

Single-Cell Microarray Analysis in Hippocampus CA1: Demonstration and Validation of Cellular Heterogeneity

Fredrik Kamme,¹ Ranelle Salunga,² Jingxue Yu,¹ Da-Thao Tran,¹ Jessica Zhu,¹ Lin Luo,¹ Anton Bittner,¹ Hong-Qing Guo,¹ Nancy Miller,³ Jackson Wan,¹ and Mark Erlander²

¹Johnson & Johnson Pharmaceutical Research and Development LLC, San Diego, California 92121, ²Arcturus Applied Genomics, Carlsbad, California 92008, and ³OmniViz, Maynard, Massachusetts 01754

Laser capture microdissection in combination with microarrays allows for the expression analysis of thousands of genes in selected cells. Here we describe single-cell gene expression profiling of CA1 neurons in the rat hippocampus using a combination of laser capture, T7 RNA amplification, and cDNA microarray analysis. Subsequent cluster analysis of the microarray data identified two different cell types: pyramidal neurons and an interneuron. Cluster analysis also revealed differences among the pyramidal neurons, indicating that even a single cell type *in vivo* is not a homogeneous population of cells at the gene expression level. Microarray data were confirmed by quantitative RT-PCR and *in situ* hybridization. We also report on the reproducibility and sensitivity of this combination of methods. Single-cell gene expression profiling offers a powerful tool to tackle the complexity of the mammalian brain.

Key words: laser capture microdissection; gene expression profiling; RNA amplification; single cell; microarray; cell type definition

Introduction

The cellular heterogeneity of the mammalian brain is vast. This complexity prevails at different levels: different regions have specialized functions reflected in their cellular compositions; within a region several different neuronal cell types may be present; and among cells of a single cell type there may be functional differences depending on factors such as projection targets or afferent input. Cellular heterogeneity may also be further exacerbated by conditions such as Alzheimer's disease in which the progression of pathological change will vary between adjacent neurons (Braak and Braak, 1991). Consequently, a central issue when studying cellular processes in the brain has been to define the cell population of interest. This in turn is inextricably associated with the definition of cell types. This has been done using morphological criteria (Ramon Y Cajal, 1884), electrophysiological properties (Delgado-Garcia et al., 1983), projection targets (Grofova, 1975), and the expression of marker genes (Vandesande and Dierickx, 1975). Often additional heterogeneity has been revealed within morphologically defined cell types using the three last techniques (Wilcox and Unnerstall, 1990; Augood et al., 1999; Peruzzi et al., 2000). Marker genes have been particularly useful in defining cell types in the brain. Most gene markers have been identified using a "candidate gene approach," in which *in situ* hybridization or immunohistochemistry has been used to localize the expression of an mRNA or protein in tissue sections.

Laser capture microdissection allows for the selective collection of cells of interest from tissue sections, and microarray analysis permits the analysis of several thousands of genes in a single sample. The combination of the two techniques, enabled by a powerful RNA amplification method, is suited for gene expression profiling in the brain. These methods have been used to

profile gene expression in groups of neurons within the CNS (Luo et al., 1999). We have continued the development of the combination of these methods to allow for single-cell gene expression profiling. A single-cell gene expression profiling technology would provide a tool to analyze cellular heterogeneity. The heterogeneity in a region of the brain could be analyzed by picking single neurons and classifying them on the basis of their gene expression profiles. Cell types would then be defined on the basis of gene expression patterns, and specific markers could be rationally identified from the data set. Furthermore, specific cell types could be further characterized by looking at the gene expression pattern of these particular cells. Hitherto, gene expression profiling of single cells has been performed using aspiration of the intracellular contents of live cells (Eberwine et al., 1992) or through manual dissection of fixed cells using a needle or scalpel (Dell et al., 1998). After RNA amplification, the expression of genes has been detected using radioactive labels, usually on limited, membrane-bound arrays (Mackler et al., 1992; Brooks-Kayal et al., 1998). Laser capture microdissection offers an alternative to manual microdissection and has a much higher throughput in cell collection. Fluorescent labels shorten the time for acquiring hybridization signals from the hybridized array from days to minutes. Consequently, the use of laser capture microdissection and fluorescent labels improves significantly throughput. Furthermore, and perhaps more importantly, the use of large, reproducible microarrays and fluorescent labels allows for cross comparison between different experiments and the building of gene expression databases.

To evaluate the performance of the assembled techniques—laser capture microdissection, RNA amplification, and microarray hybridization—and the possibility of identifying cell types on the basis of gene expression patterns using this set of technologies, we analyzed single cells captured from the hippocampus CA1 subregion of an adult rat. The region was chosen because it is an exceptionally well studied region of the brain, not the least in

Received Aug. 21, 2002; revised Feb. 13, 2003; accepted Feb. 13, 2003.

Correspondence should be addressed to Fredrik Kamme, Johnson & Johnson Pharmaceutical Research and Development LLC, 3210 Merryfield Row, San Diego, CA 92121. E-mail: fkamme@prdu.s.jnj.com.

Copyright © 2003 Society for Neuroscience 0270-6474/03/233607-09\$15.00/0

terms of gene expression. Most of the neurons are pyramidal, with ~10% of the remaining neurons being interneurons. This provided us with an opportunity to test reproducibility, sensitivity, and the ability to distinguish cell types, if interneurons were captured, using these technologies.

We found that the methods were reproducible, sensitive, and indeed capable of distinguishing among different neuronal cell types.

Materials and Methods

Laser capture microdissection

Animals used in this study were adult female Sprague Dawley rats weighing 250–300 gm. Brains were fresh frozen and cryosectioned at 12 μ m on uncoated colorfrost slides (VWR Scientific). Sections were Nissl stained using the following protocol: 100% ethanol for 1 min, 95% ethanol for 10 sec, 70% ethanol for 10 sec, 50% ethanol for 10 sec, PBS for 10 sec, 0.5% cresyl violet stain for 40 sec, 3 \times PBS for 10 sec, 70% ethanol for 10 sec, 95% ethanol for 10 sec, 95% ethanol + 1.6% acetic acid for 5–10 sec, 95% ethanol for 10 sec, 100% ethanol for 10 sec, and xylene for 1 min; they were finally left to air dry. Cells were identified as presumptive pyramidal neurons by their shape and localization in the pyramidal cell layer of hippocampus CA1. Single cells were captured using the PixCell II laser capture microdissection instrument (Arcturus, Mountain View, CA) onto standard caps (model TF100). Caps were then put in 500 μ l tubes and frozen on dry ice.

A critical issue was the purity of the single-cell captures. The main sources of contamination of concern were nonspecific lifting of cells on the capture cap, and neighboring cells, beneath, above, or adjacent to the cell of interest that were captured along with the cell of interest. Because the cap is slightly concave, the outer rim of the membrane will touch the section and may pick up tissue nonspecifically. This was solved by cutting out a small square of the membrane around the captured cell, detaching the square of membrane with the captured cell on it, and extracting the RNA by immersing the piece of film into a tube with extraction buffer. This eliminated nonspecific lifting of tissue. The risk of capturing part of an adjacent cell depends on factors such as cell morphology and thickness of the section. We chose to use 12- μ m-thick sections.

RNA extraction

The cut-out piece of film with the captured cell was put directly into 8 μ l of RLT buffer (Qiagen, Valencia, CA) supplemented with 300 ng of polyinosinic acid (Sigma, St. Louis, MO) in a 500 μ l tube. The tube containing the cell and extraction buffer was then incubated at 42°C for 20 min. An equal volume of 70% ethanol was added, and the mix was applied to an RNeasy column (Qiagen). The column was washed following the manufacturer's protocol, except that the volume of PE buffer was reduced to 100 μ l per wash. The extracted RNA was concentrated to a volume of 10 μ l. For the sensitivity assay and single-cell RT-PCR analyses, the same extraction protocol was used, but with a smaller column, Zymo-Spin I (Zymo Research, Orange, CA), which reduces the elution volume to 10 μ l.

T7 antisense RNA amplification

A modified version of the T7 antisense RNA (aRNA) amplification method (Van Gelder et al., 1990) was used. A double-stranded cDNA library containing a T7 RNA polymerase promoter site in the 5' end is made from the input mRNA and transcribed using T7 RNA polymerase. The process is repeated in a second round.

First round. T7-cDNA (0.5 μ g) synthesis primer (5'-TCTAGTCG-ACGCCAGTGAATTGTAATACGACTCACTATAGGGAGATTTTTTTT-TTTTTTTTTTTTTTTT-3') (Operon, Alameda, CA) was added. The mix was denatured at 70°C for 10 min and put on ice. cDNA was synthesized using Superscript II (200 U per reaction; Invitrogen, Carlsbad, CA) in 50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 20 mM DTT, 500 μ M deoxy NTPs (dNTPs), and 30 U of RNasin (Promega, Madison, WI) in a 20 μ l reaction for 2 hr at 42°C. The reaction was terminated by incubating at 70°C for 10 min. One microliter of the first-strand cDNA was removed for real-time PCR analysis as described below. To make the second strand, 131 μ l of H₂O, 30 μ l of 5 \times second-strand buffer [1 \times = 20 mM Tris-HCl, pH 6.9, 4.6 mM MgCl₂, 90 mM KCl, 0.15 mM β -NAD⁺, 10 mM

(NH₄)₂SO₄ (Invitrogen)], 3 μ l of 10 mM dNTPs (Amersham Biosciences, Piscataway, NJ), 20 U *Escherichia coli* DNA polymerase I [5 U/ μ l (Invitrogen)], 2 U RNase H [2 U/ μ l (Invitrogen)], and 10 U *E. coli* DNA ligase [10 U/ μ l (Invitrogen)] was added. The mix was incubated at 16°C for 2 hr. Ten units of T4 DNA polymerase [5 U/ μ l (Invitrogen)] was then added, and the mix was further incubated at 16°C for 15 min. The reaction was terminated by incubating at 70°C for 10 min. One hundred nanograms of polyinosinic acid were added to each sample, and then 750 μ l of PB buffer (Qiagen) was added. The samples were purified on a PCR purification column (Qiagen) according to the manufacturer's directions. The DNA was eluted in 1 mM Tris-HCl, pH 8, and dried down to 8 μ l. The double-stranded cDNA carrying a T7 RNA polymerase promoter was transcribed using the Ampliscribe transcription kit (Epicentre, Madison, WI). The reactions were incubated at 42°C for 3 hr. One microliter of DNase I (included in the kit) was added, and the mix was incubated for 20 min at 37°C. The resulting aRNA was cleaned up using the RNeasy kit (Qiagen). To each sample 100 ng of polyinosinic acid, 70 μ l of RLT buffer, and 50 μ l of 100% ethanol were added in sequence. The samples were then loaded onto RNeasy columns and treated according to the manufacturer's directions, except that the volume of RPE wash buffer was reduced to 150 μ l per wash. The cleaned aRNA was eluted in H₂O and dried down to 10 μ l.

Second round. One microgram of random hexamers (Amersham Biosciences) was added to the aRNA, and the sample was denatured at 70°C for 10 min and cooled on ice. Nine microliters of first-strand cocktail were added and incubated at 37°C for 2 hr. The reaction was killed at 70°C for 10 min. Two units of RNase H were added, and the reaction was incubated at 37°C for 30 min, followed by 95°C for 2 min. After a quick spin, 1 μ g of T7dT₂₁ oligo was added, and the mix was heated to 70°C for 10 min, 42°C for 10 min, and put on ice. Second-strand synthesis mix without *E. coli* DNA ligase (129 μ l) was added and incubated at 16°C for 2 hr. The double-stranded cDNA was polished by adding 10 U of T4 DNA polymerase and a further incubation was done at 16°C for 10 min. The enzymes were heat killed at 65°C for 10 min. The template was purified, concentrated, and transcribed as described for the first round. The resulting aRNA was purified on an RNeasy column and eluted in 30 μ l of H₂O. To make Cy3-labeled cDNA target, 5 μ g of random hexamers was added, and the mix was denatured at 70°C for 10 min and cooled on ice. cDNA was synthesized using Superscript II (500 U per reaction) in 50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 20 mM DTT, 500 μ M dATP, dGTP, dTTP, 40 μ M dCTP, 40 μ M Cy3-dCTP (Amersham Biosciences), and 45 U of RNasin in a 50 μ l reaction for 2 hr at 37°C. To remove the aRNA from the cDNA, the sample was digested using 10 U of RNase H and 0.1 U of RNase A (Sigma) for 10 min at 37°C. The samples were purified on PCRquick columns (Qiagen).

Real-time PCR

Real-time quantitative PCR was performed using either a Lightcycler (Roche, Indianapolis, IN) or a Smartcycler (Cepheid, Sunnyvale, CA). To monitor the T7 amplification reaction, an aliquot (1 μ l) was removed after the first and last cDNA syntheses and diluted fourfold and 300-fold, respectively. Two microliters of the dilution were used for real-time PCR analysis using the Lightcycler. On the Lightcycler, the reaction mix contained 3 mM MgCl₂ (added), 0.5 μ M each of forward primer and reverse primer, and 2 μ l of Sybr Green Mix (Roche) premixed with 0.18 μ g of Taqstart antibody (Clontech). The PCR parameters were 95°C for 30 sec, 40 cycles of 95°C for 0 sec, 55°C for 5 sec, and 72°C for 7 sec. At the end of the program a melt curve analysis was done. For the sensitivity assay and the single-cell PCR, a Smartcycler was used. On the Smartcycler the mix contained 2 U Ex-Taq (Panvera, Madison, WI), 0.2 \times Sybr Green (Molecular Probes, Eugene, OR), 0.2 mM dNTPs, 0.4 μ M each primer (Genset, La Jolla, CA; HPLC purified), 2–4 mM MgCl₂ (depending on primers), 0.12 mg/ml BSA (Sigma), 90 mM trehalose (Sigma), and 0.12% Tween 20 (Sigma) in 1 \times Ex-Taq buffer supplied with the enzyme. The PCR parameters were 95°C for 30 sec, 45 cycles of 95°C for 5 sec, 54–70°C (depending on primers) for 10 sec, and 72°C for 15 sec. At the end of each program a melt-curve analysis was done. All primers were 20 mers. PCR efficiency, E (optimally 1, mass increase after each cycle will be 2^E), for

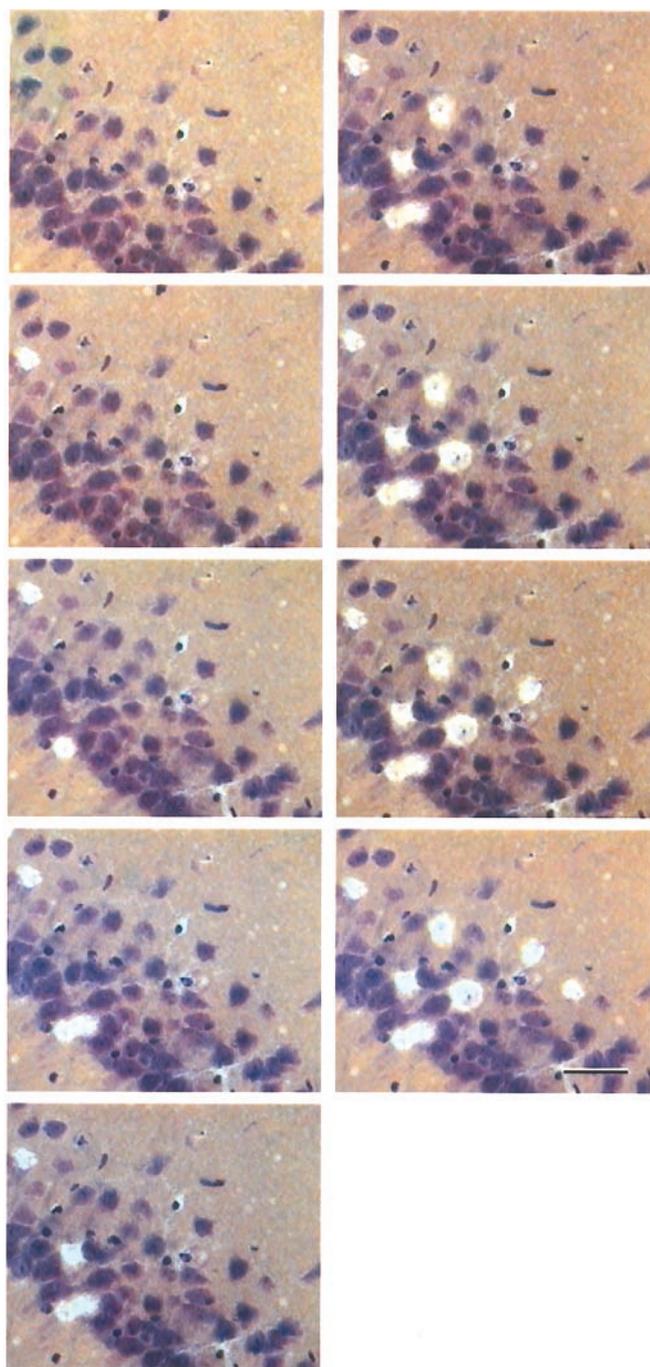


Figure 1. Sequential laser capture microdissection of cells in the pyramidal layer of hippocampus CA1. The section was stained with cresyl violet. Scale bar, 50 μ m.

the different reactions was as follows: AA858959, 0.97; AA817769, 1.0; 18S ribosomal RNA, 0.87; neuron-specific enolase (NSE), 0.99.

Plasmid standards either were obtained from our in-house clone collection or created by cloning the appropriate PCR product using a TOPO TA cloning kit (Invitrogen). Note that the Sybr Green Kit from Roche uses dUTP and therefore would require uracil-*N*-glycosidase negative bacterial strains for cloning. All clones used were sequenced to verify their identity. The plasmids were linearized, and on the basis of the A_{260} OD, a 10-fold dilution series was made, ranging from 500 fg to 5 ag per reaction. This was then used as the standard curve for the respective gene.

Microarray analysis

The arrays used were cDNA microarrays spotted using an Amersham Biosciences Generation III spotter onto Corning GAPS slides. Each array

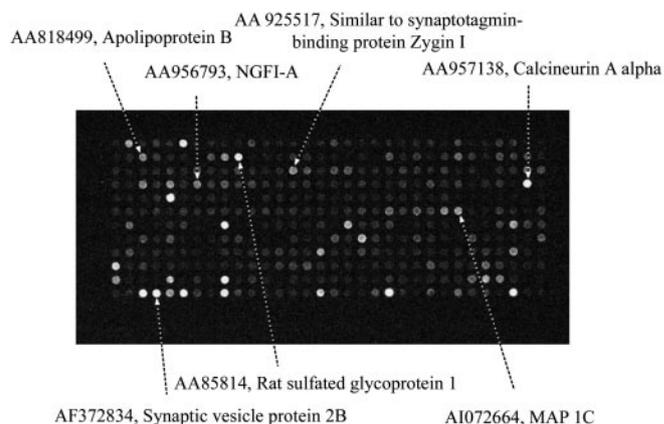


Figure 2. Partial microarray image. The microarray was hybridized to a Cy3-labeled target generated from a single cell. The identities of seven genes are highlighted.

contained 4529 clones, each spotted in duplicate; 89% were IMAGE clones purchased from Research Genetics (Invitrogen); 22 were *Arabidopsis* clones. The entire generated Cy3-labeled target was hybridized overnight at 42°C onto a single chip in a buffer containing 50% formamide and 1× Microarray Hybridization buffer (RPK-0325, Amersham Biosciences). The arrays were washed in 1× SSC/0.1% SDS at room temperature, 5 min in 1× SSC/0.1% SDS at 55°C, 5 min in 0.1× SSC/0.1% SDS at 55°C, and a final rinse in 0.1× SSC at room temperature. The arrays were scanned in a ScanArray 4000 (PerkinElmer Life Sciences, Boston, MA). Quantification was done using Imagene (Biodiscovery, Marina del Rey, CA). Microarray data from were normalized to the 75th percentile.

Sensitivity assay

Sensitivity can be assayed in different ways. A common way is to spike in exogenous transcripts into the hybridization target. To gauge the sensitivity of the entire assembled process, we decided to select genes from the output microarray data in a range of intensities from clearly expressed (14× plant gene background) to not expressed (<plant gene background). The cDNA copy number of these genes in laser-captured cells from hippocampus CA1 was then determined by quantitative RT-PCR. Eleven genes were selected from the microarray data with average expression levels across the samples that ranged in intensity from 1269 to 80 (median plant gene background value was 91). PCR primers were designed to generate a specific fragment of each gene with a length of between 213 and 278 bases. Plasmid PCR standards were generated for these genes by cloning the PCR product. Triplicate samples of 30 CA1 cells were laser captured. RNA was extracted as above and reverse transcribed using the T7dT₂₁ oligonucleotide as a primer. The abundance of each of the 11 selected genes was measured using real-time quantitative PCR on a Smartcycler.

Clustering using OmniViz

Data preparation. The data normalization involved two steps: thresholding to 100 U and subsequent ratio creation. The value of any intensity data below the threshold was increased up to 100 U. This threshold between supposedly nonexpressed and expressed genes was justified by the PCR-sensitivity assay, which consistently detected genes with a microarray signal of 96.3 or higher (see below). After thresholding, the geometric mean was calculated across experiments for each gene individually. The mean value for each gene was then used to divide the collection of experimental intensities for that gene across experiments. Thus for each gene the relative response across experiments could be compared.

Cluster analysis. The ratios were clustered in the OmniViz software package (OmniViz) using an agglomerative hierarchical clustering algorithm with complete linkage. A Euclidean metric for the pairwise comparisons was selected. The data were manually filtered to eliminate gene clusters without apparent regulation. The resulting dendrogram containing 1284 genes was cut to produce 66 clusters.

In situ hybridization riboprobe synthesis.

DNA templates for riboprobe synthesis were taken from the cDNA clone collection used for making the microarray and resequenced to verify their identities. Probe lengths were between 300 and 600 bp. The riboprobe for parvalbumin was labeled with digoxigenin using a DIG RNA Labeling Kit (Roche). One microgram of linearized template was transcribed with either T7 RNA polymerase or T3 RNA polymerase. The riboprobe was purified by ethanol precipitation, resuspended at 1 $\mu\text{g}/\mu\text{l}$ in hybridization buffer, and stored frozen. For radioactive labeling of riboprobes, 1 μg of linearized template was transcribed with T7 or T3 RNA polymerase in the presence of 150 μCi ^{35}S -UTP (DuPont NEN, Boston, MA). Probes were purified on a G-50 Sephadex Quick Spin column (Roche) and stored at -20°C until used.

Double in situ hybridization

Sections were air-dried and subsequently fixed in freshly made 4% paraformaldehyde/PBS for 20 min at room temperature. After rinsing in PBS (two times for 15 min) and $5\times$ SSC (15 min), slides were prehybridized in hybridization buffer [50% formamide (Sigma), $5\times$ SSC, 40 $\mu\text{g}/\text{ml}$ salmon sperm DNA (Eppendorf)] at 58°C for 2 hr. The digoxigenin-labeled riboprobe for parvalbumin was mixed with each ^{35}S -labeled riboprobe to make a final concentration of 500 ng/ml and 10^7 cpm/ml, respectively. The probe mix was denatured for 5 min at 75°C and cooled on ice; 70 μl of probe mix was added to each slide. The hybridization reaction was performed at 58°C for 16 hr. After hybridization, sections were washed in $2\times$ SSC for 30 min at room temperature, $2\times$ SSC with 1 mM DTT for 1 hr at 65°C , and then in $0.1\times$ SSC with 1 mM DTT for 1 hr at 65°C . Sections were then equilibrated in buffer 1 (10 mM Tris-HCl, pH 7.5, and 150 mM NaCl) and incubated with an alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche), diluted 1:500, for 2 hr at room temperature. Excess antibody was removed by two 15 min washes in buffer 1, and the sections were equilibrated for 5 min in buffer 2 (100 mM Tris-HCl, pH 9.5, 100 mM NaCl, and 50 mM MgCl_2). Color development was performed at room temperature overnight in buffer 2 containing nitroblue tetrazolium chloride and 5-bromo-4-chloro-3-indolyl-phosphate, toluidine salt (Roche). Staining was stopped by a 10 min wash in Tris/EDTA (10/1 mM, pH 8.0), and nonspecific staining was removed in 95% ethanol for 1 hr. Sections were rehydrated for 15 min in deionized water to remove the precipitated Tris and then dehydrated through successive baths of EtOH (70, 95, 100%) and air dried. For radioactive signal detection, slides were dipped in Ilford K-5 nuclear emulsion diluted 1:1 with water at 42°C and exposed at 4°C for 9 weeks. The emulsion was developed in Kodak D-19 and counterstained with YO-PRO (Molecular Probes) 1:10,000 in PBS for 20 min. Slides were imaged using a SPOT camera (Diagnostic Instruments, Sterling Heights, MI) mounted on a Nikon Optiphot microscope with a $60\times$ objective. Images were assembled using Adobe Photoshop 6.0 (San Jose, CA).

Validation of microarray data by single-cell RT-PCR

Single cells were captured from the same region of the CA1 subregion that was chosen for the microarray analysis. RNA was extracted and reverse transcribed as described above, using a mix of an oligo-dT₁₂₋₁₈ primer (50 ng per reaction) and a 21 mer primer specific for 18S ribosomal RNA (50 ng per reaction) in a total volume of 10 μl . The resulting cDNA was diluted 12-fold. Quantitative PCRs, using Sybr green detection on a Smartcyler, were performed for five different genes for each single cell: NSE to show that the cell picked was a neuron (PCR run in duplicate), 18S ribosomal RNA to assess RNA yield (PCR run in quadruplicate), parvalbumin to exclude parvalbumin-positive interneurons (PCR run in duplicate), and for the Rat H(+)-transporting ATPase

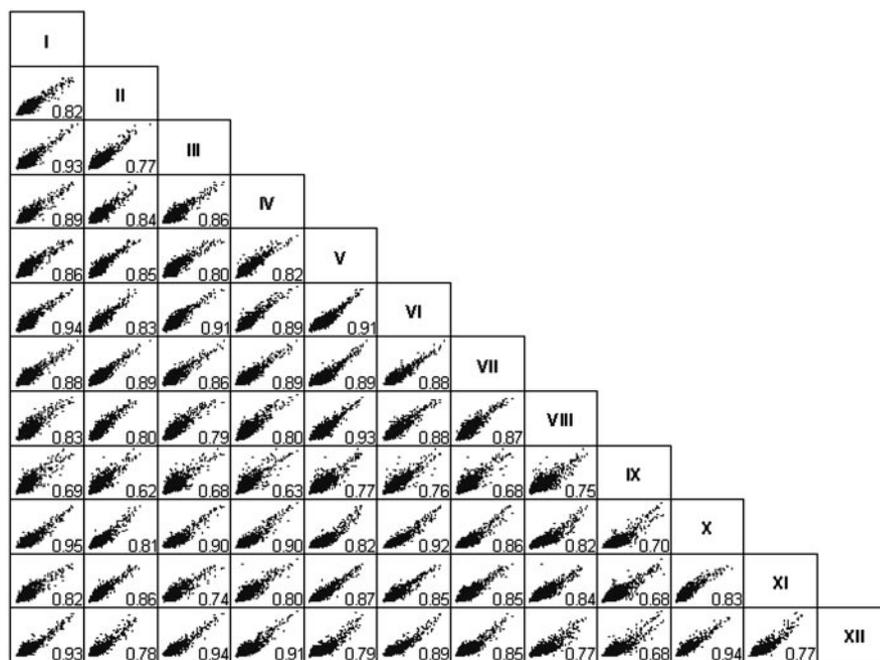


Figure 3. Compilation of scatter plots among the 12 cells, roman numerals I–XII. R^2 value for each pair-wise comparison is indicated in the respective box.

(AA858959) (PCR run in quadruplicate) and the expressed sequence tag AA817769 (PCR run in quadruplicate). For all genes, except parvalbumin, plasmid standards were used. To confirm that the melting curves from the real-time PCR analysis corresponded to the expected amplicon, products from PCR for the different genes were run on an agarose gel to check the size.

Results

Single cells were captured in the CA1 subregion of the dorsal hippocampus as exemplified in Figure 1. Fourteen cells were captured throughout the width of the pyramidal cell layer. In addition to these, two mock captures were done as negative controls, where the cap was placed in contact with the section at the same location, but no laser pulse was fired. The thicker the section, the more of the cell of interest will be captured; however, the risk of capturing unwanted material beneath or over the cell of interest also increases. We used 12 μm sections. In 2 of the 12 cells, we detected, by PCR, glial fibrillary acidic protein mRNA in the cDNA generated after two rounds of T7 amplification, indicating astrocytic contamination. In the clustering analysis shown later, the two cells, numbers V and VI, did not cluster next to each other, indicating that the overall impact of the astrocytic contamination was small. After RNA extraction and cDNA synthesis, the cells were screened for expression of NSE using quantitative PCR. Two cells were NSE negative and omitted. An additional PCR was done on the Cy3-labeled first-strand cDNA after two rounds of T7 aRNA amplification (which was the hybridization target). The average amplification-fold achieved for NSE was 5×10^5 . The two negative samples remained negative by PCR, and their hybridization images were also negative. An image of a part of a microarray hybridized to target from a single cell is shown in Figure 2. The microarray data for each cell were plotted against each other cell in scatter plots in Figure 3. One cell, number IX, had on average a lower correlation against the other cells, $R^2 = 0.7$, whereas the average R^2 among the remaining 11 cells was 0.85.

Table 1. Estimation of the sensitivity of the combination of laser capture microdissection, T7 aRNA amplification, and cDNA microarray analysis using quantitative RT-PCR

Source ID	Microarray signal	SEM of microarray signal	cDNA copy number	SEM of cDNA copy number
M11931	1269.2	137.4	2061.7	156.9
AI029254	337.5	30.9	236.3	4.7
AA900184	141.0	10.4	59.1	1.6
AA818955	121.7	16.8	0.7	0.3
AA859035	116.7	13.2	9.5	1.2
AI044508	97.3	4.0	23.9	1.3
AA925795	96.3	13.3	97.0	3.2
AI071912	81.8	2.9	0.0	
AI385137	80.2	7.9	0.0	
AA964718	80.2	2.2	0.0	
AA819293	80.0	5.5	0.0	

Eleven genes with a range of microarray expression levels from 1269.2 to 80 were quantified by quantitative RT-PCR on RNA from 30 laser captured cells from hippocampus CA1, captured in triplicate.

Single-cell gene expression profiling is sensitive

Table 1 shows a compilation of the sensitivity data. All genes with a microarray signal of ≥ 96.3 were expressed according to the PCR data, whereas none of the genes with a microarray signal of ≤ 81.8 were detected by PCR. The overall correlation between microarray signal and cDNA copy number determined by PCR was strong ($R^2 = 0.99$). Some of the genes with a microarray signal above 96.3 had a cDNA copy number ranging from 23.9 copies per cell to 0.7 copies per cell, indicating that these were rare transcripts.

Single-cell gene expression profiling reveals at least two different neuronal cell types in CA1

Figure 4 shows a hierarchical clustering of the microarray data. Genes (3201) that did not show any regulation were omitted. Each column corresponds to a single cell, roman numerals I–XII. As shown in the cluster tree, or dendrogram, cell IX clustered outside the other 11 cells, indicating differences in gene expression. Figure 5B (cluster I) shows a cluster of genes that were found to be highly expressed in cell IX but low in the other cells. One of the genes in this cluster was parvalbumin, which is a well established marker for one of the types of interneurons in the hippocampus (Kosaka et al., 1987). This suggested that cell IX was an interneuron. The OmniViz program allows the clustering result to be viewed in a proximity map or Galaxy view. The Galaxy visualization (Gedeck and Willett, 2001) projects the genes shown in the dendrogram from Figure 4 in a complementary but different way, such that genes with closely related expression will appear close to each other, and genes with unrelated expression are farther apart. The Galaxy view of the genes is based on a principal component analysis (PCA) (Mardia et al., 1979) of the gene expression profiles in conjunction with various heuristics to emphasize cluster membership. PCA is a popular statistical method that is used to reduce the dimensionality of the data (in our case we have 12 dimensions) while capturing the bulk of the variability in the data into a smaller number of new dimensions (the galaxy shows two dimensions). Each gene becomes a single point in Figure 5A and is plotted using the new coordinate system based on the PCA analysis. Unlike the cluster tree in Figure 4, the galaxy view in Figure 5A emphasizes many-to-many relationships. Two clusters that may not be placed next to one another on the cluster tree view may appear next to one another on the Galaxy view. Thus, the actual number of clusters selected and the

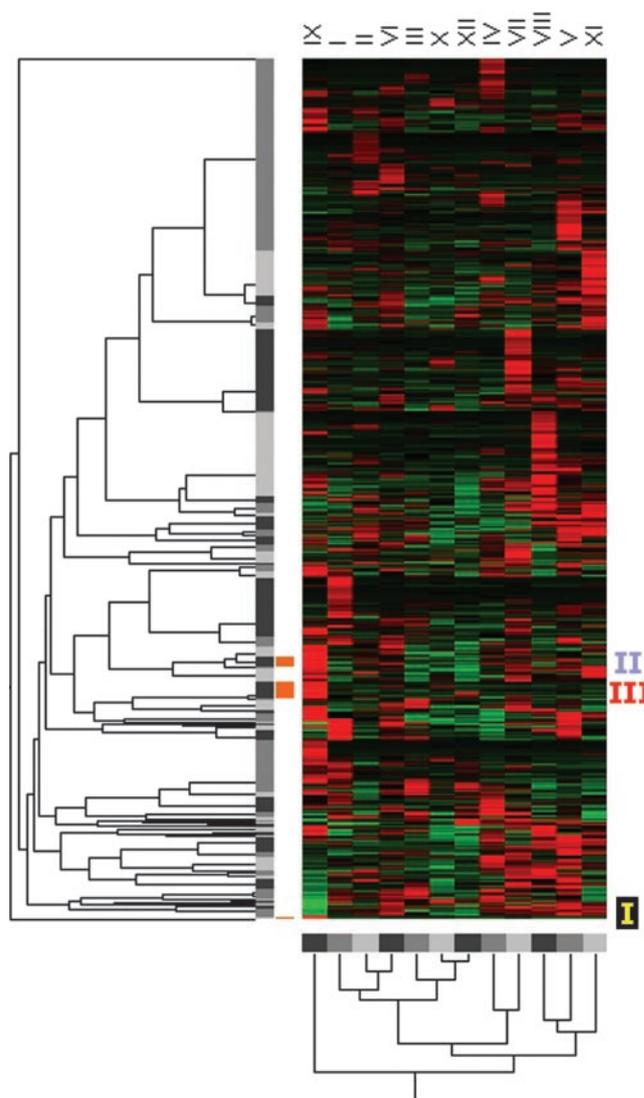


Figure 4. Hierarchical clustering of the microarray data set. Each column represents a single cell, each row a gene. Three clusters are indicated by an orange box to the left of the columns and by roman numerals I–III to the right of the columns. These three clusters are shown in Figure 5.

specific cluster membership are far less critical because the proximity map allows genes with similar expression patterns to be projected near one another on the galaxy map, whether or not they are found in the same cluster. Additional clusters with genes expressed in cell IX but not in the remaining cells were found (Fig. 5B, clusters II and III). These groups of genes were located close to each other in the Galaxy view (Fig. 5A). From two of the three clusters, one of them containing parvalbumin, five genes were selected for double *in situ* hybridization to experimentally validate the clustering result, which suggested coexpression with parvalbumin. The five selected genes were thus expected to be expressed in parvalbumin-positive cells but not in parvalbumin-negative cells. The selected genes were the GABA transporter-1 (GAT-1), neurofilament-H (NF-H), a K^+ channel subunit (NGK2-Kv4), vesicle-associated membrane protein 1 (VAMP1), and the myocyte enhancer factor-2C (MEF2C). The expression of these genes was colocalized with parvalbumin using double-labeling *in situ* hybridization. *In situ* for parvalbumin was done using digoxigenin and subsequent alkaline phosphatase detection, whereas the other genes were done using ^{35}S . As expected, parvalbumin-positive cells were found scattered in stratum pyra-

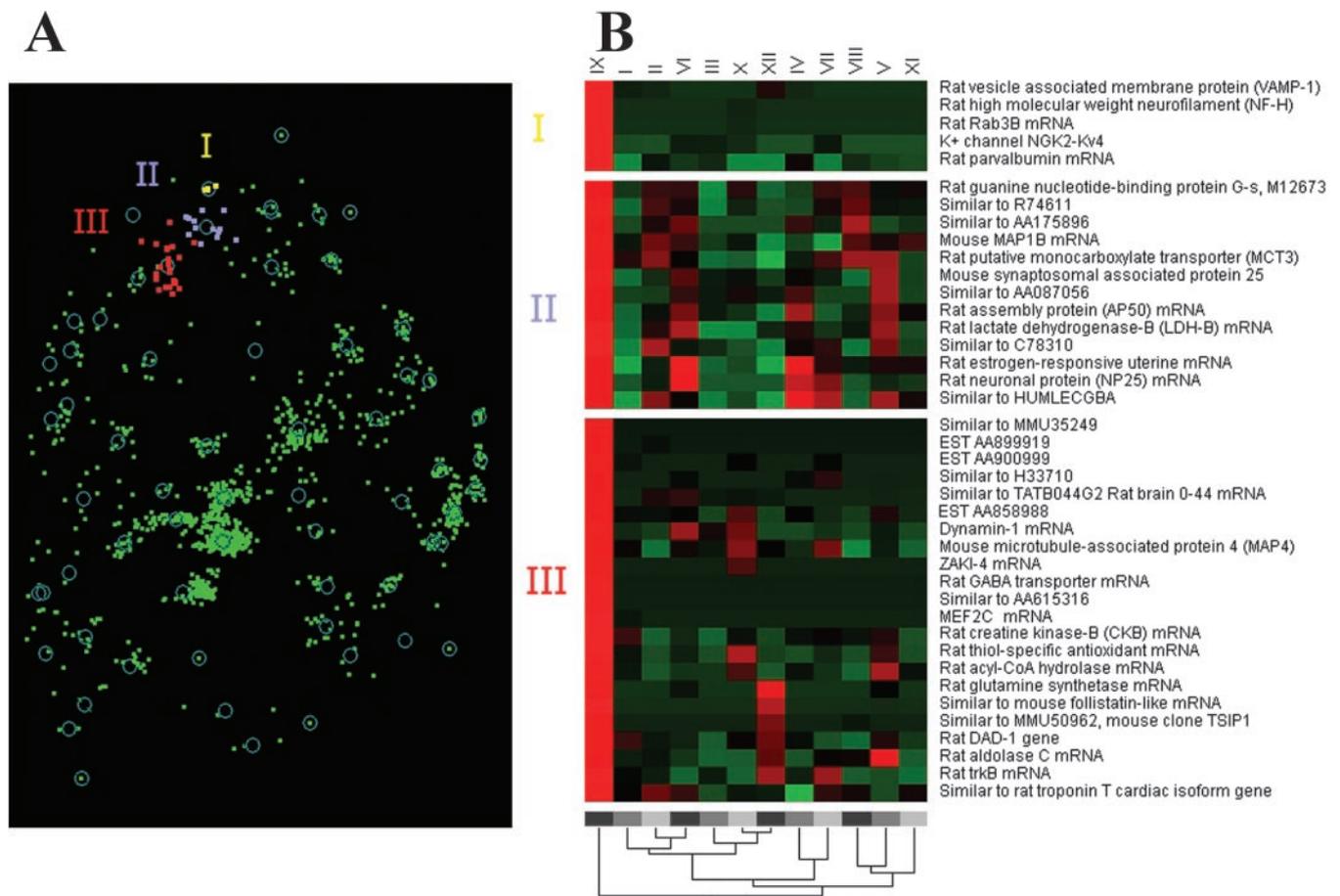


Figure 5. Galaxy and cluster tree views. *A*, A galaxy view of the clustered data set as described in Materials and Methods. Three clusters are color coded and labeled as I, II, and III. These three clusters are shown in a cluster tree view in *B*. The genes included in these three clusters are all expressed in cell IX, but not expressed, or expressed at a lower level, in the remaining 11 cells. The column dendrogram indicated that cell IX was dissimilar from the other cells.

midale and oriens of CA1 and CA3 (Kosaka et al., 1987). On average we found 47 parvalbumin-positive cells in CA1 per section. All of the five genes colocalized with parvalbumin as shown in Figure 6. Cell counts of the double *in situ* hybridizations confirmed the colocalization of the genes with parvalbumin in the CA1 region. As shown in Table 2, the sets of cells expressing NGK2-Kv4 and MEF2C were practically identical to that of parvalbumin. Both NF-H and VAMP1 defined a subset of parvalbumin-positive cells. In the case of GAT-1, all of the parvalbumin-positive cells in CA1 were positive for GAT-1, whereas 53% of the GAT-1-positive cells were positive for parvalbumin. Thus, parvalbumin defined a subset of GAT-1-positive cells.

Validation of differences between pyramidal neurons

One cluster of genes suggested that a group of genes was expressed at a higher level in some pyramidal neurons and at a lower level in others (Fig. 7). This indicated nonrandom differences in gene expression within pyramidal neurons. To validate this finding we captured additional single cells in hippocampus CA1 and performed quantitative RT-PCR for two of the genes in Figure 7, Rat H(+)-transporting ATPase (AA858959) and an expressed sequence tag (AA817769), to show that their expression correlated and that there were differences among individual pyramidal cells. For each single cell, RT-PCR was done for NSE, 18S ribosomal RNA, parvalbumin, AA858959, and AA817769. Cells were screened for NSE and parvalbumin expression. Of 11 cells captured, one did not express detectable levels of NSE and was omit-

ted. No cells expressed parvalbumin. The values for AA817769 and AA858959 were normalized to the 18S ribosomal RNA value to correct for different RNA yields from the single cells. The PCR data for the 10 cells is shown in Figure 8. For the two genes, the PCR data were normalized to the maximum value for each gene within the data set, which was defined as 100%. There was a significant variation in the expression of the two genes within the 10 cells. Furthermore, the expression of AA817769 and AA858959 correlated with each other between the single cells. The three cells with the highest expression of AA858959 also had the highest expression of AA817769. In contrast, NSE expression did not covary with the two genes. The cells with the highest expression of NSE were cells 8, 9, and 10. The data show that the expression of AA858959 and AA817769 varied between individual CA1 pyramidal cells and that the expression of the two genes correlated with each other.

Discussion

We have shown that single cells can be captured, their RNA extracted and amplified using a T7 aRNA amplification system, labeled using direct incorporation of Cy3, and hybridized to cDNA glass microarrays with high sensitivity and relatively high reproducibility.

We decided to estimate the sensitivity of the entire process by selecting a set of genes with a range of expression from clearly expressed (14× plant gene background) to probably not expressed (< background). These genes were then quantified inde-

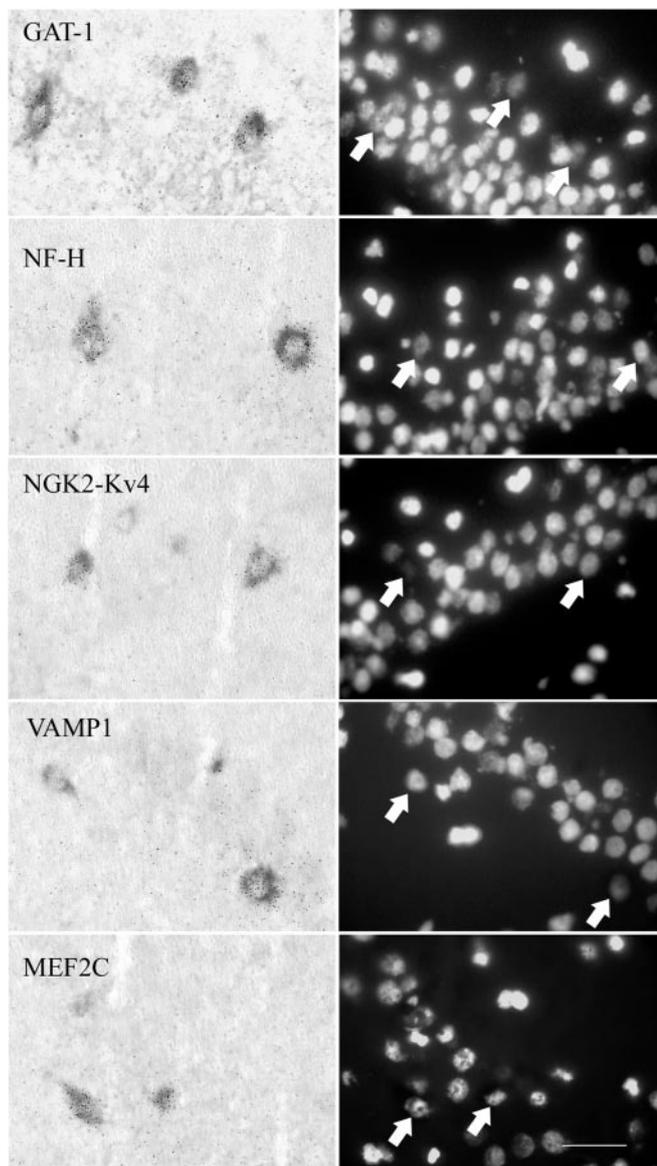


Figure 6. Microarray data validation using double-labeling *in situ* hybridization. Five genes selected from two of the cluster described in Figure 5 were analyzed by *in situ* hybridization together with parvalbumin. Presence of parvalbumin mRNA is indicated by light-gray staining, and the colocalizing gene is indicated by black silver grains. The right column shows a nuclear counterstain, using YO-PRO, of the same field of view as the respective *in situ* hybridization. Scale bar, 50 μ m.

pendently in laser-captured tissue from the same region using quantitative PCR. The data show that all of the five genes with an expression level between 141 and 96 were expressed, whereas none of the four genes with an expression level between 81 and 80 were expressed. The correlation between microarray signal and cDNA copy number as determined by PCR was strong: $R^2 = 0.99$. Three of the expressed genes above, with a microarray signal ranging from 122 to 97, had a cDNA copy number, as determined by PCR, of 23.9–0.7 copies per cell. cDNA copy numbers will not exactly reflect mRNA copy numbers, because of factors such as imperfect priming and reverse transcription. However, using a collection of 11 genes, we expected to reproduce the overall average reverse transcription efficiency of the sample. The data thus strongly suggest that even rare transcripts can be detected by T7 aRNA amplification and microarray hybridization in this single-cell approach.

Table 2. Cell counts of double *in situ* hybridizations in CA1

Gene	% of parvalbumin-positive cells colabeled	% of labeled cells that also were parvalbumin positive
GAT-1	100	53
NF-H	57	94
VAMP1 ^a	77	100
NGK-Kv4	100	96
MEF2C	100	98

Percentage of parvalbumin-positive cells that were also positive for the gene specified in column 1 is shown in column 2. Percentage of cells positive for each of the genes in column 1 and also positive for parvalbumin is shown in column 3.

^aThere was a low general level of expression of VAMP1 throughout the striatum pyramidale of CA1. Expression in parvalbumin-positive cells was however clearly higher and used as the threshold for counting VAMP1-positive, parvalbumin-negative cells.

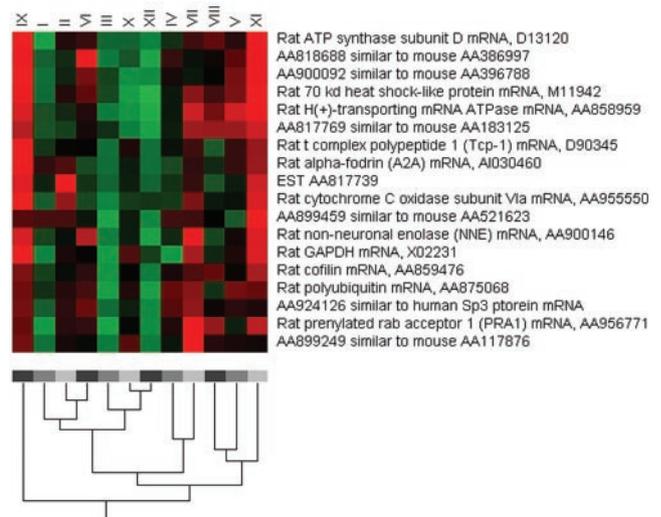


Figure 7. A cluster indicating differences in gene expression among the pyramidal neurons. The column dendrogram splits the cells in two major groups, in addition to the interneuron. Two genes were selected for validation using RT-PCR: Rat H(+)-transporting ATPase mRNA, AA858959, and AA817769 similar to mouse AA183125.

Cluster analysis identified a group of genes expressed in one of the 12 neurons, cell number IX, but not in the other neurons. Among this group of genes was parvalbumin, which is an established marker for one of the types of interneurons in CA1 (Kosaka et al., 1987; Danos et al., 1991). Five additional genes were selected out of this group for double *in situ* hybridization with parvalbumin to validate the clustering result. For all of the genes, the colocalization with parvalbumin in CA1 was confirmed. In CA1, where the initial cells were laser captured, the sets of cells expressing parvalbumin and either NGK2-Kv4 or MEF2C were practically identical. Expression of NF-H and VAMP1 correlated strongly with parvalbumin expression. Finally, parvalbumin-positive cells were a subset of GAT-1-positive cells, consistent with the view of parvalbumin as a marker for one of the types of interneurons (Kosaka et al., 1987). These data clearly confirm the clustering result that showed these genes to be enriched in the parvalbumin-positive cell. We have thus demonstrated the feasibility to use single-cell gene expression profiling to identify cell types in a mixed cell population. Single cells are picked from a population of cells that appears homogeneous; gene expression profiles are then generated, and the different cell types are defined by cluster analysis of their gene expression patterns. Marker genes for these different cell types are then rationally selected from the data set. Furthermore, the description of genes selectively expressed by defined cell types may aid in the biological characterization of these cells. Clustering is used frequently, particularly in

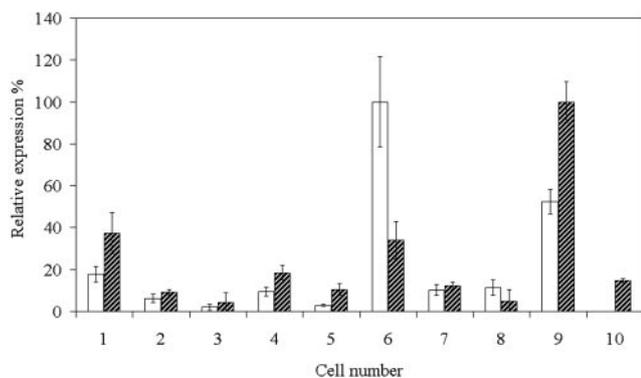


Figure 8. Quantitative RT-PCR confirmed the covariation in expression of AA858959 and AA817769 between individual CA1 pyramidal cells. Ten single cells were analyzed. Samples were analyzed in quadruplicate. The expression levels for each gene are normalized to the maximum expression for that gene within the set, which is defined as 100%. White columns, AA817769; striped columns, AA858959. Error bars indicate \pm SD.

time course experiments, to assign functions to genes by extrapolating the functions of characterized genes in a cluster to uncharacterized members of the same cluster (Eisen et al., 1998; Wen et al., 1998). This is a guilt-by-association argument based on the premise that coregulation implies common function, an argument that has been shown to be prone to false assignments (Hughes et al., 2000). In the present study, clustering was used to differentiate cell types and select marker genes. This does not require any assumptions regarding the functions of genes within a certain cluster.

The clustering result in Figure 7 suggested that there were differences even among pyramidal CA1 neurons, and because several genes were found clustered in the same group, that this was a nonrandom event. The cells were all NSE positive with a pyramidal-like morphology and captured from the pyramidal layer of CA1. Therefore, they were most likely pyramidal neurons. The gene expression differences were subtler and not as pronounced as those seen between pyramidal neurons and the parvalbumin-positive interneuron. The maximum difference in microarray signal was at most three- to fourfold. We therefore decided to use real-time quantitative PCR to validate this cluster. The PCR results supported the clustering results and showed that there were differences in gene expression among individual CA1 pyramidal neurons and that at least some of these differences were nonrandom. It has been described that neurons *in vivo* and *in vitro* exhibit differences in gene expression among single cells within a given cell type (Mackler et al., 1992; Sheng et al., 1995; Zawar et al., 1999). Differences between individual cells may be caused by stochastic processes (Elowitz et al., 2002). In the present study we found a group of genes that covaried in expression between individual CA1 pyramidal cells and were able to confirm this finding by RT-PCR. This shows that these differences were nonrandom. There are several possible reasons for nonrandom differences in gene expression between single pyramidal CA1 neurons, e.g., differences in projection targets, afferent input, or participation in place field encoding. It has been shown that selected neurons in CA1 respond to spatial experiences of the rat by changes in gene expression (Guzowski et al., 1999). Another possibility is that dynamic change in gene expression is an intrinsic property of gene regulatory networks (Kauffman, 1993). In any case, the nonrandom differences that we have

found between individual cells may correspond to different functional states. Single-cell gene expression profiling may therefore provide information on functional states that otherwise would be masked by averaging a population of cells.

Eberwine et al. (1992) first suggested using single-cell gene expression profiling to molecularly define cell types. We have now demonstrated the feasibility of using laser capture microdissection and fluorescent microarray analysis, both relatively high-throughput methods, to generate single-cell gene expression profiles and subsequently to molecularly define cell types in the brain on the basis of data clustering. This will be a particularly useful tool when working with complex tissues such as the CNS. With an improvement in throughput, particularly of the T7 amplification system that is currently the overall rate-limiting step, researchers will be able to describe dynamic processes, such as disease progression, in complex tissues with unprecedented resolution.

References

- Augood SJ, Waldvogel HJ, Munkle MC, Faull RL, Emson PC (1999) Localization of calcium-binding proteins and GABA transporter (GAT-1) messenger RNA in the human subthalamic nucleus. *Neuroscience* 88:521–534.
- Braak H, Braak E (1991) Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 82:239–259.
- Brooks-Kayal AR, Jin H, Price M, Dichter MA (1998) Developmental expression of GABA(A) receptor subunit mRNAs in individual hippocampal neurons *in vitro* and *in vivo*. *J Neurochem* 70:1017–1028.
- Danos P, Frotscher M, Freund TF (1991) Non-pyramidal cells in the CA3 region of the rat hippocampus: relationships of fine structure, synaptic input and chemical characteristics. *Brain Res* 546:195–202.
- Delgado-Garcia JM, Lopez-Barneo J, Serra R, Gonzalez-Baron S (1983) Electrophysiological and functional identification of different neuronal types within the nucleus ambiguus in the cat. *Brain Res* 277:231–240.
- Dell D, Raghupathi R, Crino PB, Morrison III B, Eberwine JH, McIntosh TK (1998) Amplification of mRNAs from single, fixed, TUNEL-positive cells. *Biotechniques* 25:566–570.
- Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, Zettel M, Coleman P (1992) Analysis of gene expression in single live neurons. *Proc Natl Acad Sci USA* 89:3010–3014.
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868.
- Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
- Gedeck P, Willett P (2001) Visual and computational analysis of structure-activity relationships in high-throughput screening data. *Curr Opin Chem Biol* 5:389–395.
- Grofova I (1975) The identification of striatal and pallidal neurons projecting to substantia nigra. An experimental study by means of retrograde axonal transport of horseradish peroxidase. *Brain Res* 91:286–291.
- Guzowski JF, McNaughton BL, Barnes CA, Worley PF (1999) Environment-specific expression of the immediate-early gene *Arc* in hippocampal neuronal ensembles. *Nat Neurosci* 2:1120–1124.
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburty K, Simon J, et al. (2000) Functional discovery via a compendium of expression profiles. *Cell* 102:109–126.
- Kauffman SA (1993) *Origins of order*. New York: Oxford UP.
- Kosaka T, Katsumaru H, Hama K, Wu JY, Heizmann CW (1987) GABAergic neurons containing the Ca^{2+} -binding protein parvalbumin in the rat hippocampus and dentate gyrus. *Brain Res* 419:119–130.
- Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG (1999) Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* 5:117–122.
- Mackler SA, Brooks BP, Eberwine JH (1992) Stimulus-induced coordinate changes in mRNA abundance in single postsynaptic hippocampal CA1 neurons. *Neuron* 9:539–548.

- Mardia KV, Kent JT, Bibby JM (1979) *Multivariate analysis*. London: Academic.
- Peruzzi D, Sivaramakrishnan S, Oliver DL (2000) Identification of cell types in brain slices of the inferior colliculus. *Neuroscience* 101:403–416.
- Ramon Y Cajal S (1884) *Manual de Histología normal y técnica micrográfica*. Valencia: Aguilar.
- Sheng HZ, Lin PX, Nelson PG (1995) Combinatorial expression of immediate early genes in single neurons. *Brain Res Mol Brain Res* 30:196–202.
- Vandesande F, Dierickx K (1975) Identification of the vasopressin producing and of the oxytocin producing neurons in the hypothalamic magnocellular neurosecretory system of the rat. *Cell Tissue Res* 164:153–162.
- Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 87:1663–1667.
- Wen X, Fuhrman S, Michaels GS, Carr DB, Smith S, Barker JL, Somogyi R (1998) Large-scale temporal gene expression mapping of central nervous system development. *Proc Natl Acad Sci USA* 95:334–339.
- Wilcox BJ, Unnerstall JR (1990) Identification of a subpopulation of neuropeptide Y-containing locus coeruleus neurons that project to the entorhinal cortex. *Synapse* 6:284–291.
- Zawar C, Plant TD, Schirra C, Konnerth A, Neumcke B (1999) Cell-type specific expression of ATP-sensitive potassium channels in the rat hippocampus. *J Physiol (Lond)* 514:327–341.