cells differentiate into a wide variety of cell types, including neurons and glial cells of the peripheral nervous system, pigment cells, smooth muscle and cartilage and bone in the head (Le Douarin and Kalcheim, 1999). What is the strategy used by these cells to generate such diversity?

The following are two extreme alternative hypotheses: i) Do neural crest cells represent a homogeneous, multipotent population of precursor cells which generate

different derivatives influenced by environmental factors; or rather, ii) are there pre-specified progenitors with predetermined limited developmental capacities?

Numerous experiments suggested that at least some neural crest cells are pleuripotent (Bronner-Fraser and Fraser, 1988; Bronner-Fraser and Fraser, 1989; Serbedzija et al., 1994), generating sensory and sympathetic neurons and their associated glia. However in these experiments, other NCCs generated only a subset of these derivatives. Others (Henion and Weston, 1997) also implied the occurrence of early fate restriction. Does the existence of restricted fates truly reflect precursor specification? Or is it simply due to stochastic variations in the behavior of a homogeneous, multipotent population?

Reexamination of these issues was required. I have used a genetic system to permanently label a subpopulation of neural crest cells marked by the expression of the proneural gene *Neurogenin-2*. The strategy involved inducible Cre recombinase-dependent expression of a reporter gene in *Neurogenin-2* expressing cells and their progeny. *Neurogenin-2* is a bHLH DNA-binding factor necessary for the development of some sensory neurons (Ma et al., 1999) but dispensable for the generation of autonomic neurons (Fode et al., 1998) and is expressed early in the ontogeny of some neural crest cells.

By comparing the fate of these cellular derivates to the bulk crest population marked by expression of *Wnt1*, I concluded that at least some neural crest progenitor cells are strongly biased for a sensory rather than an autonomic sublineage. Specifically, *Ngn2*$^+$ progenitors were four times more likely than *Wnt1*$^+$ neural crest cells to contribute to sensory rather than sympathetic ganglia. Within dorsal root ganglia, however, both

Ngn2- and *Wnt1*-expressing cells were equally likely to generate neurons or glia. These data suggest that *Ngn2* marks a neural crest subpopulation with a predictable fate bias, early in migration. Moreover, such neuronal subtype specification is acquired before commitment to a particular neuronal or glial fate. These data, together with other work (Lu et al., 2002; White et al., 2001; Zhou and Anderson, 2002) challenge the hitherto textbook view of neural cell lineage segregation, which stated that neuronal and glial precursors were segregated first, before neuronal subtype specification occurred (Anderson, 2001; Gage, 2000; Temple and Qian, 1996).

Parts of this chapter have been reported elsewhere (Zirlinger, M., Lo, L., McMahon, J., McMahon, A. P., and Anderson, D. J. (2002). Transient expression of the bHLH factor Neurogenin-2 marks a subpopulation of neural crest cells biased for a sensory but not a neuronal fate. Proc. Natl. Acad. Sci. USA, *In press*.)

.

At the end of this chapter I briefly discuss our attempts to obtain Cre-dependent reporter expression in the adult brain.

We thank C. Schuurmans and F. Guillemot for providing the *Ngn2* genomic construct and *Ngn2-lacZ* mice, P. Soriano for Rosa26-loxp reporter mice, B. Kennedy, S. Pease and the staff of TAFCIT for expert help in the generation and maintenance of genetically modified mice, L. Reichardt and F. Rice for anti-Trk antibodies, and C. Birchmeier for the anti-BFABP antibody. We thank G. Kreiman for help with mathematical simulations, J. Yamada for genotyping, S. Pintchovski and G. Mosconi for help with experiments.

3.2Transient expression of the bHLH factor Neurogenin-2 marks a subpopulation of neural crest cells biased for a sensory but not a neuronal fate

The neural crest presents a model system for studying neural cell lineage segregation. Neural crest cells (NCCs) generate the sensory and autonomic neurons of the peripheral nervous system, and their associated glia, after migrating from the dorso-lateral neural tube (Le Douarin and Kalcheim, 1999). Single-cell lineage marking experiments in vivo have indicated that some pre-migratory trunk NCCs are pleuripotent, generating neurons and glia in both sensory and autonomic (sympathetic) ganglia (Bronner-Fraser and Fraser, 1988; Bronner-Fraser and Fraser, 1989; Serbedzija et al., 1994). However in these experiments, other NCCs generated only a subset of these derivatives, e.g., in sensory but not autonomic ganglia. Because the marked cells were selected at random, it could not be distinguished whether NCCs are homogeneous and pleuripotent, but exhibit heterogeneity in their fates due to stochastic variation, or rather are heterogeneous and comprise subpopulations with deterministic fate restrictions that could not be prospectively identified. Lineage-marking studies performed on migrating neural crest cells in vitro have revealed evidence of rapid restriction to neuronal or glial fates (Henion and Weston, 1997), but the neuronal subtype(s) involved were not examined.

Evidence of molecular heterogeneity among early NCCs has been suggested by the often transient expression of various antigenic markers and genes (e.g., Barbu et al., 1986; reviewed in Le Douarin and Kalcheim, 1999). However, without some way to convert the expression of such markers into a permanent lineage tracer, it remained

unclear whether this heterogeneity reflected early fate-specification. Here we have achieved such a conversion by using a conditional form of Cre recombinase to permanently mark a subset of NCCs that transiently express the proneural bHLH transcription factor *Neurogenin2* (*Ngn2*) (Gradwohl et al., 1996; Sommer et al., 1996). *Ngn2* is required for the differentiation of a subset of sensory neurons (Ma et al., 1999), but is dispensable for autonomic neurons and for Schwann (glial) cells in peripheral nerves (Fode et al., 1998). The genetic requirement for *Ngn2* in sensory neurons, however, leaves open the question of whether it is expressed by progenitors restricted to a sensory neuron fate (Fig. 1, A4), or rather by cells with a broader developmental potential (Fig. 1, A1-A3).

To address this question, we have employed a conditional, binary system for fate-mapping of *Ngn2*-expressing progenitor cells in vivo based on Cre-lox-mediated DNA recombination (Danielian et al., 1998; Zinyk et al., 1998). We have expressed a 4-hydroxy tamoxifen (4-OH Txf)-inducible form of Cre recombinase, CreER^{TM} (Brocard et al., 1997; Danielian et al., 1998; Metzger et al., 1995), from *Ngn2* genomic regulatory elements in mice (Fig. 1B). Such *Ngn2-CreER^{TM}* mice are then crossed to mice carrying a ubiquitously expressed, Cre-dependent *lacZ* reporter gene (Soriano, 1999). In embryos from this intercross, *lacZ* expression can be activated only in *Ngn2*-expressing progenitors. Because activation of the reporter gene involves a cell-heritable DNA rearrangement event, *lacZ* expression will persist in the progeny of transiently *Ngn2*-expressing cells (Fig. 1C). Since 4-OH Txf has a half-life of only 0.5-2 hrs in vivo (Danielian et al., 1998), activation of CreER^{TM} can be further restricted to a relatively narrow developmental time window when the ligand is injected. Thus the final

expression pattern of the *lacZ* reporter will exclusively identify the progeny of cells expressing *Ngn2-CreER*^{TM} during this time window, and will not include cells that express *Ngn2* at later times (Fig. 1C, large lower oval).

3.2.1 Material and methods

Mouse manipulations: Homologous recombination in ES cells was employed to replace the *Ngn2* coding sequence with *CreER*^{TM} (Danielian et al., 1998), via the same strategy used to generate *Ngn2* knockouts (Fode et al., 1998). *Ngn2-CreER*^{TM} mice bred into a C57Bl6/J background were crossed to Rosa26-loxp reporter mice (Soriano, 1999) to generate embryos for analysis. In other experiments, mice harboring a *Wnt1-CreER*^{TM} transgene (Danielian et al., 1998) were bred to the same line of reporter mice. 4-OH Txf was injected intraperitoneally (1 mg/mouse) to activate Cre-ER^{TM}. Injections to activate CreER^{TM} at E8.75-E9.0 were performed between 5 PM and midnight (counting the morning the vaginal plug was identified as E0.5), and to activate it at E9.5 were performed between 8 AM and noon the following day. Embryos were analyzed at E12.5, or in some experiments at E10.5.

Histology: Embryos were fixed in 4% PFA for 1-2 hr at RT, embedded, frozen and sectioned at 15 μm. For double-labeling (Durbec and Rougon, 2001), sections were first X-gal-stained (Beddington et al., 1989) overnight at room temperature, post-fixed in 4% PFA, pre-blocked, and then incubated for 1.5 hr at RT with the appropriate primary antibodies: mouse monoclonal anti-Isl-1 (IgG1, Developmental Studies Hybridoma Bank,1:1 dilution); rabbit anti-TH (Sigma, 1:1000 dilution); rabbit anti- BFABP (C. Birchmeier, 1:1000 dilution); rabbit anti-trk-A (L. Reichardt, 1:2000 dilution); rabbit

anti-trk-B (L. Reichardt, 1:1000 dilution); goat anti-trk-C (L. Reichardt, 1:500 dilution). Antibody staining was developed with DAB using the Vecstatin ABC kit (Vector labs, Inc.). Fluorescent secondary antibodies were Alexa red anti-rabbit (Molecular Probes) or FITC-conjugated anti-mouse IgG1 (Southern Biotech), (1:250 dilution).

3.2.2 Overview of the genetic strategy employed

Current molecular techniques allow precise genetic modifications in the mouse. Not only can specific nucleotide sequences be targeted in the genome but also genetic switches can be designed to turn genes on (or off) at desired times. A powerful tool for the creation of modified animals is the Cre-site specific DNA recombinase of bacteriophage P1 (reviewed in (Sauer, 1998)). Cre is a 38-kDa product that recognizes a 34-bp site called loxP and efficiently catalyzes reciprocal conservative DNA recombination between pairs of loxP sites, which results in excision of the DNA between them. The loxP site consists of two 13-bp inverted repeats flanking an 8-bp nonpalindromic core region. Importantly, no accessory host factor or DNA topological requirements are needed for efficient Cre-mediated DNA recombination. A useful way of controlling expression of a transgene is to design it as a "dormant" transgene so that it can be only activated under specific circumstances. One way of doing this is by inserting a STOP signal flanked by loxP sites, between the transgene and its promoter. The STOP signal can be removed by Cre-mediated excision, for example, by intercrossing with a second mouse expressing Cre in specific tissues. The STOP sequence consists of a spacer DNA (from the yeast HIS3 gene), the small intron and polyadenylation signal from SV40, followed by a gratuitous ATG translation start and 5' splice donor signal to

suppress correct expression from any residual transcription of the desired downstream gene (Lakso et al., 1992). Thus, this cassette prevents expression of the downstream gene until the STOP signal is removed. The tissue specificity of expression for the recombinationally activated dormant transgene is a function both of the promoter specificity of the target dormant transgene and of the promoter specificity of the Cre transgene, so that the resulting expression pattern of the dormant transgene is confined to the intersection of the two expression patterns. For instance, intercross of a transgenic mouse that expresses the dormant transgene ubiquitously and one that expresses Cre recombinase in a spatially or developmentally limited fashion will provide activation of the dormant transgene that is dictated solely by the promoter driving Cre. In this case, the dormant transgene was simply a lacZ reporter gene, which allowed for lineage analysis in the mouse. Cre expression and ensuing recombination removed the STOP sequence to allow reporter gene expression not only in the cells in which recombination took place, but also in their progeny (Soriano, 1999). Furthermore, another degree of specificity was achieved by introducing a temporal control of Cre recombinase functional expression. A fusion protein was created between Cre and the ligand-binding domain of the estrogen receptor (Brocard et al., 1997; Danielian et al., 1998; Schwenk et al., 1998), which could only be activated upon addition of ligand. Moreover, the estrogen-binding domain had a point mutation that provided higher affinity for tamoxifen (a synthetic ligand) than to endogenous estrogen. The use of both steroid agonists that do not appreciably target endogenous steroid receptors and of ligand-binding domains responsive only to exogenously added ligand allowed for the precise temporal control of reporter gene expression in transgenic mice.

3.2.3 Results

In E9.5 embryos, NGN2 is expressed in both the dorsal neural tube (Fig. 2A, C; white arrowheads), and in a subset of migrating NCCs marked by Sox10 (Britsch et al., 2001; Kuhlbrodt et al., 1998) (Fig. 2A, C; white arrows). $NGN2^{+}Sox10^{+}$ NCCs were exclusively found in the dorsal part of the neural crest migration stream (Fig. 2B, arrows), and never detected near the sympathetic ganglia (SG) (Fig. 2B, C; open arrowheads). Similarly, examination of embryos from a conventional *Ngn2lacZ* knock-in line (Fode et al., 2000; Scardigli et al., 2001) at E10.5 revealed numerous $lacZ^{+}$ cells in the DRG (Fig. 2D, black arrowheads) but none in the vicinity of the SG (Fig. 2D, white arrowheads). These data could indicate that *Ngn2*-expressing cells never give rise to sympathetic neurons, or that *Ngn2* is very transiently expressed in precursors of such autonomic neurons. To resolve this issue, we permanently marked the progeny of *Ngn2*-expressing cells by injecting pregnant mothers carrying embryos from an *Ngn2-CreER*^{TM} *x Rosa-loxPSTOPloxP-lacZ* intercross with 4-OH Txf, at E9.0 (n=6 embryos) and E9.5 (n=15 embryos). In such embryos, a very small proportion (<5%) of $lacZ^{+}$ neural crest-derived cells were observed in the SG, identified by counter-staining with antibody to tyrosine hydroxylase (TH; Fig. 3C, D; arrows). Thus some *Ngn2*-expressing cells do appear to generate sympathetic neurons.

To determine whether the relative contribution of *Ngn2*-expressing cells to DRG vs. SG was quantitatively different from the NCC population as a whole, we performed a similar analysis on embryos containing a *Wnt1-CreER*^{TM} transgene (Danielian et al., 1998). *Wnt1* is expressed in the dorsal neural tube and roof-plate from the onset of neural

crest migration (Serbedzija and McMahon, 1997), and *Wnt1*-expressing NCCs populate all of the tissues derived from the neural crest (Jiang et al., 2000). Injections of embryos from *Wnt1-CreER*^{TM} x *Rosa-loxPSTOPloxP- lacZ* intercrosses with 4-OH Txf were performed at similar times to that in the *Ngn2-CreER*^{TM} experiments (E8.75-E9.0, and E9.5), and the embryos were analyzed at E12.5 (Fig. 3B).

To quantify the results, we counted the total number of lacZ^{+} cells in the DRG plus SG in every fourth section through the trunk region. There were many more (~18-fold) lacZ^{+} neural crest-derived cells in *Wnt1-CreER*^{TM}-expressing embryos, presumably reflecting both the expression of *Wnt1* in more cells than *Ngn2*, and at higher levels from the multicopy transgene. Therefore, we first normalized the data by calculating the percentage of labeled ganglionic cells in either the DRG or SG for each embryo, and then taking the mean of these values across all embryos of a given genotype (n=21 *Ngn2-CreER*^{TM} embryos; n=4 *Wnt1-CreER*^{TM} embryos). As expected, in *Wnt1-CreER*^{TM} embryos the average percent of cells in the DRG was ~5-fold higher than that in the SG (Table 1), reflecting the larger size of the sensory ganglia. Furthermore, this percentage was higher when 4-OH Txf injection was carried out at E9.5 (6.7-fold) than at E8.75-E9.0 (3.7-fold; Table 1), consistent with the fact that NCCs continue to populate the DRGs after they have stopped colonizing the SGs (Serbedzija et al., 1990; Weston and Butler, 1966).

Despite this overall DRG-bias of the bulk NCC population, however, in *Ngn2-CreER*^{TM} embryos the percentage of labeled cells in the DRG was ~20-fold higher than in the SG (Table 1). Thus, cells derived from *Ngn2*-expressing progenitors contributed to sensory ganglia rather than sympathetic ganglia at a frequency 3- to 4-fold higher than

that of *Wnt1*-expressing progenitors, at both early and late injection times (Table 1). To ensure that this difference was not an artifact of the smaller number of labeled cells in *Ngn2-CreER*^{TM} embryos (~67 cells per embryo in DRG+SG, vs. 1,258 cells per embryo in DRG+SG of *Wnt1-CreER*^{TM} embryos), we performed a Monte Carlo simulation to randomize the data, using the actual number of cells counted in each of the 21 embryos examined but assuming that the distribution of cells between the DRG and SG was actually the same as that measured in *Wnt1-CreER*^{TM} embryos (83% in DRG, 17% in SG, a ratio of 4.8 (Table1)). After 10,000 iterations of this simulation, the probability that the ratio [%cells in DRG]/[%in SG] of ~20 that we actually measured in *Ngn2-CreER*^{TM} embryos (Table1) was due to chance was < 0.001. Consistent with this, the percentage of *Wnt1*-derived cells in SG was ~3.5-fold higher than the percentage of *Ngn2*-derived cells in these autonomic ganglia. Thus, by several measures NCCs expressing *Ngn2* are more likely to colonize sensory ganglia than are those expressing *Wnt1*. This difference is likely to be an underestimate, since the *Wnt1*-expressing NCC population itself presumably contains *Ngn2*-expressing cells.

At the time of 4-OH-Txf injection, NGN2^{+} cells are located in both the dorsal and ventral neural tube, as well as in the neural crest (Fig. 2A-C). Strikingly, in embryos analyzed at E12.5 there were no labeled neurons in the dorsal neural tube; rather, they were found in neural crest derivatives and in the ventral neural tube (Fig. 2G, H). That reporter activation did occur in some dorsal neural tube cells at E9.5 is supported by the fact that lacZ^{+} cells could still be detected in this location in embryos analyzed at E10.5 (Fig. 2F), and by the persistence at E12.5 of a few labeled progenitor-like cells with an elongated morphology and endfoot in the dorsal ventricular zone (Fig. 2H, I, white

arrow). These data suggest that most $NGN2^+$ cells in the dorsal neural tube at E9.5 may be premigratory neural crest or dual crest-CNS progenitor cells (Bronner-Fraser and Fraser, 1988; Bronner-Fraser and Fraser, 1989). However we cannot formally exclude that some of these cells die, or migrate ventrally.

We next asked whether there was any bias in the differentiation of *Ngn2*-expressing cells to neuronal vs. glial fates in the sensory ganglia. To do this, we performed double-labeling for lacZ expression and the pan-neuronal markers Isl1 (Fig. 3E, F) or NeuN (Fig. 3G) to identify neurons, or anti-brain fatty acid binding protein (BFABP) antibody staining to identify peripheral glia (Britsch et al., 2001) (Fig. 3H, I). Double-labeled $lacZ^+Isl1^+$ (or $lacZ^+NeuN^+$) neurons were clearly distinguishable from $lacZ^+Isl1^-$ non-neuronal cells (Fig. 3E-G; arrows vs. arrowheads, respectively). Expression of lacZ in these latter, non-neuronal cells was co-localized with that of BFABP; this was particularly clear in the dorsal roots and peripheral nerve (Fig. 3H, I; arrows) where there are no neuronal cell bodies. Quantification indicated that the percentage of $lacZ^+$ neurons and non-neuronal DRG cells was statistically indistinguishable in embryos injected with 4-OH Txf at either E9.0 or E9.5 (Table 2). A similar result was obtained in Wnt1-CreER^{TM} embryos (52±3% $LacZ^+Isl^+$ and 48±3% $LacZ^+Isl^-$ cells; n = 2 embryos). Therefore, $Ngn2^+$ expressing cells give rise to both neurons and non neuronal cells in similar proportions, as does the bulk NCC population. These data indicate that while *Ngn2*-expressing NCCs are strongly biased towards generating sensory derivatives, they are not biased towards generating neuronal vs. glial derivatives.

Finally, we examined the distribution of *Ngn2*-derived cells among different sensory neuron subtypes by double-labeling using antibodies to the NGF receptor trkA, the BDNF receptor trkB, or the NT-3 receptor, trkC. These receptors are predominantly, although not exclusively, expressed by nociceptive, mechanoceptive, and proprioceptive sensory neurons, respectively, (reviewed in ref. Snider, 1994). We estimated that ~68% of $lacZ^+$ neurons in the DRG are $trkA^+$ (Fig. 4A-F, arrows; n=10 embryos), a number remarkably close to the percentage of all neurons that express the receptor at this stage (~70%) (Mu et al., 1993). A much smaller proportion of $lacZ^+$ cells appeared to be $trkB^+$ or $trkC^+$ (Fig. 4C-F, arrows), consistent with the fact that these receptors are expressed by a smaller proportion of all sensory neurons at this stage (Mu et al., 1993). Thus, these data suggest that *Ngn2*-expressing NCCs contribute to multiple classes of sensory neurons in DRG, without any apparent bias towards a particular subtype.

3.2.4 Discussion

Classical lineage tracing studies have left open the question of whether the premigratory and early migrating neural crest is a homogeneous population of pleuripotent cells, or whether it is heterogeneous and contains both pleuripotent and intrinsically fate-restricted cells. The results presented here identify at least one subpopulation of NCCs, namely that marked by expression of *Ngn2*, which exhibits a predictable bias in its differentiated fate from an early stage in migration. *Ngn2*-expressing cells are ~20 times more likely to generate sensory than autonomic (sympathetic) derivatives, and this bias is 3 to 4-fold greater than that of bulk NCCs

marked by expression of *Wnt1*. This conclusion is supported even when taking into account the fact that there are more labeled NCCs in *Wnt1-CreER*^{TM} embryos.

The fact that even a few *Ngn2*-expressing NCCs give rise to autonomic derivatives indicates that *Ngn2* does not mark inevitable commitment to a sensory fate, but rather a strong bias to such a fate. This bias may reflect an inherently probabilistic influence of *Ngn2* on sensory fate specification, or the influence of additional unknown but deterministic factors. For example, forced expression of *Ngns* promotes sensory differentiation both in vivo (Perez et al., 1999) and in cultured PNS progenitor cells, but in vitro this effect can be overridden by environmental signals that promote autonomic neurogenesis (Lo et al., 2002). These data suggest that there are other determinants of sensory identity required in addition to *Ngns* to commit cells to a sensory fate. Perhaps the small fraction of *Ngn2CreER*^{TM}-expressing NCCs that generated autonomic derivatives represents those cells in which such putative sensory identity co-determinants had not yet been expressed at the time of 4-OH Txf injection. Alternatively, commitment to a sensory fate may require prolonged or enhanced *Ngn2* expression, analogous to the commitment to a neural precursor fate of *Drosophila* epithelial cells expressing high levels of proneural genes (Cubas et al., 1991).

The identification of *Ngn2*-expressing cells as strongly biased to a sensory fate suggests that those marked NCCs that generated only sensory derivatives in the chick lineage tracing experiments (Bronner-Fraser and Fraser, 1988; Bronner-Fraser and Fraser, 1989; Serbedzija et al., 1994) may have been intrinsically restricted to this fate, and could be equivalent to or overlapping with the *Ngn2*-expressing population. The fact, moreover, that *Ngn2* is expressed in pre-migratory as well as in migrating neural crest

cells (Fig. 2A and refs. Ma et al., 1999; Sommer et al., 1996) further suggests that such a developmental bias may be acquired by NCCs at a relatively early stage in their ontogeny, perhaps even before they exit the neural tube (Fig. 5A). However, we cannot exclude that such premigratory $NGN2^+$ cells contribute most of the sympathetic progeny derived from the *Ngn2*-expressing population. Thus, the exact time and place at which restriction to a sensory fate is imposed remain to be determined. Whatever the case, our results suggest that from the earliest stages of migration, the neural crest is not a homogeneous population of pleuripotent cells, but rather contains deterministically distinct subpopulations with predictable fate biases.

Why should there be a subpopulation of *Ngn2*-expressing NCCs pre-specified for a sensory fate? In previous studies, we have shown that *Ngn2* is required in a subpopulation of early differentiating sensory precursors that preferentially generate large-diameter sensory neurons (Frank and Sanes, 1991), including the trkC+ and trkB+ subclasses (Ma et al., 1999). In contrast, *Ngn1* is required by a different, later-differentiating subpopulation of sensory precursors that generates predominantly smaller-diameter, trkA+ neurons (Frank and Sanes, 1991; Ma et al., 1999). The early sensory fate-specification of *Ngn2*-expressing NCCs may reflect a need to ensure rapid differentiation of these cells at the earliest stages of sensory ganglion formation. At apparent odds with this view is our observation that the majority of sensory neurons derived from *Ngn2*-expressing cells are trkA+. However, this may simply reflect the fact that *Ngn2* is also expressed by the later-differentiating, *Ngn1*-dependent subpopulation, even though these precursors do not require *Ngn2* function (Ma et al., 1999). Consistent with this idea, there is evidence for cross-activation of *Ngn2* expression by *Ngn1* in the

PNS (Ma et al., 1998). Because the *Ngn1*-dependent precursors constitute the majority of the DRG precursor population, and because these precursors can compensate for the defective differentiation of *Ngn2*-dependent precursors in *Ngn2-/-* mutants (Ma et al., 1999), it is highly unlikely that any differences in the fate of *Ngn2-CreERTM*-expressing cells would be detectable in an *Ngn2-/-* background.

Our results further suggest that *Ngn2*-expressing NCCs are not committed to a neuronal fate, despite their bias for a sensory identity. Consistent with this conclusion, CNS precursors isolated on the basis of *Ngn2* expression can generate glia as well as neurons in vitro (Nieto, 2001), although it was not determined whether these cells are restricted to generating specific neuronal subtypes. Previously, we showed that post-migratory neural crest stem cells in the sciatic nerve are restricted to autonomic fates but can still generate both neurons and glia (White et al., 2001). Those studies, however, left open the converse question of whether all sensory progenitors also generate autonomic derivatives, or whether there exists a complementary population of sensory-restricted multipotent neuro-glial progenitors. In vitro studies have provided evidence of sensory-restricted progenitors, but whether these cells also generated glia could not be determined (Greenwood et al., 1999). The present results indicate that sensory-restricted precursors generate neurons and glia with equal probabilities, in a similar fashion as the bulk neural crest population marked by *Wnt1* expression. Taken together, these data imply that, at least in the PNS, multipotential neural progenitors become specified for certain aspects of neuronal subtype identity (Fig. 5B, P_S and P_A) before they have become committed to neuronal or glial fates.

This conclusion challenges current models of neural cell lineage diversification, in which progenitors are typically depicted as first committing to neuronal or glial fates (Fig. 5C, P_N and P_G), before acquiring particular neuronal subtype identities (reviewed in refs. Anderson, 2001; Gage, 1998). Evidence for such restricted neuronal and glial precursors has been provided by in vitro studies of embryonic spinal cord-derived progenitor cells (Kalyani et al., 1998; Mayer-Proschel et al., 1997; Rao and Mayer-Proschel, 1997; Rao et al., 1998). However, recent studies suggest that in vivo, spinal cord progenitors are specified for particular neuronal and glial subtype identities before choosing between neuronal and glial fates (Lu et al., 2002; Zhou and Anderson, 2002). Thus, while some finer aspects of neuronal subtype identity may not be acquired until after precursors commit to a neuronal fate, at least certain broad subclasses (e.g., sensory vs. autonomic; motoneuron vs. interneuron) are specified while progenitors still possess both neuronal and glial capacities. Such a decision logic could reflect the dual role of bHLH proneural genes in controlling the neuron vs. glial fate decision (reviewed in ref. Vetter, 2001), and in neuronal identity determination (e.g., ref. Gowan et al., 2001). The need to coordinate the latter role with other transcriptional programs that specify neuronal identity (Jessell, 2000; Scardigli et al., 2001) may necessitate that such programs are established at the time proneural genes are first expressed, when neural progenitors are competent for, but not committed to, a neuronal fate (Lo et al., 1997).

Table 3-1 Relative contributions of *Ngn2*- and *Wnt1*-derived neural crest cells to sensory and sympathetic ganglia.

Injxn stage	Rel. DRG col.: Ngn2-CreER[a]	Rel. DRG col.: Wnt1-CreER[a]	DRG Ratio [Ngn2/Wnt1][b]
E8.75-E9.0	11.2	3.7	3.0
E9.5	27.6	6.7	4.1
Combined	19.5	4.8	4.0

The percentage of cells in the DRG or SG was first calculated for each embryo as [(# cells in DRG)/(# cells in DRG + # cells in SG)] x 100%, or [(# cells in SG)/(# cells in DRG + # cells in SG)] x 100%, respectively. The mean of this percentage was then calculated for all embryos of a given genotype. The mean percentage of cells in the DRG, or in the SG, in *Ngn2-CreER* embryos (n=21) was statistically significantly different from that in *Wnt1-CreER* embryos (n=4) at all stages ($p<0.01$). [a]The relative colonization of the DRG compared to the SG was then calculated as the [mean % of cells in DRG/ mean % of cells in the SG]. [b]The relative preference of Ngn2-derived cells to colonize the DRG compared to Wnt1-derived cells is calculated as the ratio of the numbers in the first two columns.

Table 3-2 Relative contribution of *Ngn2*-derived neural crest cells to neuronal vs. non-neuronal derivatives in sensory ganglia.

Stage of injxn **(no. of embryos)**	Neurons	Non-neuronal cells
E9 (n=5)	56±6%[a]	44±6%[a]
E9.5 (n=11)	52±5%	48±5%
combined	53±4%	47±4%

The percentage of lacZ-labeled cells that were neurons ($Isl1^+$) or non-neuronal cells ($Isl1^-$) was measured in the DRG of *Ngn2-CreER*^{TM} embryos injected at the indicated stages; the numbers represent the mean±SEM. [a]There was no statistically significant difference in the percentage of neurons vs. non-neuronal cells at either stage. Similar results were obtained in *Wnt1-CreER*^{TM} embryos (see text).

3.2.5 Figure legends

Figure 3-1 Mapping the fate of *Ngn2*-expressing cells.

(A) *Ngn2* is required for the development of a subset of sensory neurons (N_S) (Fode et al., 1998; Ma et al., 1999), but may be transiently expressed by progenitors with a broader developmental potential (examples illustrated in 1-4). N_a=autonomic neuron; G=glia. (B) Strategy for inducible activation of the *Rosa26-lacZ* reporter gene by 4-OH Txf-inducible Cre expressed from the *Ngn2* locus. (C) Because 4-OH Txf has a half-life in vivo of only 0.5-2 hrs (Danielian et al., 1998), CreER^{TM} will be active in *Ngn2*-expressing cells only at the time of injection of the steroid ligand. The progeny of these cells can be detected by expression of *lacZ* at later times (E12.5; blue ovals). Other cells expressing *Ngn2* at these later times will not express *lacZ* due to the absence of 4-OH Txf (E12.5, white oval).

Figure 3-2 Fate of transiently *Ngn2*-expressing cells in the neural tube.

(A-C) Double-label immunofluorescence staining of a section through the caudal trunk region of an E9.5 mouse embryo with antibodies to NGN2 (A, green) and the pan-neural crest marker Sox10 (B, red). White arrowheads indicate NGN2$^+$Sox10$^-$ cells in the dorsal neural tube, open arrow indicates NGN2$^+$Sox10$^-$ motoneuron precursors in the ventral neural tube; white arrows indicate NGN2$^+$Sox10$^+$ neural crest cells in the dorsal migration pathway, open arrowheads indicate NGN2$^-$Sox10$^+$ cells migrating ventrally towards the dorsal aorta (da). (D, E) Analysis of conventional *Ngn2lacZ* knock-in embryos (Fode et al., 2000; Scardigli et al., 2001). (D) At E10.5 numerous lacZ$^+$ cells

are visible in the DRG (black arrowheads) and ventral nerve roots (arrows) but not in the region of the SG (white arrowheads). (E) By E12.5, expression in the DRG is no longer detectable (arrowheads) and there is extensive expression in the dorso-lateral neural tube (arrows). (F-J) Analysis of *Ngn2CreER^{TM}* x *RosaSTOPlacZ* embryos injected with 4-OH Txf at E9.5 and analyzed at E10.5 (F) or E12.5 (G-J). (F) Some *Ngn2-CreER^{TM}*-expressing cells were in the dorsal neural tube at the time of Cre activation (arrows). (G, H) By E12.5 there are very few lacZ^{+} cells in the dorsal neural tube and numerous labeled cells in DRG (white arrowhead) and dorsal roots (black arrowhead); contrast this with the pattern seen using the conventional *Ngn2lacZ* reporter (E). Arrow in (G) indicates motorneurons. (H-J) occasional elongated, lacZ^{+} cells with an endfoot in the ventricular zone are seen at E12.5 in the dorsal (H, I white arrow) and ventral (H, J black arrow) regions of the spinal cord; I, J are higher-magnification views of the dorsal and ventral regions of the section shown in (H). Such cells are not seen in conventional *Ngn2-lacZ* knock-in mice at the same age (cf. E vs. H).

Figure 3-3 Distribution of progeny derived from *Ngn2CreER^{TM}*-expressing NCCs in the PNS.

All panels are from *Ngn2-CreER^{TM}* x*RosaSTOPlacZ* embryos injected at E9.5 and analyzed at E12.5, except for (B) which is from a control *Wnt1-CreER^{TM}* x *RosaSTOPlacZ* embryo similarly injected and analyzed. Arrows in (A-B) indicate DRG, arrowheads SG. (C, D) Example of a rare section from *Ngn2-CreER^{TM}* embryos with a lacZ^{+} cell in the SG (arrow), visualized by counter-staining for TH (brown); (D) shows a higher-magnification view of the SG. Arrowhead indicates DRG. See Table 1 for quantification. (E-G) Double-labeling reveals that lacZ^{+} cells in the DRG include both

neuronal ($Isl1^+$ or $NeuN^+$, arrows) and non-neuronal ($Isl1^-$ or $NeuN^-$, arrowheads) cells. NT indicates neural tube. The magnification in (G) is slightly higher than in (F). (H, I) *Ngn2*-expressing NCCs generate $BFABP^+$ glial cells in dorsal root (large arrow) and peripheral nerve (H, small arrows) as well as in the DRG. Neurons and non-neuronal cells are generated in statistically indistinguishable proportions (see Table 2).

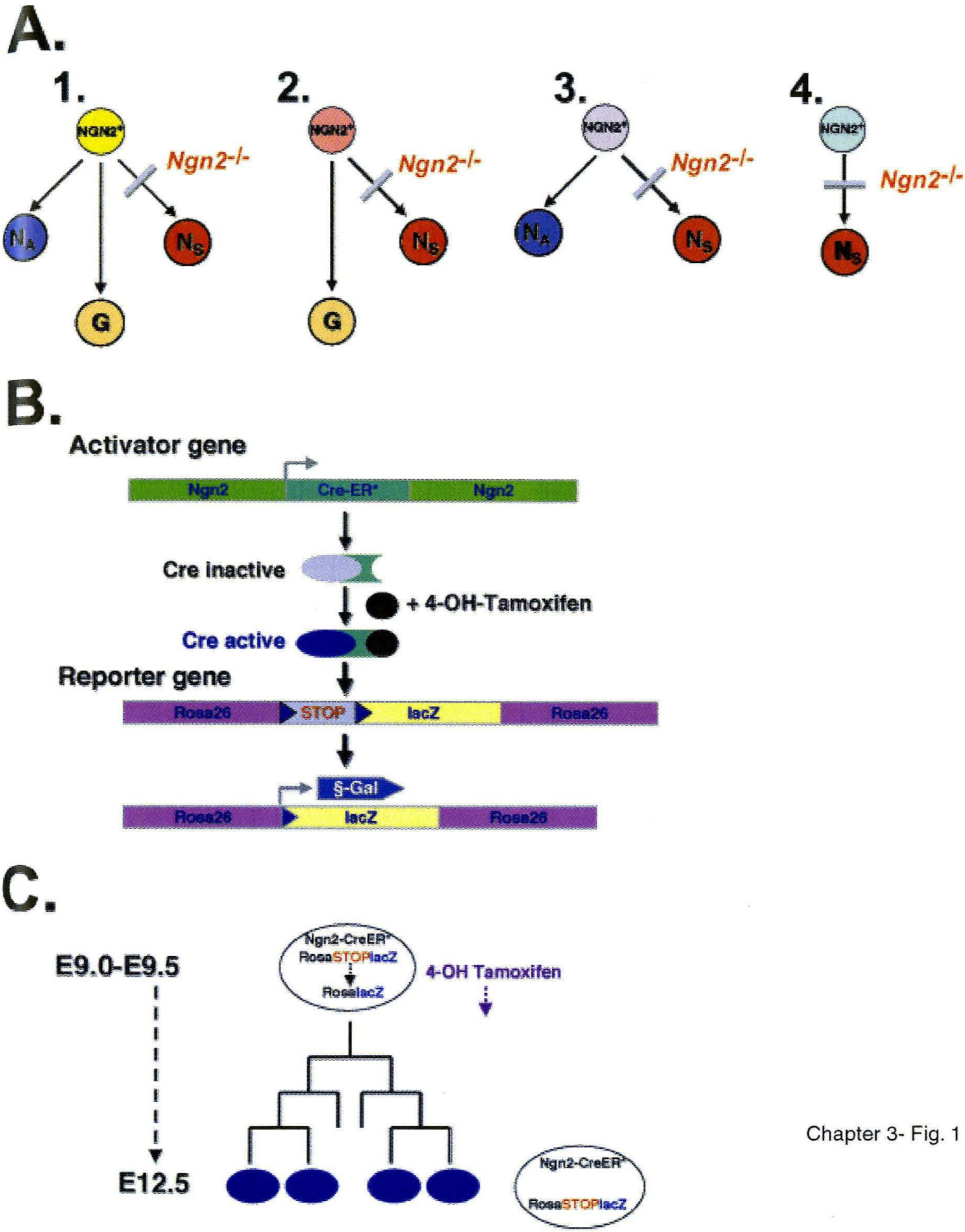
Figure 3-4 *Ngn2*-expressing NCCs generate multiple sensory neuron subtypes in the DRG.

Sections through DRG of E12.5 *Ngn2-CreER^{TM}* embryos injected with 4-OH Tamoxifen at E9.5. (A, B) Sections double-labeled to reveal $lacZ^+$ (blue) and $trkA^+$ (brown) cells at low (A) and high (B) magnification. Arrows indicate double-positive cells, arrowheads a $lacZ^+$, $trkA^-$ cell. (C, D) Double-labeling for $lacZ^+$ and $trkB^+$ cells. (E, F) Double-labeling for $lacZ^+$ and $trkC^+$ cells. In each case, analysis in multiple focal planes confirmed the co-localization of Xgal staining and anti-trk antibody staining.

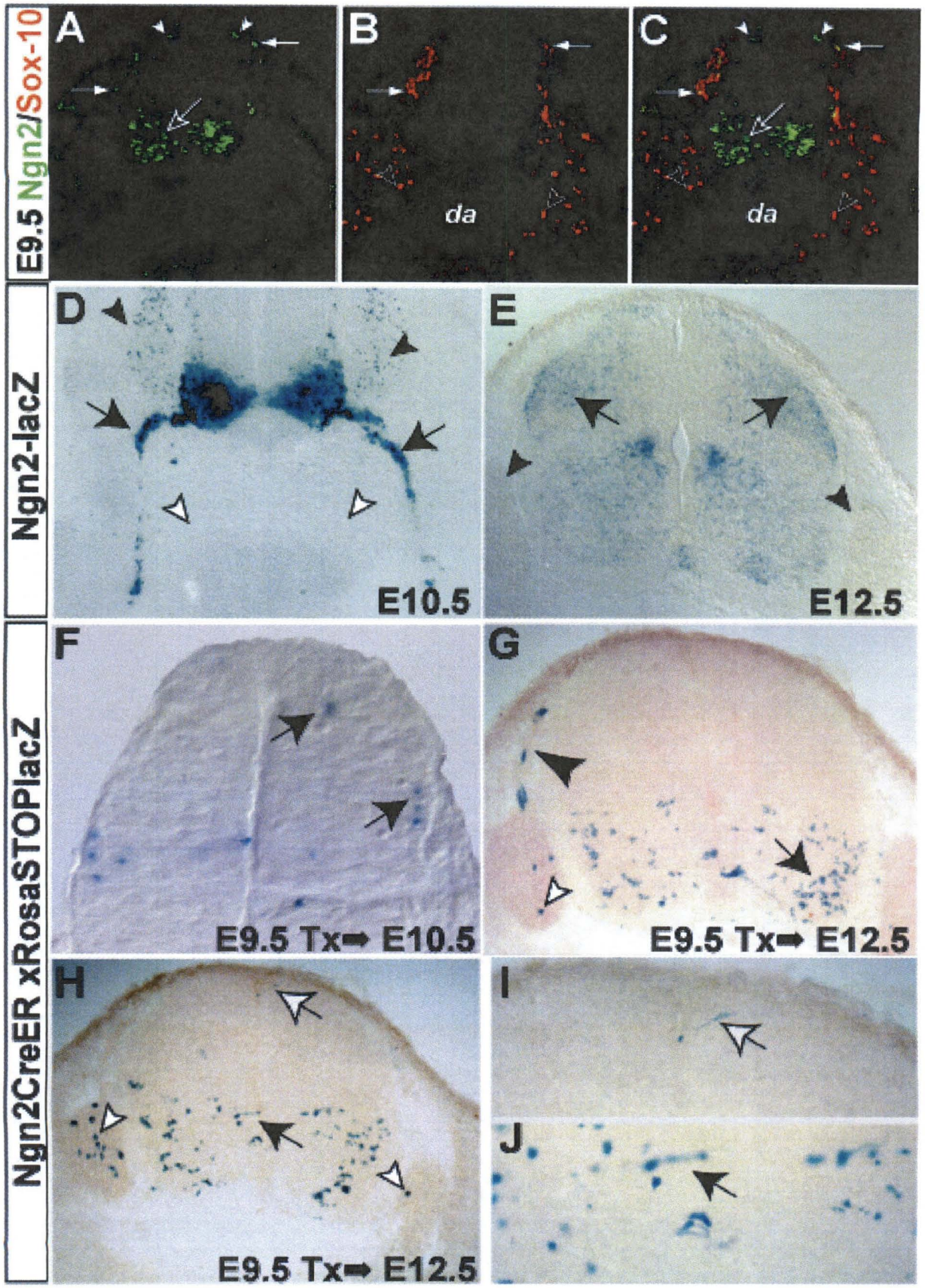
Figure 3-5 Summary diagram.

(A) Illustration of experimental results. Transient activation of Cre recombinase in *Ngn2*-expressing cells in the dorsal neural tube and/or migrating neural crest at E9.0 – 9.5 leads to progeny that by E12.5 have predominantly populated the DRG and surrounding nerve roots; both neurons (blue circles) and glia (blue diamonds) are found in this location. (B-D) Possible modes of neural crest lineage segregation. (B) Neural crest cells (NCC) first generate autonomic- and sensory-restricted progenitors (P_A and P_S, respectively). These progenitors then each generate neurons (N_A, N_S) and glia (G_A and G_S). (C) NCCs first generate neuronal- and glial-restricted progenitors (P_N and P_G,

respectively), which then differentiate into autonomic and sensory neurons or glia. (D) NCCs directly generate autonomic and sensory neurons and glia without restricted intermediates. (E) Wnt1 expression marks NCCs, while NGN2 is predominantly expressed by sensory-restricted progenitors, favoring the model in (B). Dotted arrow indicates that some $NGN2^{+}$ cells can still generate autonomic derivatives at low frequency.



Chapter 3- Fig. 1



Chapter 3- Fig. 2

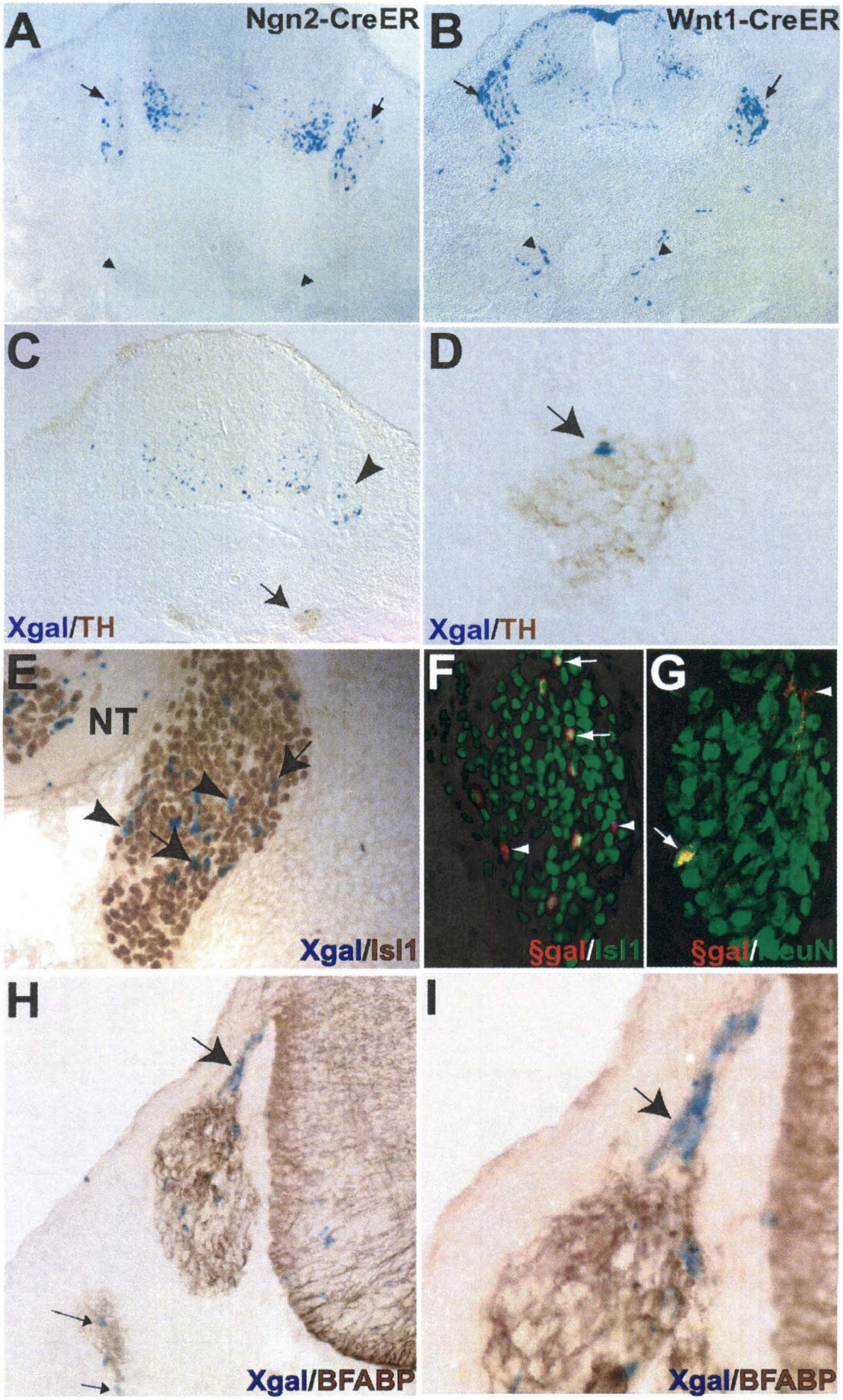
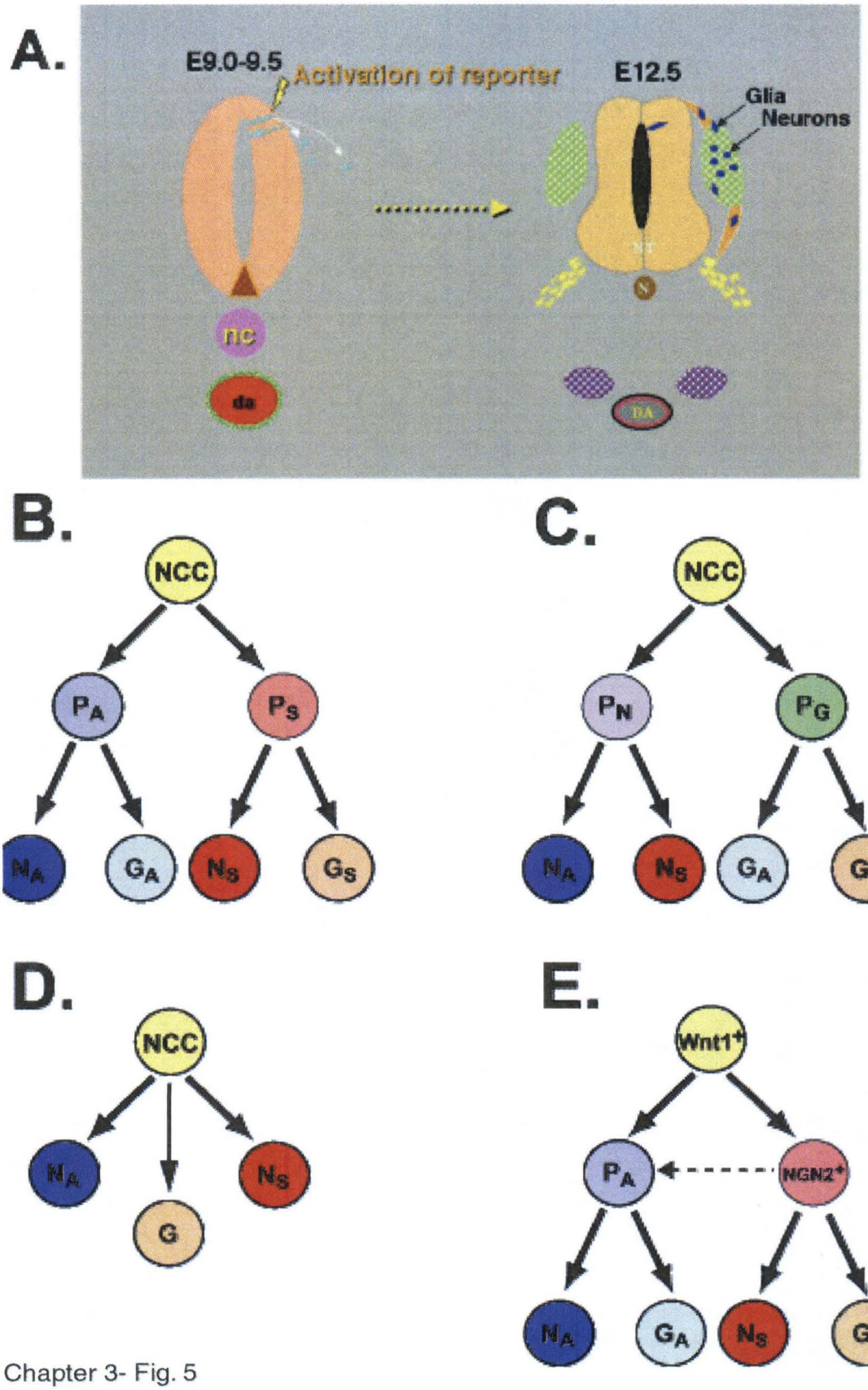

A
Ngn2-CreER
B
Wnt1-CreER
C
Xgal/TH
D
Xgal/TH
E
NT
Xgal/Isl1
F
§gal/Isl1
G
§gal/NeuN
H
Xgal/BFABP
I
Xgal/BFABP

Fig. 3

Chapter 3- Fig. 4



Chapter 3- Fig. 5

3.2.6 Experiments with Ngn2-CreER^{TM} adult mice

We designed these transgenic mice for two purposes originally. On the one hand, for embryonic studies like those described above, and on the other, for manipulations in adult mice with the hope to conditionally inhibit adult neurogenesis. This latter application failed because we were not able to induce lacZ expression in reporter mice when 4-OH Txf was injected in the adult. Since I had previously observed that some *Ngn2*-expressing cells in the subgranular zone of the adult dentate gyrus were dividing (not shown), the goal was to ablate these cells by crossing Ngn2-CreER^{TM} animals to transgenic mice expressing Cre-dependent thymidine kinase (analogous to reporter mice expressing Cre-dependent lacZ). For cell ablation, a drug, ganciclovir, must be provided, so that only dividing thymidine kinase-expressing cells undergo cell death because ganciclovir is turned into a toxic product by this enzyme (Heyman et al., 1989). Thus, in principle, we should have been able to temporally control (first by means of 4-OH Txf-Cre activation and then by ganciclovir treatment) the ablation of dividing, *Ngn2*-expressing cells. This could serve to study the role, if any, of adult neurogenesis (reviewed in Gould et al., 1999; Kempermann and Gage, 2000).

There are multiple reasons that could explain why reporter activation was not observed in the adult, even after multiple high doses of 4-OH Txf injections [18]. It is possible that the level of *Ngn2* produced in the adult (and thus, of CreER^{TM}) is simply not high enough to elicit Cre-mediated DNA recombination. Previous successful Cre-dependent DNA recombination in the adult was observed when multicopy transgenic promoters were driving Cre expression (for example, Brocard et al., 1997), but in general

[18] I administered ~ 5-10 times the dose reported in (Brocard et al., 1997) that obtained Cre activation in adult skin (see next).

these studies do not show recombination in brain (but see below). On the other hand, knocking-in of Cre recombinase induced DNA recombination in one case (Hebert and McConnell, 2000), but the enzyme was not fused to any regulatory domain, like the estrogen-responsive one utilized in our studies. Thus the strength of an unmodified recombinase may be higher than the fusion we used. In this regard, a recent report quantitated that tamoxifen-induced excision had a 5-10% efficiency rate compared to a general unmodified Cre deletor line (Guo et al., 2002). They also observed different extents of recombination in various tissues, of which muscle and pancreas had the most extensive, even though a general strong promoter (consisting of the CMV early enhancer and chicken β-actin promoter) was employed (Guo et al., 2002).

Alternatively, the degree of Cre activation by 4-OH Txf may have been insufficient. A number of different mutated estrogen responsive domains have been used in transgenic mice (Indra et al., 1999), with varying recombination efficiencies. The fusion used for the present study was of the "first generation" of CreER^{TM} fusions synthesized, before improved versions were produced by specific mutations in other positions of the ER domain.

There is at least one recent study reporting the successful application of the CreER^{TM}/loxP recombination technology in the brain (Weber et al., 2001), where a strong neuronal promoter (prion protein promoter) was employed. It is noted by the authors that strong positional effects were obtained in mice generated by pronuclear injection, since Cre expression varied extensively in different transgenic lines. In general, a complicating factor in the use of these technologies is the positional effect (Lacy et al.,

1983), which produces mosaicism in the target tissue or inappropriate transgene expression, depending on the site of transgene insertion in the genome.

Other attempts of inducible recombination in adult brain took advantage of Cre and progesterone receptor fusions. The principle is the same as in CreER^{TM} fusions, but activation of the recombinase is produced by the synthetic steroid RU486 instead. One study (Kellendonk et al., 1999) compared activation efficiencies of Cre fused with either estrogen or progesterone ligand binding domains *in vitro*, and concluded that the latter were more promising for use in brain. It is thus possible that Cre fusions with other ligand-activated domains, such as progesterone, are more effective for targeting transgene expression in brain. As examples, specific neuronal populations, such as cerebellar Purkinje (Kitayama et al., 2001) or granule (Tsujita et al., 1999) cells have been already targeted by RU486 administration in transgenic mice.

To summarize, either because of low transgene copy numbers driven by the NGN2 promoter, or because of low effective concentrations of active Cre (i.e. only partially activated by 4-OH Txf), or a combination of both, we could not observe lacZ induction in adult brains of Ngn2-CreER^{TM}, Rosa-loxPSTOPloxP-lacZ double transgenic mice.

4 CONCLUDING REMARKS

I have chosen to study neural diversity to uncover the molecular details of the nervous system. I have attacked the problem from many angles by applying a combination of techniques. Several conclusions can be drawn from this work, some of which can be better appreciated if we place it under perspective.

4.1 A technological revolution: microarrays and biological research

In principle, microarray technology has a tremendous potential to accelerate basic science research. It provides the opportunity to simultaneously scrutinize expression of thousands of genes under different behavioral or pathological conditions or in mutant backgrounds (Lockhart and Barlow, 2001). This has revolutionized biological research recently, as illustrated by the existence of >1600 citations of "microarray" in Pubmed since late 1995, when the first papers were published (e.g., Schena et al., 1995).

Some of the results presented here represent in a way "proof of principle" experiments. I foresee a near future in which these types of studies will be carried out systematically, with the help of automated systems. It is probably a good analogy to compare the leap accomplished in the field of genetics when automated sequencers were introduced to the revolution caused by the development of microarray technology. Obtaining sequence or expression data for a single gene used to involve plenty of time

and intensive work, whereas nowadays, collecting such data for thousands of genes at once is becoming customary in many (wealthy) labs and can be generated in a matter of days.

4.2 Application of microarray technology to the study of the brain

Despite the benefits of this recently developed technology, it was not *a priori* clear that it could be applied to the study of the nervous system, given the brain cellular complexity (Serafini, 1999). In addition, oligonucleotide microarray technology is confined to the interrogation of known genes or ESTs of reported sequences. Again, it was not clear either that the representation of genes on these types of arrays would comprise a comprehensive cohort of brain genes (Cao and Dulac, 2001). The results presented here show that microarray technology can indeed be applied to the study of brain regions. They highlight, though that a thorough validation of results is required by an independent method. *In situ* hybridization proved to be an adequate method of confirmation and also provided detailed anatomical information of gene expression at a high level of resolution, fundamental for the analysis of the brain.

The dilution of potentially enriched transcripts of low abundance among genes more broadly expressed is a serious concern. As discussed in chapter 2, starting the analysis with more homogeneous tissue (obtained by LCM), circumvented part of this problem. Several new genes were identified in my second screen that appeared not to be particularly enriched in the amygdala in the first screen, and were thus missed in the

original study. Perhaps the next level of refinement will require starting with single cells to further enrich for rare transcripts and solve the dilution problem.

It is also possible that the use of cDNA microarrays (which are not restricted to the interrogation of known genes or ESTs) augments the set of transcripts analyzed in brain profiling studies. However, as sequence databases are expanding continually, the representation of genes (and predicted genes) on updated oligonucleotide arrays may be appropriate enough for conclusive molecular analyses of brain regions.

4.3 Molecular signature of brain regions

In general, there were few genes (~ 0.3- 1.0 %) differentially expressed across different brain regions or within the amygdala. As already pointed out in the discussion in chapter 2, this is somewhat contrary to the intuitive idea that the complexity of the brain is due in part to the actions of genes differentially expressed. However, apart from the genes that are most likely missed due to the dilution problem or because they are not yet represented on the arrays, it is also possible that the molecular differences across brain regions are mainly based on different combinatorial collections of genes. Moreover, many genes could be differentially processed in each brain region (Cao and Dulac, 2001), which would be largely undetected by current oligonucleotide microarrays. Recent genome-wide analyses of alternative splicing indicate that 40-60% of human genes have alternative splice forms (Lander et al., 2001), resulting in an amazing potential functional diversity.

It will be very interesting to design splicing-specific microarrays (like the ones made for yeast (Clark et al., 2002)) for gene profiling in the brain. Another intriguing

possibility, so far largely unexplored, is that differentially transcribed non-coding RNA sequences (Eddy, 2002; Eddy, 1999) play a role in generating neural functional diversity.

4.4 Data analysis

As previously discussed, independent validation of microarray data was indispensable. In addition, the results obtained also underscore the need for careful data analysis for the selection of candidate genes with potential differential expression. The custom programs that we developed for this purpose are only examples of such tools. There are a number of other programs useful for microarray data analysis, such as clustering algorithms like Genecluster and Genespring, or other software, such as Bullfrog, (all described in Appendix 1). It is important to note that clustering algorithms only consider values of average differences and disregard all the other parameters calculated by the Affymetrix software. These programs can be useful when large collections of genes are analyzed and a global picture of expression data is sought. However, to find genes that are highly enriched in one region with respect to all others it is important to set specific criteria for enrichment, which cannot be done with clustering algorithms. Our program permits the explicit specification of multiple criteria for enrichment and also, in contrast to clustering algorithms, allows for specifying constraints about the ratio of expression in one brain region relative to all others. It is interesting that the principles behind our program are very similar to the ones implemented by the Bullfrog program, independently developed by David Lockhart (founding president of Affymetrix Inc.) and others.

4.5 From molecular boundaries to functional specialization

Although for the wrong reasons, Franz Gall suggested already in the eighteenth century that different regions of the brain are specialized to fulfill distinct and specific functions (Finger, 2000). For example, the amygdala is known to play a fundamental role in the processing of emotional stimuli (Davis, 1992; LeDoux, 1995). The seminal work of Brodmann at the beginning of the last century described different regions of the human brain based on cellular morphology and network anatomy (Kandel et al., 2000). Remarkably, it has become evident that many of these boundaries described by cytoarchitectonic criteria also mark rather clear functional boundaries as observed by electrophysiological, lesion and functional imaging studies. The so-called Brodmann map is commonly used nowadays, particularly in functional imaging studies, to refer to specific regions of the brain.

Strikingly, the studies presented in Chapters 1 and 2 showed that most boundaries of gene expression in the amygdala were concordant with subregional limits defined by cytoarchitectonic criteria. By extension, most brain regions may harbor genes whose expression domains are contained within subregion boundaries.

I propose that a further refinement in the delineation of brain structures can be obtained based on molecular signatures, as illustrated in Chapter 1. In other words, gene expression domains may uncover sub-boundaries not obviously marked by classical histological staining techniques, which may lead to the revelation of functional roles previously unnoticed.

In general, the use of molecular markers to delimit specific brain regions opens the possibility for studying the functional architecture of the brain at an unprecedented level of resolution. Specifically, region-enriched genes obtained using these techniques could potentially provide direct tools to experimentally manipulate and dissect specific brain functions, as discussed below, through the use of transgenic technologies.

4.6 Immediate applications of these studies

4.6.1 Identification of markers for transgenic studies:

The genes identified in these screens may prove to be useful for gene targeting experiments in transgenic mice. For instance, they could be used to drive expression of a reporter gene or axonal tracer. This could define the sites of expression of their gene products and also map axonal projections. Alternatively, a desired gene could be used to alter the physiology of specific cells or to simply ablate them, such as a 'silencing' gene that decreases neuronal excitability (Johns et al., 1999; Nadeau et al., 2000), or a toxic chemical or toxin (Hara et al., 2001; Nirenberg and Cepko, 1993; Nirenberg and Meister, 1997), respectively. This manipulation could be used to define the role that specific neurons play under different behaviors. As already discussed in chapter 2, combinations of "semi-specific" genes (having more than one restricted site of expression) could be employed to achieve selective targeting. For example, there are a number of binary systems that require the presence of two components to activate transcription of a heterologous gene. These include the Cre-loxP recombination system (reviewed in Chapter 3), the tetracycline (Gossen and Bujard, 1992; Gossen et al., 1995) and the

ecdysone (Christopherson et al., 1992; No et al., 1996) inducible systems (reviewed in Saez et al., 1997). In principle one could build a Venn diagram of gene expression in the brain and take advantage of transcripts that have overlapping expression domains in the region of interest to design transgenic mice. I have identified a number of pairs of amygdala-enriched genes that could be used for this purpose (see Table 4 Chapter 2). Along these lines it is important to mention the recent development of techniques for efficiently designing bacterial artificial chromosomes (BACs) for use in BAC transgenic mice (reviewed in Heintz, 2001). With this method, transgenic mice can be created faster than with conventional transgenic techniques. It will most likely prove useful for the genetic manipulation of genes or neurons under different behaviors once specific genes (or combination of semi-specific genes) are further identified.

4.6.2 Analysis of regulatory regions

Another interesting application of these studies is in the analysis of gene regulatory regions. Transcripts that are coexpressed in one brain region are likely to share regulatory sequences in their enhancer regions. Microarray data can be invaluable for this analysis since it can potentially inform of all the transcripts expressed in each brain area. By careful examination of sequences flanking the transcriptional initiation (or, if not available, translational start) of these genes, it may be possible to find conserved motifs in genes similarly expressed. This principle has been successfully applied in yeast (Pilpel et al., 2001; Tavazoie et al., 1999) and plants (Harmer et al., 2000). Regulatory sequences responsible for gene expression control during progression of the cell cycle or light-dark phases were identified in each case, respectively. Since the yeast and plant genomes are

more compact than the mouse's, it is easier to perform regulatory region analysis in these organisms, in which only ~1kb of upstream sequences are considered. In mouse it is necessary to analyze larger flanking regions, since enhancer sequences can be located several tens of kilobases upstream (or downstream) of the site of transcriptional initiation (Lewin, 2000). The task of identifying common motifs is thus computationally more intensive in higher organisms. Very few comparisons have been done to date. However, the job will be soon tremendously simplified with the public availability of the sequenced genome and microarray gene expression data. Identifying brain nucleus-specific regulatory regions would predict the identification of further genes bearing the same regulatory regions that may be expressed in the same area. In addition, these computational studies may help design promoter constructs for transgenic mice.

4.7 Mechanisms of generation of diverse neurons and glia

My thesis work also deals with the study of neural crest lineage diversification. Methodologically and intellectually, this work was rather different from the screens carried out to molecularly characterize the amygdala. In this case, a specific question was asked: is the neural crest a homogeneous cell population or rather, is it formed partly by cells with predictable fate biases? It was concluded in Chapter 3 that some neural crest cells are biased towards the generation of sensory rather than autonomic neurons in spite of producing neurons and glia indistinguishably. The demonstration of such diversification is important because it guides the way in which the problem of neural cell lineage segregation is further studied. Also, as mentioned in that chapter, together with

other recent work (Lu et al., 2002; White et al., 2001; Zhou and Anderson, 2002), this challenged the hitherto view that neurons and glia arise from separate neuronal or glial precursors. Thus it seems that general aspects of neuronal subtype identity (i.e. sensory vs. autonomic) can be acquired when progenitor cells are still capable of generating both neurons and glia.

4.8 Long-term Applications

The combination of hypotheses and techniques employed in this thesis will probably prove a very robust method for the study of nervous system development and function. For example, microarrays could uncover the molecular signature of not only every brain region, but also of every neural cell population during development, provided fine dissecting techniques, such as LCM, permit the isolation of specific cell types. One could thus effectively obtain a molecular brain atlas of the developing and adult brain, such that the information generated could serve to study problems at the level of detail similar to the work described in Chapter 3. There is evidence that the analysis of transcripts differentially expressed by specific neuronal populations during neurogenesis can shed light on their development (Kuhar et al., 1993). Specifically, this work nicely shows how the identification of molecular markers uncovered phases in cerebellar histogenesis that were previously unrecognized.

Last, the examples presented here pave the way for conducting hypothesis-driven approaches to global profiling of gene expression that will complement the findings obtained by more general screens undertaken with no preconceptions. Hence, broad, systems neuroscience questions will be finally addressed at molecular resolution.

5 REFERENCES

Abel, T., Nguyen, P. V., Barad, M., Deuel, T. A., Kandel, E. R., and Bourtchouladze, R. (1997). Genetic demonstration of a role for PKA in the late phase of LTP and in hippocampus-based long-term memory. Cell 88, 615-26.

Abu-Abed, S., MacLean, G., Fraulob, V., Chambon, P., Petkovich, M., and Dolle, P. (2002). Differential expression of the retinoic acid-metabolizing enzymes CYP26A1 and CYP26B1 during murine organogenesis. Mech Dev 110, 173-7.

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature 372, 669-672.

Albright, T. D., Jessell, T. M., Kandel, E. R., and Posner, M. I. (2000). Neural Science: A century of progress and the mysteries that remain. Cell 100, S1-S55.

Anderson, D. J. (2001). Stem cells and pattern formation in the nervous system: the possible versus the actual. Neuron 30, 19-35.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25-9.

Barbu, M., Ziller, C., Rong, P. M., and Le Douarin, N. M. (1986). Heterogeneity in migrating neural crest cells revealed by a monoclonal antibody. Journal of Neuroscience 6, 2215-2225.

Beddington, R. S., Morgernstern, J., Land, H., and Hogan, A. (1989). An in situ transgenic enzyme marker for the midgestation mouse embryo and the visualization of inner cell mass clones during early organogenesis. Development 106, 37-46.

Blanchard, D. C., and Blanchard, R. J. (1972). Innate and conditional reactions to threat in rats with amygdaloid lesions. J. Comp. Physiol. Psychol. 81, 281-290.

Britsch, S., Goerich, D. E., Riethmacher, D., Peirano, R. I., Rossner, M., Nave, K. A., Birchmeier, C., and Wegner, M. (2001). The transcription factor Sox10 is a key regulator of peripheral glial development. Genes Dev. 15, 66-78.

Brocard, J., Warot, X., Wendling, O., Messaddeq, N., Vonesch, J. L., Chambon, P., and Metzger, D. (1997). Spatio-temporally controlled site-specific somatic mutagenesis in the mouse. Proc. Natl. Acad. Sci. USA 94, 14559-63.

Bronner-Fraser, M., and Fraser, S. (1988). Cell lineage analysis shows multipotentiality of some avian neural crest cells. Nature 335, 161-164.

Bronner-Fraser, M., and Fraser, S. (1989). Developmental potential of avian trunk neural crest cells in situ. Neuron 3, 755-766.

Bronner-Fraser, M., Stern, C. D., and Fraser, S. (1991). Analysis of neural crest cell lineage and migration. J Craniofac Genet Dev Biol 11, 214-22.

Brown, M. P., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., Jr., and Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc. Natl. Acad. Sci. USA 97, 262-7.

Brown, P., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nature Genetics 21, 33-37.

Caffe, A., and Vanleeuwen, F. (1983). Vasopressin-immunoreactive cells in the dorsomedial hypothalamic region, medial amygdaloid nucleus and locus coeruleus of the rat. Cell and Tissue Research 233, 23-33.

Canteras, N. S., Simerly, R. B., and Swanson, L. W. (1994). Organization of projections from the ventromedial nucleus of the hypothalamus: a Phaseolus vulgaris-leucoagglutinin study in the rat. J Comp Neurol 348, 41-79.

Cao, Y., and Dulac, C. (2001). Profiling brain transcription: neurons learn a lesson from yeast. Curr. Opin. Neurobiol. 11, 615-20.

Chemelli, R. M., Willie, J. T., Sinton, C. M., Elmquist, J. K., Scammell, T., Lee, C., Richardson, J. A., Williams, S. C., Xiong, Y., Kisanuki, Y., Fitch, T. E., Nakazato, M., Hammer, R. E., Saper, C. B., and Yanagisawa, M. (1999). Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. Cell 98, 437-51.

Christopherson, K. S., Mark, M. R., Bajaj, V., and Godowski, P. J. (1992). Ecdysteroid-dependent regulation of genes in mammalian cells by a Drosophila ecdysone receptor and chimeric transactivators. Proc Natl Acad Sci U S A 89, 6314-8.

Clark, G. A. (1995). Emotional learning. Fear and loathing in the amygdala. Curr. Biol. 5, 246-8.

Clark, T. A., Sugnet, C. W., and Ares, J. M. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296, 907-910.

Cubas, P., de Celis, J.-F., Campuzano, S., and Modolell, J. (1991). Proneural clusters of achaete-scute expression and the generation of sensory organs in the Drosophila imaginal wing disc. Genes & Dev. 5, 996-1008.

Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K., and McMahon, A. P. (1998). Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. Curr. Biol. 8, 1323-1326.

Dasilva, S., Cox, J., Jonk, L., Kruijer, W., and Burbach, J. (1995). Localization of transcripts of the related nuclear orphan receptors Coup-TF-I and Arp-1 in the adult mouse brain. Molecular Brain Research 30, 131-136.

Davis, M. (1992). The role of the amygdala in fear and anxiety. Annu. Rev. Neurosci. 15, 353-375.

Durbec, P., and Rougon, G. (2001). Transplantation of mammalian olfactory progenitors into chick hosts reveals migration and differentiation potentials dependent on cell commitment. Mol Cell Neurosci 17, 561-76.

Eddy, S. R. (2002). Computational genomics of noncoding RNA genes. Cell 109, 137-140.

Eddy, S. R. (1999). Noncoding RNA genes. Curr Opin Genet Dev 9, 695-9.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95, 14863-8.

Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A., and Liotta, L. A. (1996). Laser capture microdissection. Science 274, 998-1001.

Esclapez, M., Tillakaratne, N. J., Tobin, A. J., and Houser, C. R. (1993). Comparative localization of mRNAs encoding two forms of glutamic acid decarboxylase with nonradioactive in situ hybridization methods. J. Comp. Neurol. 331, 339-62.

Finger, S. (2000). Minds behind the brain. A history of the pioneers and their discoveries (New York: Oxford University Press).

Flint, J. (1997). Freeze! Nat Genet 17, 250-1.

Fode, C., Gradwohl, G., Morin, X., Dierich, A., LeMeur, M., Goridis, C., and Guillemot, F. (1998). The bHLH protein NEUROGENIN 2 is a determination factor for epibranchial placode-derived sensory neurons. Neuron 20, 483-494.

Fode, C., Ma, Q., Casarosa, S., Ang, S. L., Anderson, D. J., and Guillemot, F. (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. Genes Dev. 14, 67-80.

Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. Science 251, 767-73.

Frank, E., and Sanes, J. R. (1991). Lineage of neurons and glia in chick dorsal root ganglia: analysis in vivo with a recombinant retrovirus. Development 111, 895-908.

Franklin, K., and Watson, G. (1997). The Mouse Brain in Stereotaxic Coordinates: Academic Press).

Fraser, S. E., and Bronner-Fraser, M. E. (1991). Migrating neural crest cells in the trunk of the avian embryo are multipotent. Development 112, 913-920.

Gage, F. (2000). Mammalian neural stem cells. Science 287, 1433-1438.

Gage, F. H. (1998). Stem cells of the central nervous system. Curr. Op. Neuro. 8, 671-676.

Gallagher, M., and Chiba, A. A. (1996). The amygdala and emotion. Current Opinion in Neurobiology 6, 221-227.

Gautvik, K., deLecea, L., Gautvik, V., Danielson, P., Tranque, P., Dopazo, A., Bloom, F., and Sutcliffe, J. (1996). Overview of the most prevalent hypothalamus-specific mRNAs, as identified by directional tag PCR subtraction. Proc. Natl. Acad. Sci. U.S.A. 93, 8733-8738.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. USA 89, 5547-5551.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science 268, 1766-9.

Gould, E., Tanapat, P., Hastings, N. B., and Shors, T. J. (1999). Neurogenesis in adulthood: a possible role in learning. Trends Cogn Sci 3, 186-192.

Gowan, K., Helms, A. W., Hunsaker, T. L., Collisson, T., Ebert, P. J., Odom, R., and Johnson, J. E. (2001). Cross-inhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. Neuron 31, 219-232.

Gradwohl, G., Fode, C., and Guillemot, F. (1996). Restricted expression of a novel murine atonal-related bHLH protein in undifferentiated neural precursors. Dev. Biol. 180, 227-241.

Grattan, D. R. (2001). The actions of prolactin in the brain during pregnancy and lactation. Prog Brain Res 133, 153-71.

Greenwood, A. L., Turner, E. E., and Anderson, D. J. (1999). Identification of dividing, determined sensory neuron precursors in the mammalian neural crest. Development 126, 3545-3559.

Grigoriou, M., Tucker, A., Sharpe, P., and Pachnis, V. (1998). Expression and regulation of Lhx6 and Lhx7, a novel subfamily of LIM homeodomain encoding genes, suggests a role in mammalian head development. Development 125, 2063-2074.

Guo, C., Yang, W., and Lobe, C. G. (2002). A Cre recombinase transgene with mosaic, widespread tamosxifen-inducible action. Genesis 32, 8-18.

Hara, J., Beuckmann, C. T., Nambu, T., Willie, J. T., Chemelli, R. M., Sinton, C. M., Sugiyama, F., Yagami, K., Goto, K., Yanagisawa, M., and Sakurai, T. (2001). Genetic ablation of orexin neurons in mice results in narcolepsy, hypophagia, and obesity. Neuron 30, 345-54.

Harmer, S. L., Hogenesch, J. B., Straume, M., Chang, H. S., Han, B., Zhu, T., Wang, X., Kreps, J. A., and Kay, S. A. (2000). Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science 290, 2110-3.

Hebert, J. M., and McConnell, S. K. (2000). Targeting of cre to the Foxg1 (BF-1) locus mediates loxP recombination in the telencephalon and other developing head structures. Dev Biol 222, 296-306.

Heintz, N. (2001). BAC to the future: the use of bac transgenic mice for neuroscience research. Nat Rev Neurosci 2, 861-70.

Henion, P. D., and Weston, J. A. (1997). Timing and pattern of cell fate restrictions in the neural crest lineage. Development 124, 4351-4359.

Henrique, D., Adam, J., Myat, A., Chitnis, A., Lewis, J., and Ish-Horowicz, D. (1995). Expression of a Delta homologue in prospective neurons in the chick. Nature 375, 787-790.

Heyman, R. A., Borrelli, E., Lesley, J., Anderson, D., Richman, D. D., Baird, S. M., Hyman, R., and Evans, R. M. (1989). Thymidine kinase obliteration: creation of transgenic mice with controlled immune deficiency. Proc Natl Acad Sci U S A 86, 2698-702.

Hokfelt, T., Martensson, R., Bjorklund, A., Kleinau, S., and Goldstein, M. (1984). Distributional maps of tyrosine-hydroxilase immunoreactive neurons in the rat brain. In Handbook of Chemical Neuroanatomy, A. Bjorklund and T. Hokfelt, eds.: Elsevier), pp. 277-379.

Horseman, N. D., Zhao, W., Montecino-Rodriguez, E., Tanaka, M., Nakashima, K., Engle, S. J., Smith, F., Markoff, E., and Dorshkind, K. (1997). Defective mammopoiesis, but normal hematopoiesis, in mice with a targeted disruption of the prolactin gene. Embo J 16, 6926-35.

Indra, A. K., Warot, X., Brocard, J., Bornert, J. M., Xiao, J. H., Chambon, P., and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. Nucleic Acids Res 27, 4324-7.

Jessell, T. M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nature Reviews Genetics 1, 20-29.

Jiang, X., Rowitch, D. H., Soriano, P., McMahon, A. P., and Sucov, H. M. (2000). Fate of the mammalian cardiac neural crest. Development 127, 1607-16.

Johns, D. C., Marx, R., Mains, R. E., O'Rourke, B., and Marban, E. (1999). Inducible genetic suppression of neuronal excitability. J Neurosci 19, 1691-7.

Kacharmina, J. E., Crino, P. B., and Eberwine, J. (1999). Preparation of cDNA from single cells and subcellular regions. Methods Enzymol 303, 3-18.

Kalyani, A. J., Piper, D., Mujtaba, T., Lucero, M. T., and Rao, M. S. (1998). Spinal cord neuronal precursors generate multiple neuronal phenotypes in culture. J. Neurosci. 18, 7856-7868.

Kandel, E., Schwartz, J., and Jessell, T. (2000). Principles of Neural Science, Fourth Edition Edition (McGraw-Hill).

Kellendonk, C., Tronche, F., Casanova, E., Anlag, K., Opherk, C., and Schutz, G. (1999). Inducible site-specific recombination in the brain. J Mol Biol 285, 175-82.

Kempermann, G., and Gage, F. H. (2000). Neurogenesis in the adult hippocampus. Novartis Found Symp 231, 220-35.

Kitayama, K., Abe, M., Kakizaki, T., Honma, D., Natsume, R., Fukaya, M., Watanabe, M., Miyazaki, J., Mishina, M., and Sakimura, K. (2001). Purkinje cell-specific and inducible gene recombination system generated from C57BL/6 mouse ES cells. Biochem Biophys Res Commun 281, 1134-40.

Kluver, H., and Bucy, P. C. (1939). Preliminary analysis of the temporal lobes in monkeys. Arch. Neurol. Psychiatry 42, 979-1000.

Kuhar, S. G., Feng, L., Vidan, S., Ross, M. E., Hatten, M. E., and Heintz, N. (1993). Changing patterns of gene expression define four stages of cerebellar granule neuron differentiation. Development 117, 97-104.

Kuhlbrodt, K., Herbarth, B., Sock, E., Hermans-Borgmeyer, I., and Wegner, M. (1998). Sox10, a novel transcriptional modulator in glial cells. J. Neurosci. 18, 237-250.

Lacy, E., Roberts, S., Evans, E. P., Burtenshaw, M. D., and Costantini, F. D. (1983). A foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and expression in inappropriate tissues. Cell 34, 343-58.

Lakso, M., Sauer, B., Mosinger, B., Jr., Lee, E. J., Manning, R. W., Yu, S.-H., Mulder, K. L., and Westphal, H. (1992). Targeted oncogene activation by site-specific recombination in transgenic mice. Proc. Natl. Acad. Sci. USA 89, 6232-6236.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Le Douarin, N. M., and Kalcheim, C. (1999). The Neural Crest, 2nd Edition (Cambridge: Cambridge University Press).

LeDoux, J. E. (1994). Emotion, memory and the brain. Scientific American June, 50-57.

LeDoux, J. E. (1995). Emotion: Clues from the brain. Annu. Rev. Psychol. 46, 209-235.

Lee, C. K., Klopp, R. G., Weindruch, R., and Prolla, T. A. (1999). Gene expression profile of aging and its retardation by caloric restriction. Science 285, 1390-3.

Lee, J. E., Hollenberg, S. M., Snider, L., Turner, D. L., Lipnick, N., and Weintraub, H. (1995). Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. Science 268, 836-844.

Lewin. (2000). Genes VII: (Oxford University Press).

Lipshutz, R., Fodor, S., Gingeras, T., and Lockhart, D. (1999). High density synthetic oligonucleotide arrays. Nature Genetics 21, 20-24.

Lo, L., Sommer, L., and Anderson, D. J. (1997). MASH1 maintains competence for BMP2-induced neuronal differentiation in post-migratory neural crest cells. Curr. Biol. 7, 440-450.

Lo, L. C., Dormand, E. L., Greenwood, A., and Anderson, D. J. (2002). In Press. Development.

Lockhart, D., Dong, H., Byrne, M., Follettie, M., Gallo, M., Chee, M., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnology 14, 1675-1680.

Lockhart, D. J., and Barlow, C. (2001). Expressing what's on your mind: DNA arrays and the brain. Nat Rev Neurosci 2, 63-8.

Long, A. D., Mangalam, H. J., Chan, B. Y., Tolleri, L., Hatfield, G. W., and Baldi, P. (2001). Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12. J Biol Chem 276, 19937-44.

Lu, R., Sun, T., Zhu, Z., Ma, N., Garcia, M., Stiles, C. D., and Rowitch, D. H. (2002). Common Developmental Requirement for Olig Function Indicates a Motor Neuron/Oligodendrocyte Connection. Cell 109, 75-86.

Luo, L., Salunga, R. C., Guo, H., Bittner, A., Joy, K. C., Galindo, J. E., Xiao, H., Rogers, K. E., Wan, J. S., Jackson, M. R., and Erlander, M. G. (1999). Gene expression profiles of laser-captured adjacent neuronal subtypes. Nat Med 5, 117-22.

Ma, Q., Chen, Z. F., Barrantes, I. B., de la Pompa, J. L., and Anderson, D. J. (1998). Neurogenin 1 is essential for the determination of neuronal precursors for proximal cranial sensory ganglia. Neuron 20, 469-482.

Ma, Q., Fode, C., Guillemot, F., and Anderson, D. J. (1999). NEUROGENIN1 and NEUROGENIN2 control two distinct waves of neurogenesis in developing dorsal root ganglia. Genes & Dev. 13, 1717-1728.

MacLean, G., Abu-Abed, S., Dolle, P., Tahayato, A., Chambon, P., and Petkovich, M. (2001). Cloning of a novel retinoic-acid metabolizing cytochrome P450, Cyp26B1, and comparative expression analysis with Cyp26A1 during early murine development. Mech Dev 107, 195-201.

Mahadevappa, M., and Warrington, J. A. (1999). A high-density probe array sample preparation method using 10- to 100-fold fewer cells. Nat Biotechnol 17, 1134-6.

Maren, S., and Fanselow, M. S. (1996). The amygdala and fear conditioning: has the nut been cracked? Neuron 16, 237-40.

Mayer-Proschel, M., Kalyani, A. J., Mujtaba, T., and Rao, M. S. (1997). Isolation of lineage-restricted neuronal precursors from multipotent neuroepithelial stem cells. Neuron 19, 773-85.

Mazella, J., Botto, J. M., Guillemare, E., Coppola, T., Sarret, P., and Vincent, J. P. (1996). Structure, functional expression, and cerebral localization of the levocabastine-sensitive neurotensin/neuromedin N receptor from mouse brain. J Neurosci 16, 5613-20.

McDonald, A. (1992). Cell Types and Intrinsic Connections of the Amygdala. In The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction, J. Aggleton, ed. (Wiley-Liss, Inc), pp. 67-96.

Merali, Z., McIntosh, J., and Anisman, H. (1999). Role of bombesin-related peptides in the control of food intake. Neuropeptides 33, 376-86.

Merali, Z., McIntosh, J., Kent, P., Michaud, D., and Anisman, H. (1998). Aversive and appetitive events evoke the release of corticotropin-releasing hormone and bombesin-like peptides at the central nucleus of the amygdala. J Neurosci 18, 4758-66.

Metzger, D., Clifford, J., Chiba, H., and Chambon, P. (1995). Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. Proc. Natl. Acad. Sci. USA 92, 6991-6995.

Milner, R. J., and Sutcliffe, J. G. (1983). Gene expression in rat brain. Nucleic Acids Res 11, 5497-520.

Mirnics, K., Middleton, F. A., Marquez, A., Lewis, D. A., and Levitt, P. (2000). Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. Neuron 28, 53-67.

Mu, X., Silos-Santiago, I., Carrol, S. L., and Snider, W. D. (1993). Neurotrophin receptor genes are expressed in distinct patterns in developing dorsal root ganglia. J. Neurosci. 13, 4029-4041.

Nadeau, H., McKinney, S., Anderson, D. J., and Lester, H. A. (2000). ROMK1 (Kir1.1) causes apoptosis and chronic silencing of hippocampal neurons. J. Neurophysiol. 84, 1062-75.

Nakanishi, S., Maeda, N., and Mikoshiba, K. (1991). Immunohistochemical localization of an Inositol 1,4,5- trisphosphate receptor, P400, in neural tissue - Studies in developing and adult mouse brain. Journal of Neuroscience 11, 2075-2086.

Nelson, D. R. (1999). Cytochrome P450 and the individuality of species. Arch. Biochem. Biophys. 369, 1-10.

Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., and Nebert, D. W. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics 6, 1-42.

Nieto, M., Schurmans, C., Britz, O., and Guillemot, F. (2001). Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. Neuron 29, 401-413.

Nirenberg, S., and Cepko, C. (1993). Targeted ablation of diverse cell classes in the nervous system in vivo. J Neurosci 13, 3238-51.

Nirenberg, S., and Meister, M. (1997). The light response of retinal ganglion cells is truncated by a displaced amacrine circuit. Neuron 18, 637-50.

No, D., Yao, T. P., and Evans, R. M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A 93, 3346-51.

Ohki-Hamazaki, H., Sakai, Y., Kamata, K., Ogura, H., Okuyama, S., Watase, K., Yamada, K., and Wada, K. (1999). Functional properties of two bombesin-like peptide receptors revealed by the analysis of mice lacking neuromedin B receptor. J Neurosci 19, 948-54.

Ohki-Hamazaki, H., Watase, K., Yamamoto, K., Ogura, H., Yamano, M., Yamada, K., Maeno, H., Imaki, J., Kikuyama, S., Wada, E., and Wada, K. (1997). Mice lacking bombesin receptor subtype-3 develop metabolic defects and obesity. Nature 390, 165-9.

Ohyama, H., Zhang, X., Kohno, Y., Alevizos, I., Posner, M., Wong, D. T., and Todd, R. (2000). Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. Biotechniques 29, 530-6.

Omura, T. (1999). Forty years of cytochrome P450. Biochem. Biophys. Res. Commun. 266, 690-8.

Ormandy, C. J., Binart, N., Helloco, C., and Kelly, P. A. (1998). Mouse prolactin receptor gene: genomic organization reveals alternative promoter usage and generation of isoforms via alternative 3'-exon splicing. DNA Cell Biol. 17, 761-70.

Perez, S. E., Rebelo, S., and Anderson, D. J. (1999). Early specification of sensory neuron fate revealed by expression and function of neurogenins in the chick embryo. Development 126, 1715-1728.

Pettibone, D. J., Hess, J. F., Hey, P. J., Jacobson, M. A., Leviten, M., Lis, E. V., Mallorga, P. J., Pascarella, D. M., Snyder, M. A., Williams, J. B., and Zeng, Z. (2002). The effects of deleting the mouse neurotensin receptor NTR1 on central and peripheral responses to neurotensin. J Pharmacol Exp Ther 300, 305-13.

Pi, X., Voogt, J. L., and Grattan, D. R. (2002). Detection of prolactin receptor mRNA in the corpus striatum and substantia nigra of the rat. J Neurosci Res 67, 551-8.

Pilpel, Y., Sudarsanam, P., and Church, G. M. (2001). Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 29, 153-9.

Pirker, S., Schwarzer, C., Wieselthaler, A., Sieghart, W., and Sperk, G. (2000). GABA(A) receptors: immunocytochemical distribution of 13 subunits in the adult rat brain. Neuroscience 101, 815-50.

Pitkänen, A., Savander, V., and LeDoux, J. E. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. Trends Neurosci. 20, 517-523.

Price, J., Russchen, F., and Amaral, D. (1987). The limbic region. II: The amygdaloid complex. In Handbook of Chemical Neuroanatomy, A. Bjorklund, T. Hokfelt and S. LW, eds. (Elsevier), pp. 279-388.

Puelles, L., Kuwana, E., Puelles, E., and Rubenstein, J. (1999). Comparison of the mammalian and avian telencephalon from the perspective of gene expression data. European Journal of Morphology 37, 139-150.

Rampon, C., Jiang, C. H., Dong, H., Tang, Y. P., Lockhart, D. J., Schultz, P. G., Tsien, J. Z., and Hu, Y. (2000). Effects of environmental enrichment on gene expression in the brain. Proc. Natl. Acad. Sci. USA 97, 12880-4.

Rao, M. S., and Mayer-Proschel, M. (1997). Glial-restricted precursors are derived from multipotent neuroepithelial stem cells. Dev. Biol. 188, 48-63.

Rao, M. S., Noble, M., and Mayer-Proschel, M. (1998). A tripotential glial precursor cell is present in the developing spinal cord. Proc. Natl. Acad. Sci. USA. 95, 3996-4001.

Roberts, G. (1992). Neuropeptides: Cellular Morphology, Major Pathways, and Functional Considerations. In The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction, J. Aggleton, ed. (Aggleton), pp. 115-142.

Roberts, G., Woodhams, P., Polak, J., and Crow, T. (1982). Distribution of neuropeptides in the limbic system of the rat: The amygdaloid complex. Neuroscience 7, 99-131.

Rogan, M. T., and LeDoux, J. E. (1996). Emotion: systems, cells, synaptic plasticity. Cell 85, 469-75.

Rosen, J. B., and Schulkin, J. (1998). From normal fear to pathological anxiety. Psychol Rev 105, 325-50.

Rotenberg, A., Mayford, M., Hawkins, R. D., Kandel, E. R., and Muller, R. U. (1996). Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus. Cell 87, 1351-61.

Saez, E., No, D., West, A., and Evans, R. M. (1997). Inducible gene expression in mammalian cells and transgenic mice. Curr Opin Biotechnol 8, 608-16.

Sandberg, R., Yasuda, R., Pankratz, D., Carter, T., Del Rio, J., Wodicka, L., Mayford, M., Lockhart, D., and Barlow, C. (2000). Regional and strain-specific gene

expression mapping in the adult mouse brain. Proc. Natl. Acad. Sci. U.S.A. 97, 11038-11043.

Sauer, B. (1998). Inducible gene targeting in mice using the Cre/lox system. Methods 14, 381-92.

Sauer, B. (1993). Manipulation of transgenes by site-specific recombination: use of Cre recombinase. Methods Enzymol. 225, 890-900.

Scardigli, R., Schuurmans, C., Gradwohl, G., and Guillemot, F. (2001). Crossregulation between Neurogenin2 and pathways specifying neuronal identity in the spinal cord. Neuron 31, 203-217.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467-70.

Schuler, L. A., Lu, J. C., and Brockman, J. L. (2001). Prolactin receptor heterogeneity: processing and signalling of the long and short isoforms during development. Biochem. Soc. Trans. 29, 52-6.

Schwab, M., Druffel-Augustin, S., Gass, P., Jung, M., Klugmann, M., Bartholomae, A., Rossner, M., and Nave, K. (1998). Neuronal basic helix-loop-helix proteins (NEX, neuroD, NDRF): Spatiotemporal expression and targeted disruption of the NEX gene in transgenic mice. Journal of Neuroscience 18, 1408-1418.

Schwenk, F., Kuhn, R., Angrand, P. O., Rajewsky, K., and Stewart, A. F. (1998). Temporally and spatially regulated somatic mutagenesis in mice. Nucleic Acids Res 26, 1427-32.

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., and Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. Nature 385, 254-7.

Serafini, T. (1999). Finding a partner in a crowd: neuronal diversity and synaptogenesis. Cell 98, 133-6.

Serafini, T. (1999). Of neurons and gene chips. Curr Opin Neurobiol 9, 641-4.

Serbedzija, G. N., Bronner-Fraser, M., and Fraser, S. E. (1994). Developmental potential of trunk neural crest cells in the mouse. Development 120, 1709-18.

Serbedzija, G. N., Fraser, S. E., and Bronner-Fraser, M. (1990). Pathways of trunk neural crest cell migration in the mouse embryo as revealed by vital dye labeling. Development 108, 605-612.

Serbedzija, G. N., and McMahon, A. P. (1997). Analysis of neural crest cell-migration in Splotch mice using a neural crest-specific lacZ reporter. Dev. Biol. 185, 139-147.

Sherlock, G. (2000). Analysis of large-scale gene expression data. Curr. Opin. Immunol. 12, 201-5.

Simone, N. L., Bonner, R. F., Gillespie, J. W., Emmert-Buck, M. R., and Liotta, L. A. (1998). Laser-capture microdissection: opening the microscopic frontier to molecular analysis. Trends Genet. 14, 272-6.

Snider, W. D. (1994). Functions of the neurotrophins during nervous system development - what the knockouts are teaching us. Cell 77, 627-638.

Sommer, L., Ma, Q., and Anderson, D. J. (1996). neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal

determination genes taht reveal progenitor cell heterogeneity in the developing CNS and PNS. Mol. Cell. Neurosci. 8, 221-241.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat. Genet. 21, 70-71.

Supplement, N. G. (1999). The Chipping Forecast. In Nature Genetics, pp. 1-60.

Swanson, L., and Petrovich, G. (1998). What is the amygdala? Trends in Neurosciences 21, 323-331.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S., and Golub, T. S. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA 96, 2907-2912.

Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., and Church, G. M. (1999). Systematic determination of genetic network architecture. Nat Genet 22, 281-5.

Temple, S., and Qian, X. (1996). Vertebrate neural progenitor cells: subtypes and regulation. Curr. Op. Neurobiol. 6, 11-17.

Tsien, J. Z., Chen, D. F., Gerber, D., Tom, C., Mercer, E. H., Anderson, D. J., Mayford, M., Kandel, E. R., and Tonegawa, S. (1996). Subregion and cell type-restricted gene knockout in mouse brain. Cell 87, 1317-1326.

Tsien, J. Z., Huerta, P. T., and Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. Cell 87, 1327-38.

Tsujita, M., Mori, H., Watanabe, M., Suzuki, M., Miyazaki, J., and Mishina, M. (1999). Cerebellar granule cell-specific and inducible expression of Cre recombinase in the mouse. J. Neurosci. 19, 10318-23.

Vetter, M. (2001). A turn of the helix: preventing the glial fate. Neuron 29, 559-62.

Vincent, J. P., Mazella, J., and Kitabgi, P. (1999). Neurotensin and neurotensin receptors. Trends Pharmacol. Sci. 20, 302-9.

Wang, Y., Krushel, L. A., and Edelman, G. M. (1996). Targeted DNA recombination in vivo using an adenovirus carrying the cre recombinase gene. Proc. Natl. Acad. Sci. USA 93, 3932-6.

Weber, P., Metzger, D., and Chambon, P. (2001). Temporally controlled targeted somatic mutagenesis in the mouse brain. Eur. J. Neurosci. 14, 1777-83.

Weston, J. A., and Butler, S. L. (1966). Temporal factors affecting localization of neural crest cells in the chicken embryo. Dev. Biol. 14, 246-266.

White, P. M., Morrison, S. J., Orimoto, K., Kubu, C. J., Verdi, J. M., and Anderson, D. J. (2001). Neural crest stem cells undergo cell-intrinsic developmental changes in sensitivity to instructive differentiation signals. Neuron 29, 57-71.

Wodicka, L., Dong, H., Mittmann, M., Ho, M., and Lockhart, D. (1997). Genome-wide expression monitoring in Saccharomyces cerevisiae. Nature Biotechnology 15, 1359-1367.

Yamada, K., Santo-Yamada, Y., Wada, E., and Wada, K. (2002). Role of bombesin (BN)-like peptides/receptors in emotional behavior by comparison of three strains of BN-like peptide receptor knockout mice. Mol. Psychiatry. 7, 113-7.

Zhang, L., Zhou, W., Velculescu, V., Kern, S., Hruban, R., Hamilton, S., Vogelstein, B., and Kinzler, K. (1997). Gene expression profiles in normal and cancer cells. Science 276, 1268-1272.

Zhao, Y., Guo, Y., Tomac, A., Taylor, N., Grinberg, A., Lee, E., Huang, S., and Westphal, H. (1999). Isolated cleft palate in mice with a targeted mutation of the LIM homeobox gene Lhx8. Proc. Natl. Acad. Sci. USA 96, 15002-15006.

Zhou, Q., and Anderson, D. J. (2002). The bHLH Transcription Factors OLIG2 and OLIG1 Couple Neuronal and Glial Subtype Specification. Cell 109, 61-73.

Zinyk, D. L., Mercer, E. H., Harris, E., Anderson, D. J., and Joyner, A. L. (1998). Fate mapping of the mouse midbrain-hindbrain constriction using a site-specific recombination system. Curr. Biol. 8, 665-668.

Zirlinger, M., Kreiman, G., and Anderson, D. J. (2001). Amygdala-enriched genes identified by microarray technology are restricted to specific amygdaloid subnuclei. Proc. Natl. Acad. Sci. USA 98, 5270-5.

Zirlinger, M., Lo, L., McMahon, J., McMahon, A. P., and Anderson, D. J. (2002). Transient expression of the bHLH factor Neurogenin-2 marks a subpopulation of neural crest cells biased for a sensory but not a neuronal fate. Proc. Natl. Acad. Sci. USA, *In press*.