In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus

Highlights

- Amplified seqFISH enables in situ detection of hundreds of genes in single cells in tissues

- Combinatorial expression patterns of genes define cell classes in the mouse brain

- Spatial transcriptomics defines regions within the hippocampus

- Heterogeneity in cell class compositions increases along the dorsal to ventral axis

Authors

Sheel Shah, Eric Lubeck, Wen Zhou, Long Cai

Correspondence

lcai@caltech.edu

In Brief

Shah et al. demonstrated multiplexing of 250 genes in situ in mouse brain slices using amplified seqFISH. They found that there are distinct subregions of the hippocampus consisting of different combinations of cell classes, defined by the expression patterns in single cells.

In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus

Sheel Shah,[2,3,4] Eric Lubeck,[2,4] Wen Zhou,[1] and Long Cai[1,5,*]
[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[3]UCLA-Caltech Medical Scientist Training Program, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, USA
[4]Co-first author
[5]Lead Contact
*Correspondence: lcai@caltech.edu
http://dx.doi.org/10.1016/j.neuron.2016.10.001

SUMMARY

Identifying the spatial organization of tissues at cellular resolution from single-cell gene expression profiles is essential to understanding biological systems. Using an in situ 3D multiplexed imaging method, seqFISH, we identify unique transcriptional states by quantifying and clustering up to 249 genes in 16,958 cells to examine whether the hippocampus is organized into transcriptionally distinct subregions. We identified distinct layers in the dentate gyrus corresponding to the granule cell layer and the subgranular zone and, contrary to previous reports, discovered that distinct subregions within the CA1 and CA3 are composed of unique combinations of cells in different transcriptional states. In addition, we found that the dorsal CA1 is relatively homogeneous at the single cell level, while ventral CA1 is highly heterogeneous. These structures and patterns are observed using different mice and different sets of genes. Together, these results demonstrate the power of seqFISH in transcriptional profiling of complex tissues.

INTRODUCTION

The mouse brain contains $\sim 10^8$ cells arranged into distinct anatomical structures. While cells in these complex structures have been traditionally classified by morphology and electrophysiology, their characterization has been recently aided by gene expression studies. In particular, the Allen Brain Atlas (ABA) provides a systematic gene expression database using in situ hybridization (ISH) of the entire mouse brain one gene at a time (Dong et al., 2009; Fanselow and Dong, 2010; Thompson et al., 2008). This comprehensive reference provides regional gene expression information but lacks the ability to correlate the expression of different genes in the same cell. More recently,

single-cell RNA sequencing (RNA-seq) has identified many cell types based on gene expression profiles (Darmanis et al., 2015; Tasic et al., 2016; Zeisel et al., 2015). However, while single-cell RNA-seq provides useful information on multiple genes in individual cells, it has relatively low detection efficiencies and requires cells to be removed from their native environment, resulting in the loss of spatial information. These different approaches can lead to contradictory descriptions of cellular organization in the brain and other biological systems.

In the hippocampus, recent RNA-seq data suggest that the CA1 region is composed of cells with a continuum of expression states (Cembrowski et al., 2016; Zeisel et al., 2015), while ABA analysis indicates that sub-regions within the CA1 have distinct expression profiles (Thompson et al., 2008). To resolve the two conflicting descriptions of hippocampal organization, a method to profile transcription in situ in the hippocampus with single-cell resolution is needed. Here, we demonstrate a general technique that enables the mapping of cells and their transcription profiles with single-molecule resolution in tissue, allowing unprecedented resolution of cellular transcription states for molecular neuroscience (Figure 1A).

A great deal of progress has been made recently in developing highly quantitative methods to profile the transcriptome of single cells. Building upon single-molecule fluorescence in situ hybridization (smFISH) (Femino et al., 1998; Raj et al., 2006), Lubeck and Cai (2012) devised a general method to highly multiplex single-molecule in situ mRNA imaging regardless of transcript density using super-resolution microscopy (Betzig et al., 2006; Rust et al., 2006; Lubeck and Cai, 2012). However, the spectral barcoding methods used in these previous works is difficult to scale up beyond 20–30 genes because of the limited number of fluorophores (Fan et al., 2001; Lubeck and Cai, 2012).

To overcome the scalability problem, a temporal barcoding scheme was developed that uses a limited set of fluorophores and scales exponentially with time (Lubeck et al., 2014). Specifically, sequential probe hybridizations on the mRNAs in fixed cells impart a unique pre-defined temporal sequence of colors, generating in situ mRNA barcodes. The multiplex capacity scales as F^n, where F is the number of fluorophores and n is the number of rounds of hybridization. Thus, one can increase the multiplex
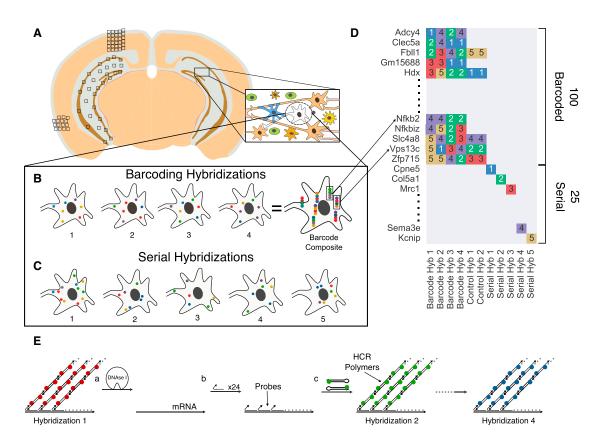
Figure 1. Overview of Sequential Barcode FISH in Brain Slices

(A) A coronal section from a mouse brain was mounted on a slide and imaged in all boxed areas. Each image was taken at 60× magnification.

(B) Example of barcoding hybridizations from one cell in a field from (A). The same points are re-probed through a sequence of 4 hybridizations (numbered). The sequence of colors at a given location provides a barcode readout for that mRNA ("barcode composite"). The genes represented by these barcodes are identified through referencing a lookup table abbreviated in (D) and quantified to obtain single-cell expression levels. In principle, the maximum number of transcripts that can be identified with this approach scales to F^n, where F is the number of fluorophores and n is the number of hybridizations. Error correction adds another round of hybridization.

(C) Serial smHCR is an alternative detection method in which 5 genes are quantified in each hybridization and this process is repeated n times. Serial hybridization scales as $F*n$.

(D) Schematic for multiplexing 125 genes in single cells. 100 genes are multiplexed in 4 hybridizations by seqFISH barcoding. This barcode scheme is tolerant to loss of any round of hybridization in the experiment. 25 genes are serially hybridized 5 genes at a time by 5 rounds of hybridization. Each number represents a color channel in smHCR. As a control, 5 genes are measured by both double rounds of smHCR as well as barcoding in the same cell.

(E) SmHCR amplifies signal from individual mRNAs. After imaging, DNase strips the smHCR probes from the mRNA, enabling rehybridization on the same mRNA (step a). The "color" of an mRNA can be modulated by hybridizing probes that trigger HCR polymers labeled with different dyes (step b). Signals from the mRNAs are amplified following hybridization by adding the complementary hairpin pair (step c). The DNase smHCR cycle is repeated on the same mRNAs to construct a predefined barcode over time.

capacity by increasing the number of rounds of hybridization with a limited pool of fluorophores. We called this approach sequential fluorescence in situ hybridization (seqFISH) (Lubeck et al., 2014). In parallel, in situ sequencing methods were developed to directly sequence transcripts in tissue sections, but these methods suffer from low detection efficiency (<1%) (Ke et al., 2013; Lee et al., 2014). Recently, Chen et al. (2015b) expanded the error correction method in the original seqFISH demonstration by using a Hamming distance 2 based error correcting barcode system called MERFISH. However, this implementation requires larger transcripts (>6 kb) and many more rounds of hybridization than the method described here (Chen et al., 2015b). Furthermore, seqFISH and its variants have only previously been applied in cell culture systems due to the difficulty of smFISH detection in

tissue. Here, we demonstrate an improved version of seqFISH in complex tissues by including signal amplification (Shah et al., 2016) and a time-efficient error correction scheme (Figures 1A–1D and Table S1), allowing us to resolve the structural organization of the hippocampus with single-cell resolution.

RESULTS

Signal Amplification and Error Correction Enable Robust Detection of mRNAs in Tissues

To overcome the autofluorescence and scattering inherent to brain tissues, we used an amplified version of smFISH, called single-molecule hybridization chain reaction (smHCR) (Figures 1E and S1A) (Choi et al., 2014; Shah et al., 2016). smHCR
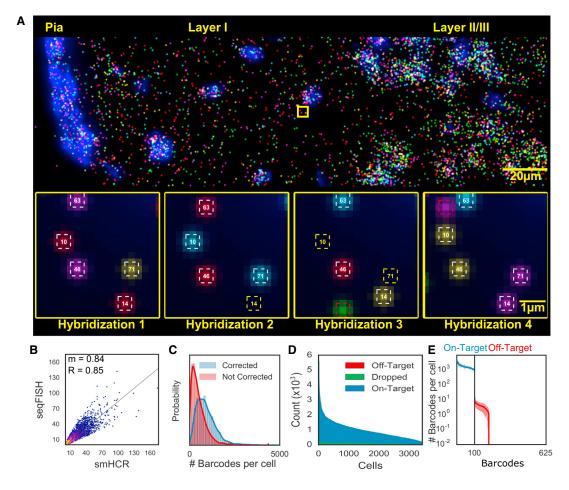
Figure 2. SeqFISH Generates Accurate In Situ Quantification of mRNA Levels

(A) Image of seqFISH barcoding 100 genes in the outer layer of the mouse cortex. RNA dots in the image are z projected over 15 μm. Individual mRNA points are shown across 4 hybridizations in the inset images. White squares correspond to identified barcodes, yellow squares correspond to missing transcripts in a particular hybridization, and red squares correspond to spurious false positives and are not counted in any barcode measurements. Numbers in the squares correspond to barcode indices.

(B) seqFISH correlates with smHCR counts. Five target mRNAs were measured by both barcoded seqFISH and smHCR in the same cells, providing absolute counts of the transcripts. The two techniques correlate with an R = 0.85 and a slope (m) of 0.84 (n = 3851 measurements). The 2D histogram intensity shows the distribution of points around the regression line. A high density of points is seen along the regression line.

(C) Error correction results in a median gain of 373 (25%) counts per cell (n = 3,497). Red and blue curves correspond to the total barcode counts per cell before and after error correction.

(D) Dropped and off-target barcodes represent a small source of error in seqFISH. 100 on-target barcodes and 525 off-target barcodes are measured per cell. Dropped barcodes arise from ambiguous cases in which two or more overlapping dots appear within the same region.

(E) Off-target barcodes are rarely observed and contribute minimally to the expression profile in single cells. Each of the 100 on-target barcodes (blue) and 525 off-target barcodes (red) are quantified per cell. The mean is shown with shaded regions corresponding to 1 SD (n = 41 imaged regions).

amplified signal 22.1 ± 11.5-fold (mean ± SD, n = 1338; Figure S1B) compared to smFISH, enabling robust and rapid detection of individual mRNA molecules in tissues and facile alignment of spots between hybridizations (Figure 2A). Single transcripts can be detected and localized in 3D with just 24 probes in tissues, enabling detection of transcripts < 1 kb in size, with a fidelity comparable to the smFISH gold standard (Figures S1C and S1D) but with signals 20-fold brighter (Shah et al., 2016). smHCR DNA polymers can also be digested by DNase and the mRNA can be re-hybridized in brain slices, allowing HCR-seqFISH to be robustly implemented (Figure 2A). We note the smHCR enables true 3D imaging in tissues, whereas the previ-

ous sequential FISH demonstrations (Lubeck et al., 2014; Chen et al., 2015b) were performed only in flat cell cultures.

Furthermore, we improved upon our existing barcode system by implementing a time-efficient error correction scheme. The major source of error in seqFISH is the loss of signal due to mis-hybridization. We introduced an extra round of hybridization to correct for loss of signal during any round of hybridization (Figure 1D) (Supplemental Information). For example, using 5 fluorophores and 4 rounds (instead of 3 rounds) of hybridization to code for 125 genes, we can still uniquely assign barcodes to genes even when signal from any single round of hybridization is missing. Although MERFISH can tolerate 2 errors per barcode, it requires

16 rounds of hybridization to code 140 genes (Chen et al., 2015b). This large number of hybridizations can lead to more experimental error and analysis complexity. In comparison, our simple error correction accounts of the most common error, dropped signal, while using a minimal number of hybridizations (4 rounds for 125 genes and 5 rounds for 625 genes). Also, the fewer rounds of hybridizations decrease the total imaging time, which is rate-limiting for tissue experiments. HCR-seqFISH with this simpler error-correction scheme allows efficient and accurate quantification of transcription profiles in tissues.

Using this HCR-seqFISH method, we surveyed the regional and sub-regional transcriptional heterogeneity within the temporal and parietal cortex and hippocampus of the mouse brain by imaging similar coronal sections collected from three different animals. Two similar sections from separate mice were profiled with probes for 125 genes, while one additional brain slice was imaged for 249 genes. In each of the coronal slices, between 60 and 80 fields of view were imaged, each 216 μm × 216 μm × 15 μm, in the cortex and hippocampus (Figures 1A and S1E). For the 125 gene set, 56 of the genes (Figure 1D and Table S1) were selected because they showed spatially heterogeneous expression based on the ABA (Lein et al., 2007), another 44 were selected from a list of transcription factors, and 25 marker genes were selected from single-cell RNA-seq datasets (Zeisel et al., 2015). 100 of these genes were barcoded by 4 rounds of hybridization (Figure 1B). The remaining 25 high-abundance genes were measured individually using 5-color smHCR in 5 serial rounds of hybridizations (Figure 1C). This hybrid approach of measuring medium expression genes with barcoding seqFISH and high copy number genes serially in subsequent hybridizations allows a large dynamic range of transcript levels to be profiled in the same cell.

seqFISH Is an Accurate and Efficient Method to Multiplex RNA In Situ

To determine the accuracy of the seqFISH method in quantifying mRNA levels in single cells in tissue, we compared the copy number of 5 of the 100 target genes measured by barcoding to the copy number found by smHCR detection in the same cell (Figures 2B and S2A) in 15 μm brain sections. We found that the copy number of the RNAs per cell as measured by barcoding and smHCR agreed with an R value of 0.85 and a slope of 0.84 (n = 3,851). As smHCR matches smFISH transcript quantitation (Shah et al., 2016), the barcoded seqFISH method can quantify mRNA molecules in single cells with 84% efficiency compared to the gold standard of smFISH. In comparison, single-cell RNA-seq measurements are 5%–20% efficient based on spike-in controls, and in situ sequencing is less than 1% efficient (Darmanis et al., 2015; Klein et al., 2015; Lee et al., 2014; Macosko et al., 2015; Tasic et al., 2016; Zeisel et al., 2015; Ståhl et al., 2016). This high efficiency of detection results from a low transcript drop rate and a high barcode recovery rate due to the error correction round of hybridization. In our experiment, 78.9% of barcodes (n = 2,115,477 barcodes) were detected in all 4 hybridization rounds, and 21.1% were observed in 3 out of the 4 hybridizations (Figure 2C), indicating that the probability of detecting a given mRNA molecule is 94% in each round of hybridization (Figure S2B).

To quantify the amount of false positive signal due to misalignment of barcodes and nonspecific binding of probes, we measured the amount of off-target barcodes that were detected. With four rounds of hybridizations and 5 fluorophores, there were $5^4 = 625$ unique codes. The 100 barcodes assigned to detect mRNAs were observed in aggregate at 914.8 ± 570.5 counts per cell (mean ± SD, n = 3,439). In comparison, the 525 remaining off-target barcodes that were not used were detected at 4.6 ± 4.7 (mean ± SD, n = 3,439) counts per cell (Figure 2D). False positives, due to chance alignment of nonspecifically bound spots, contributed minimally to the barcode readouts because of this three order of magnitude difference in detected barcodes (on target versus off target). The false positives we observed fall only on barcodes hamming distance one away from on-target barcodes, and minimally contribute to undercounting on-target barcodes (Figure 2E). Furthermore, even the most frequent off-target barcode was observed 65.57 times less frequently than the most infrequent mRNA coding barcode (Figures 2E and S2). Even though during each round of hybridization, 24.8% ± 0.4% (mean ± SE, n = 4 rounds of hybridization) of the spots were nonspecifically bound probes, barcode miss-assignments did not occur frequently because non-specifically bound probes do not reappear in the same location after digestion with DNase and re-hybridization (Figure 2A). Together, the quantifications of false positive and false negative barcodes demonstrate that this method is highly efficient and accurate at detecting RNAs in situ in single cells within tissues.

Cell Clusters Are Based on Combinatorial Expression Profiles

We imaged the expression of 125 genes in coronal sections from two mice for a total of 14,908 cells (Figure S1E). Cortical and hippocampal cells were segmented based on DAPI and Nissl staining. A tessellation algorithm was developed to accurately segment densely packed cells in the hippocampus. To avoid capturing mRNA from neighboring cells, we contracted by 10% the borders of cells determined by the segmentation algorithm.

To group the single cell data into distinct transcriptional states, we Z score normalized the copy number of each transcript in every cell (Figure 3A and Tables S6 and S7) and hierarchically clustered the cells to identify cells with similar expression patterns (Table S2, Figure S3). While these clusters do not necessarily represent canonical cell types, many of these clusters contain clear transcriptional markers of known cell types previously identified by single-cell RNA-seq (Figure 3B) (Zeisel et al., 2015; Tasic et al., 2016). For example, cell clusters 12 and 13 contained clear expression of *Gja1*, which marks out astrocytes (Zeisel et al., 2015; Tasic et al., 2016). Cluster 12 also expresses *Mfge8*, while cluster 13 did not, indicating two distinct subpopulations of astrocytes (Figure 3B). There are further subclusters within each of the astrocyte populations with different spatial localization patterns (Figures S3C–S3E). Cluster 11 cells expressed *Laptm5*, a known microglia marker (Zeisel et al., 2015; Tasic et al., 2016). Cluster 3 expressed interneuron genes, while clusters 1–2 and 4–5 expressed genes associated with pyramidal neurons (Zeisel et al., 2015; Tasic et al., 2016). The major clusters were robust to down-sampling
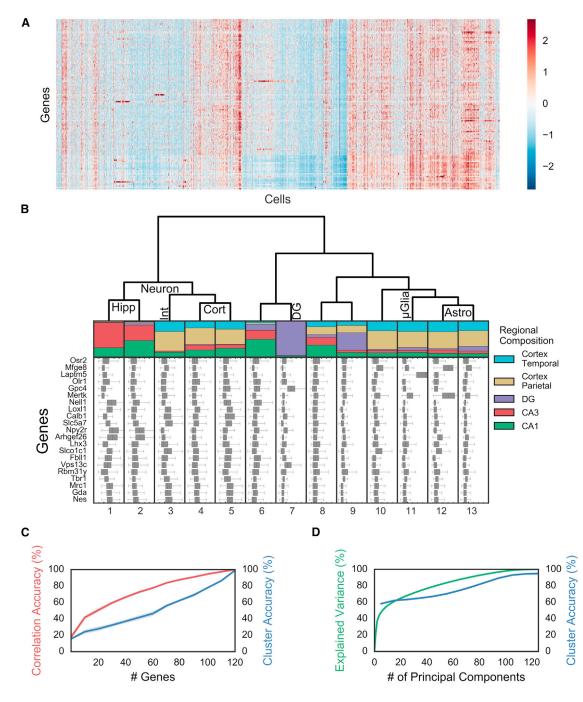
Figure 3. Distinct Clusters of Cells Exhibit Different Regional Localization in the Brain

(A) Gene expression of 14,908 cells presented as a Z score normalized heatmap.

(B) Regional compositions of 13 cell clusters are visualized as stacked bar plots with the area corresponding to the number of cells in each region. Hippocampal regions are: CA3, CA1, dentate gyrus (DG). Cortical regions: parietal and temporal. Boxplot of the Z scores of 21 representative genes are plotted for each cell class. Tick marks correspond to Z score interval of 1, with the second tick mark at Z score 0. Cell-type assignments are shown on the dendrogram. Abbreviations: hippocampus pyramidal, Hipp; cortex, Cort; dentate gyrus, DG; interneurons, Int; astrocytes, Astro; microglia, μGlia.

(C) Any random subset of 25 genes can recapitulate approximately 50% of the information in the correlation among cells (red), but a larger number of genes is required to accurately assign cells to cluster using a random forest algorithm (blue) (n = 10 bootstrap replicates; shading is 95% CI), indicating that fine structures in the data require quantitative measurements of combinatorial expression of many genes.

(D) While the first ten principal components (PCs) explain the coarse structure, a larger number of PCs are required to describe the full data. Explained variation (green) and accuracy in predicting cell identity using a random forest model (blue).

the number of cells used in clustering (Figure S4), with some of the hippocampal pyramidal and glial clusters robustly defined even with 400 cells. Similarly, principal component analysis (PCA) visualization of the data (Figure S3H) recapitulated the major clusters that correspond to astrocyte, microglia, cortical pyramidal, hippocampal pyramidal, dentate gyrus (DG) granule, and interneuron cells.

As the cluster distance between different cells is proportional to the number of differentially expressed genes in the target list, an unbiased clustering of the 125 gene data without weighting specific genes should not be interpreted directly as identifying canonical "cell types," but rather as grouping cells with different patterns of genes expression based on the current target list. We will refer to some of these clusters as pyramidal neurons or astrocytes for ease of notation, but strictly speaking, they are cells clusters with expression patterns similar to those of neurons or astrocytes.

Cell Clusters Show Distinct Regional Localization

Many neuronal clusters mapped to distinct regions in the brain (Figure 3B). Several classes of pyramidal cells (clusters 1–2) showed exclusive localization to the hippocampus, while other classes (4–5) showed predominantly cortical localization. There was also a class of cells (cluster 7) that were almost exclusively present in the DG. Interestingly, these clusters segregated based solely on gene expression profiles without adding any spatial information into the clustering algorithm. These differences in transcriptional states of neurons could be due to intrinsic differences in the cells or due to different local environment and activity patterns.

In contrast, astrocyte, microglia, and other non-neuronal cell clusters were generally uniformly present in all areas of the brain (Figure 3B). However, subclusters of astrocytes did localize to different regions of the brain preferentially (Figure S3E), with subcluster 12.3 localized preferentially to the cortex, while 12.1 subcluster was uniformly distributed. Similarly, cluster 9 cells contain subclusters (9.3, 9.5, and 9.6) that localize exclusively to the DG, while another subcluster (9.1) localizes almost exclusively to the cortex. The regional localization of neurons is especially pronounced with cluster 1 and 2 localized almost exclusively to the hippocampus, with some of the subclusters localized predominantly to the CA3. Furthermore, while pyramidal cell clusters 4 and 5 are preferentially cortically localized, the few hippocampal cells in these clusters form their own subclusters (4.4 and 5.4) (Figure S3E). In cluster 6 cells, many subclusters with distinct expression profiles are localized almost exclusively in the CA1, CA3, or the DG (Figure S3C). In contrast, cluster 7 cells show a relatively homogeneous regionalization pattern but further subdivide based on combinatorial expression patterns (Figure S3D). Subclusters of cluster 9 also show significant regionalization, as subclusters 9.1, 9.3, 9.5, and 9.6 show localization to the subgranular zone (SGZ) (Figure S3E). Overall, cell clusters with similar expression profiles exhibited similar spatial localizations across the brain with a correlation coefficient of 0.67 (Figure S3G), indicating the existence of archetypal regional expression patterns and potential spatial markers in the brain. These results show that the tissue-optimized HCR-seqFISH approach can directly identify a variety of transcriptional states and quantify broad spatial patterns of expression.

Combinatorial Expression Patterns Define Fine Clusters

While certain cell clusters contain strong expression of marker genes, not all clusters are defined based on a few genes. How much power do individual genes or groups of genes have in explaining the observed cell clusters? To understand this, we examined whether subsets of genes can recapitulate the observed clusters (Figures 3C and 3D). We found that any set of 25 genes recovers about half of the correlation structure in the cell-to-cell correlation map (Figures 3C and S3I; n = 10 bootstrap replicates). The fact that the selection of any 25 genes can explain the gross patterns in the data is likely due to the high correlations among the expression patterns of genes, as shown in the gene-to-gene correlation map (Figure S3J). Thus, a small subset of the measured genes can provide sufficient information to infer the gross transcriptional states of the cells. Similarly, low-coverage single-cell sequencing methods such as drop-seq and inDrop (Klein et al., 2015; Macosko et al., 2015) may capture major categories of cell types because many highly expressed, detected genes are correlated to other genes that collectively define cell types.

At the same time, the finer correlation structure in the data, required to define the cell clusters accurately, can only be captured with accurate quantitation of many genes (Figures 3C and 3D). Consistent with this, using a "random-forest" machine-learning algorithm (Breiman, 2001) to classify cell clusters, we found that 75 genes are needed to classify cells with 50% accuracy, indicating that correct cluster assignment requires more detailed information from many genes (Figure 3C). Supporting this view, the first 10 principal components (PCs) explained 59.5% of the variation in the data, while the rest of the variation required the remaining 115 PCs (Figures 3D and S3F). The "random forest" algorithm required 10 PCs to predict the cell cluster assignments with approximately 60% accuracy (Figure 3D), but accuracy steadily increased with more PCs. These observations indicated two levels of information in the data: a coarse level, where large distinctions in cell clusters are detectable with a few genes, and a fine level, where subtle distinctions require many more genes.

These results suggest two points experimentally. First, multiplexing at the level of 20 genes by seqFISH can give broad cell cluster identification that is not available with 2–3 gene smFISH experiments. Although single marker genes are useful for inference, we find that they frequently are not sufficient for cell classification. For example, all DG specific granule cells (cluster 7) have *Gpc4* and *Vps13c* as their enriched marker genes (Figure 3B); yet, *Gpc4* and *Vps13c* are also strongly expressed in other hippocampal cells outside of the DG, as seen in both our experiments and the ABA. Thus, smFISH against *Gpc4* and *Vps13c* alone would not be sufficient to uniquely identify the DG granule cells. Furthermore, even the strongly bimodal markers that are known to define cell types (i.e., *Mgfe8*, *Gja1*, etc.) are correlated enough to overall expression profiles that cells fall into the appropriate cluster even when these genes are excluded. This point suggests that while marker genes can be essential in assigning a cell to a known cell type, they are not necessary to identify unique clusters in the dataset, provided enough measurements are made.
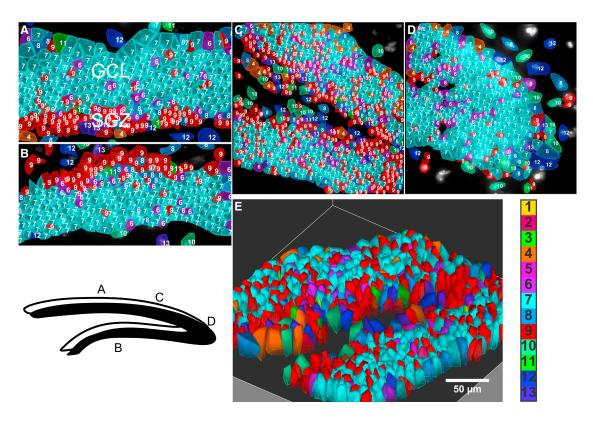
Figure 4. Spatial Layering of Cell Classes in the Dentate Gyrus

Bottom left panel: cartoon of DG with imaged regions labeled. Color key corresponds to classes in Figure 3B.

(A and B) Suprapyramidal (A) and infrapyramidal (B) blades of DG. Cells of the subgranular zone and granule cells are arranged in lamina layers in mirror symmetric patterns on the upper and lower blades.

(C) The SGZ stays on the inner layer of the DG fork.

(D) Cells are patterned in the crest.

(E) 3D image of the fork region shown in (C).

Second, accurate measurement of combinatorial expression of many genes enabled by seqFISH can allow for more specific cell cluster identification. As a comparison, in single-cell RNA-seq data, CA1 pyramidal cells are clustered into a single cluster (Zeisel et. al, 2015; Habib et al., 2016), potentially because of the relatively lower detection efficiency. In our seqFISH experiments, measuring hundreds of genes quantitatively, we can resolve several clusters and subclusters with robust regionalization within the CA1 (Figure 3B and S3C–S3E).

Cells Are Patterned in the Dentate Gyrus

To further visualize the spatial organization of cells, we mapped cluster definitions of cells back into the images. In the DG, we observed a striking lamina layering of cell classes. The two blades of the DG (Figures 4A and 4B) showed mirror arrangements of cells, with cluster 9 cells, forming the subgranular zone (SGZ), leading into a granule cell layer (GCL) dominated by a single cluster of granule cells (cluster 7) (Figure 3B). In the 125 gene dataset, the cells of the GCL were found to be dominated by expression of *Gpc4* and *Vps13c* matching ISH data from the ABA (Figure S8B). Cluster 7 was found to be further subdivided into 6 subclusters (Figure S3D). These sub-

clusters were found to have varying levels of calbindin D-28K (*Calb1*) expression, which is known to increase with granule cell maturation (Yang et al., 2015). On the other hand, the cells of the SGZ were found to be significantly enriched in astrocyte markers such as *Mfge8* and *Mertk*, which has also been observed previously (Miller et al., 2013) and in the ABA data (Figure S8A). However, these cells do not cluster with typical astrocytes (clusters 12 and 13) because their combinatorial expression patterns are different from astrocytes, consistent with their classification as a completely different population of cells.

In the fork region of the DG, the layer of cluster 9 cells appeared on the interior surface of the fork, followed by a layer of granule cells (cluster 7) (Figure 4C). A different layering pattern is seen at the crest of the DG, where astrocytes, microglia, and some other glial cells line the exterior of the crest ensheathing the GCL (Figure 4D). In both brains of the 125 gene experiments, the same cell clusters and spatial arrangements are observed. Furthermore, because the mRNAs are imaged in 3D in the 10–15 μm brain slices, we can obtain a 3D view of the expression profiles, shown in the fork regions of the DG (Figure 4E).
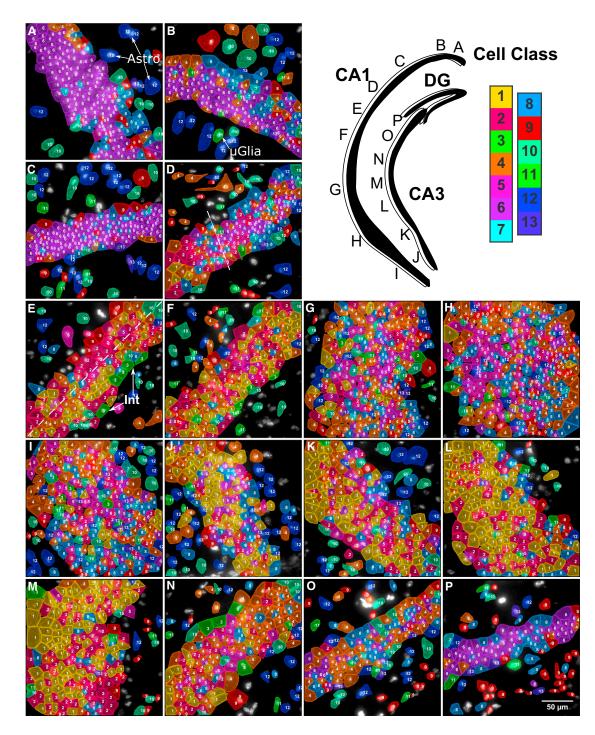
Figure 5. Subregions of the Hippocampus Are Composed of Distinct Compositions of Cell Classes

Upper right panel: cartoon of hippocampus with imaged regions labeled. Color key corresponds to the classes in Figure 3B. Images are from a single coronal section, and cell classes are based on the expression levels of 125 genes.

(A–D) These images are regions from the CA1d. Astrocytes (Astro) are marked in (A), and a microglia cell (μGlia) is also marked (B). Moving along the hippocampus from CA1 dorsal to ventral, cell classes transition from a homogeneous dorsal population (C) to a mixed population in the CA1 intermediate. The transition between the CA1d and CA1i is marked by a dotted line (D).

(E and F) These images are regions from the CA1i. A line has been drawn to show 2 laterally segregated cell classes in the CA1i (E) and arrows indicate cholinergic interneurons (Int) on the interior surface of the CA1i (E). This lateral segregation of cell classes persists throughout the CA1i (F).

(G–I) These images are regions from the CA1v. The largest amount of cellular diversity in the CA1 is seen in the ventral region starting with (G). The cell-type composition of the CA1v region remains the same throughout the region (H) down to the ventral-most area imaged (I).

(legend continued on next page)

Distinct Regions of CA1 and CA3 Are Composed of Different Combinations of Cell Clusters

While each region of the DG contains similar compositions of cells, distinct subregions within the CA1 and CA3 contained different combinations of cell classes (Figures 5 and S6F). In the CA1, there were 3 distinct regions defined by their individual cellular compositions. In the dorsal region of CA1 (CA1d), neuron cluster 6 (enriched in *Nell1*, a protein kinase C binding protein) (Table S2) was the major cell class in the pyramidal layer, with astrocyte, microglia, and other cells (clusters 10–13) intercalating into the stratum pyramidale (SP) (Figures 5A–5C). Transitioning into the CA1 intermediate region (CA1i) (Figure 5D), pyramidal cell cluster 4 displaced cell cluster 6 as the dominant cell, with the co-appearance of cluster 1 and 2 pyramidal cells.

As the middle of the CA1i region was reached, a small amount of cluster 4 pyramidal cells remained, while cluster 1 and 2 pyramidal cells dominated (Figures 5E and 5F). Clusters 1 and 2 are enriched in *Nell1* (EGF-like protein), *Npy2r* (neuropeptide Y receptor), *Slc4a8* (sodium bicarbonate transporter), and *B3gat2* (glucuronosyltransferase) (Table S2). The CA1i region displayed a characteristic spatial organization where glial cells line the outermost regions, while pyramidal cell clusters 1 and 2 longitudinally partitioned the pyramidal layer. This separation of the inner versus the outer layers of CA1 matches that observed previously (Dong et al., 2009). Furthermore, interneurons (cluster 3) were found to preferentially line the inner edge of the pyramidal layer in the CA1i region (Figures 5E and 5F). This patterning of interneurons, particularly subcluster 3.1 cells, which were enriched in *Slc5a7*, a choline transporter, was consistent with the patterning of cholinergic interneurons observed with ChAT-GFP labeling (Yi et al., 2015). Finally, the largest amount of heterogeneity in the CA1 was seen in the ventral CA1 region (CA1v), where cell clusters 3, 5, and 10 began to mix in with clusters 1 and 2 (Figures 5G–5I).

Similarly, the CA3 was found to have transcriptionally distinct regions with different pyramidal cell compositions and abrupt transitions. The ventral most region of CA3 contained a high level of heterogeneity of pyramidal cell clusters (Figures 5J and 5K), while the intermediate region of CA3 contained a mixture of cell clusters 1 and 2 (Figures 5L and 5M). As the CA3 progressed toward the hilus of the DG, the cell types transitioned first to primarily cluster 4 neurons (enriched in *dcx*, doublecortin, and *Col5a1*, a collagen) (Figures 5N and 5O) and then to almost exclusively cluster 6 neurons in the region most proximal to the DG hilus (Figures 5O and 5P). It is interesting to note that while cluster 6 cells appear in both the CA1 (subcluster 6.8) and CA3 (subclusters 6.1 and 6.4), sub-clusters of 6 show distant regional localization (Figure S3E), suggesting that the gene expression differences in CA1 and CA3 cells are captured in the seqFISH data. We note that similar patterns of homogeneous dorsal and heterogeneous ventral cell populations are observed when only hippocampal cells are clustered (Figure S5).

The regionalized expression patterns we observed in the hippocampus match closely to those observed in previous literature (Thompson et al., 2008; Dong et al., 2009). For example, CA1d, CA1i, and CA1v boundaries correspond to the boundaries shown in Figure 2B in Dong et al. (2009). In CA3, the subregions observed in our experiment match the CA3 subregion 4-7 in Thompson et al. (2008).

Lastly, we note that the brain slices from the two different mice probed with the 125 gene set show not only the same subregional structure (Figures 4, 5, and 6), but also the same clusters of cells (Figure 5 and 6) in the different subregions of the hippocampus (Figure S6F). In both brains, the CA1d consists of a relatively homogeneous population of cluster 6 cells, which transitions to a mixture of 1 and 2 cells in CA1i, and finally to a mixture of 1-6 and 10 cells in the CA1v (Figure S6F). These results together show that the subregions of the hippocampus are a robust feature in the organization of CA1 and CA3, consisting of cell classes with distinct expression profiles. The stereotypical nature of the spatial arrangement of these structures suggests further experiments with seqFISH and other functional assays to probe the distinct functions of the different cell clusters in the CA1 and CA3.

249 Gene Multiplex Experiments Show the Same Hippocampal Subregions

To further show that the sub-regional structure of the hippocampus is independent of the target genes, we performed a 249 gene seqFISH experiment on a third coronal section (Table S8). Of these 249 genes, only 22 genes overlapped with the 125 gene experiment set. For this set of genes, 214 were selected from a list of transcription factors and signaling pathway components, and the remaining 35 were selected from cell identity markers from another single-cell RNA-seq dataset (Tasic et al., 2016). The 214 genes were barcoded by 5 rounds of hybridization, while the remaining genes were imaged in 7 rounds of non-barcoding serial hybridization. To quantify the efficiency of this experiment, 4 genes in the barcoding set (*Smarca4*, *Sin3a*, *Npas3*, and *Neurod4*) were re-probed with smHCR. The barcoding efficiency of the 249 gene probe set was found to be 71% with an R value of 0.80 (Figure S6D). In single cells, we detect on average 2,807 ± 1,660 (mean ± SD, n = 2,050 cells) total barcodes.

The same arrangement in the DG was observed in the 249 gene experiment, despite different genes used, indicating robust identification of the layering in the DG by seqFISH (Figures 7S and 7T). In particular, the cells in the SGZ are clustered independently from cells in the GCL, similar to the layers observed in the 125 gene experiment. In the SGZ cells, we observed enrichment of *Sox11*, a key transcription factor in neurogenesis (Miller et al., 2013). Other transcription factors involved in neurogenesis, *NFIA* and *Tbr1* are also enriched in the SGZ cells as seen in our data and the ABA images (Figure S8A). The observations of this distinct layer in both the 249 and 125 gene experiments and

---

(J–M) These images show regions from the ventral and intermediate CA3. Similar cell class compositions are found in the ventral-most field of the CA3 (J) and the adjacent field (K). Field (L) shows similar cell composition as field (M) in the intermediate regions of the CA3.

(N–P) The following field (N) shows a slightly different cell class composition from the previous fields in the CA3. The next field (O) also shows another change in the cell-type composition. A final transcriptional region in the CA3 is seen in the dorsal-most region of the CA3 (P) with a cell class composition similar to that of the CA1d, but these cluster 6 cells are grouped into a distinct subcluster.
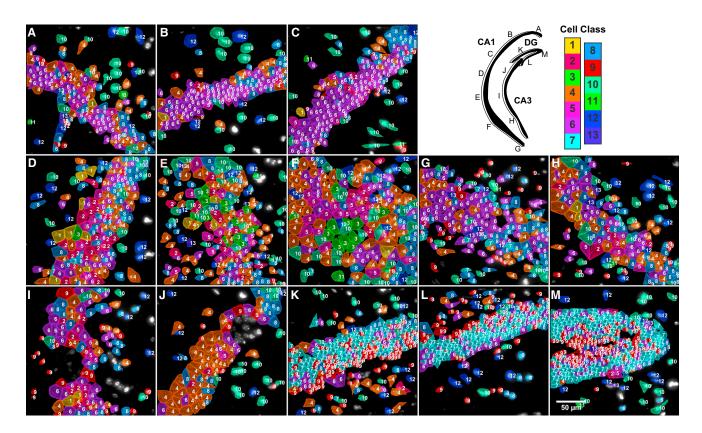
Figure 6. Mapping of Cell Types to a Second Brain Slice Based on Expression Levels of 125 Genes

Upper right panel: cartoon of hippocampus with imaged regions labeled. Color key corresponds to the classes in Figure 3B.

(A–C) Similar to the cell class compositions shown for the hippocampus in Figure 5, CA1d in this coronal section from a second mouse is composed of mostly cluster 6 cells. The dorsal-most region of the CA1 (A) shows almost exclusively class 6 cells. This homogeneity continues throughout the CA1d (B) and ends in the last field imaged in the CA1d (C).

(D–G) The cell class composition changes in the CA1i, shown in (D) and (E) and shifts to a heterogeneous profile in the CA1v (F) before the CA1v terminates (G).

(H–J) The ventral-most imaged region of the CA3 (H) and the adjacent region in the CA3 (I) form a transcriptional territory similar to that in Figure 5. The dorsal-most imaged region in the CA3 (J) shows another distinct transcriptional territory within the CA3.

(K–M) DG regions show the same cell classes in the suprapyramidal blade (K), infrapyramidal blade (L), and crest (M) of DG and layering pattern of the GCL and SGZ as shown in Figure 4.

the combined gene enrichment pattern (increased *Sox11*, *Sox9*, *NFIA*, and *Tbr1* in the 249 gene experiment and increased *Mertk* and *Mfge8* in the 125 gene experiment) suggest that many cells in this layer are involved in adult neurogenesis in the SGZ. Figure S7B shows distinctive marker gene expression in the GCL of the dentate gyrus.

In addition, the same regionalized cellular patterns are observed in CA1d, CA1i, and CA1v, where different subregions utilize different cell classes in characteristic ratios (Figure S6G). As seen with the 125 gene experiment, while the CA1d contains only a few cell classes and is relatively homogeneous, the CA1v region is made up of many different cell classes resulting in a high level of cellular heterogeneity. Furthermore, the distinction between CA1 and CA3 cell clusters is clearer in the 249 gene experiment suggesting more resolving power of spatial patterns (Figures 7A–7K). The 249 gene experiment also suggests that the CA3 may be composed of 3–4 subregions based on cell cluster composition (Figures 7L–7R). The cellular heterogeneity of the CA3 is again shown to mirror that of the CA1, where the cellular

heterogeneity increases along the dorsal to ventral axis. Cells with distinctive marker gene expression in the hippocampus are shown in Figure S7A.

DISCUSSION

Single-Cell Data Resolve Cellular Organizations in the Subregions of the CA1 and CA3

Two conflicting views of the cell types in the hippocampus have been proposed based on the analysis of the Allen Brain Atlas data (Thompson et al., 2008) as well as recent RNA-seq data (Cembrowski et al., 2016; Zeisel et al., 2015). Analysis of the ABA in situ data showed that distinct subregions of the hippocampus expressed different molecular markers, indicating that the CA1 and CA3 are "regionalized" into distinct substructures (Fanselow and Dong, 2010; Thompson et al., 2008). However, recent bulk RNA-seq experiments on the CA1 found that gene expression patterns changed gradually along the dorsal to ventral axis, contradicting the sharp boundaries observed in

Neuron 92, 342–357, October 19, 2016 351

the ABA analysis (Cembrowski et al., 2016). Further supporting this "continuous" cell type view of the hippocampus, analysis of the single-cell RNA-seq data (Zeisel et al., 2015) identified a single continuous population of cells in the CA1 region.

Our data provide a single-cell resolution picture of the spatial organization of cells in the hippocampus and reconcile both the RNA-seq and the ABA data. While our data mostly supports a regionalized view of the hippocampus, we observe that a single cell class does not in general define CA1 and CA3 subregions. Instead, we observed that different subregions of CA1 and CA3 are composed of distinct combinations of cell clusters (Figures 5, 6, and 7). For example, CA1d consists primarily of cluster 6 pyramidal cells (Figures 5A–5C), in addition to the cluster 1,2, 10, and 12 cells, while CA1v consists of a large set of cell classes including cluster 1–6 and 10 cells, but at different relative abundances (Figures 5, 6, S6F, and S6G). Due to this intermixing of cell classes in each sub-region, a bulk measurement of transcription profiles would find a lack of regionalization, but single-cell analysis with spatial resolution would identify these distinct regions based on their unique cell class compositions. Indeed, when we averaged the single-cell expression profile within each subregion of the CA1, we can reproduce the continuous correlation profiles found by bulk RNA-seq between CA1v, CA1i, and CA1d (Figure 8) (Cembrowski et al., 2016). The bulk RNA-seq observation that CA1i lacked specific marker genes can also be explained. This is in fact consistent with our findings that CA1i contained cell classes present in both CA1d and CA1v (Figures 5, 6, and 7). This organization of cell classes is observed in both the 125 gene experiments as well as in the 249 gene experiment.

It is worth noting that the complexity of cell populations observed in the CA1d versus the CA1v matches the functional differences in CA1. CA1d is responsible for spatial learning and navigation and contains a higher concentration of place cells and sends projections to dorsal subiculum and cortical retrosplenial area (Cenquizca and Swanson, 2007; Jung et al., 1994; Risold and Swanson, 1996; O'Keefe and Dostrovsky, 1971). We observed that CA1d is composed of a relatively homogeneous population of cells, predominantly of cluster 6 cells. In contrast, the ventral region is involved in a variety of cognitive tasks, such as stress response and emotional and social behavior (Cenquizca and Swanson, 2007; Jung et al., 1994; Fanselow and Dong, 2010; Kishi et al., 2006; Muller et al., 1996; Petrovich et al., 2001; Pitkänen et al., 2000; Saunders et al., 1988; Witter and Amaral, 1991; Yi et al., 2015). Correspondingly, we observed a large set of cell classes in the CA1v regions. It is intriguing to hypothesize that the different cell classes identified based on molecular profiles may correspond to neurons with distinct connectivity and functional patterns. This hypothesis can be investigated in future experiments combining anterograde tracing as well as electrophysiological recording followed by seqFISH.

SeqFISH Cell Classes versus Single-Cell RNA-Seq Cell Types

RNA-seq measurements at the whole-transcriptome level define cell types based on highly variable genes. On the other hand,
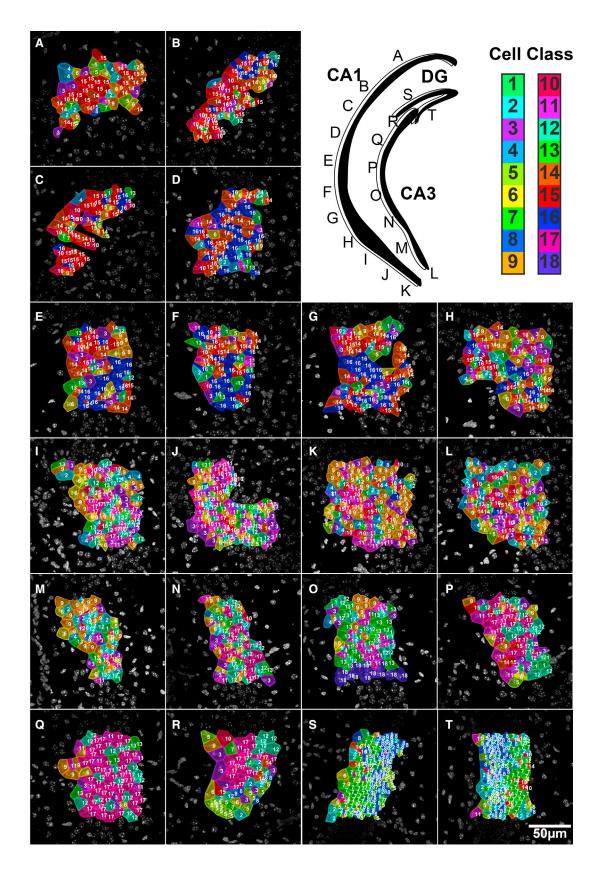
seqFISH provides highly accurate measurements of fewer genes but uses the combinatorial expression patterns to group cells into clusters. However, because only 100–200 genes are targeted in the seqFISH experiments, not all of the "cell types" are equally represented in the gene list and seqFISH cannot catalog "cell types" in the same fashion that single-cell RNA-seq can. For example, in our 125 gene experiments, we cannot resolve the distinct subpopulation of interneurons because we lacked marker genes such as *Vip* and *Sst*. seqFISH and RNA-seq provide two different, yet complementary, levels of resolution into the transcriptional profiles of cells. RNA-seq measures the transcription levels of thousands of genes, but at a lower quantitative accuracy, while seqFISH measures only hundreds of genes, but with much greater quantitative power. The differing nature of the two sets of data informs how the data should be analyzed and interpreted. Thus, seqFISH and single-cell RNA-seq have complementary roles in elucidating distinct cell subpopulations in tissues. SeqFISH could be applied to find finer distinctions within cell types found by RNA-seq or to look at the spatial patterning of cell types found by RNA-seq.

SeqFISH Provides a Generalized Method to Multiplex mRNA Imaging in Tissues

seqFISH with amplification and error correction provides a highly quantitative method to profile hundreds of mRNA species directly in single cells within their native anatomical context. Our method of stripping the probes from the RNA has many advantages. DNase digestion of probes allows false positives to be rejected, as nonspecifically bound probes do not colocalize between different rounds of hybridization (Figure 2A). In addition, the same region of the transcript can be hybridized in every round, allowing seqFISH to efficiently target mRNAs shorter than 1 kb, enabling targeting of most genes. Lastly, seqFISH allows exponential scaling of barcode numbers; thus, 4–5 rounds of hybridization can code for hundreds of transcripts with a simple error correction scheme. Theoretically, the entire transcriptome can be coded for with error correction by using 8–9 rounds of hybridization with seqFISH. These advantages of HCR-seqFISH allow robust multiplexed RNA detection in tissues, shown here in the mouse brain.

Ultimately, the multiplexing capability of seqFISH is limited by the amount of optical space within a cell, and not by the coding capacity of the method (Supplemental Information). We showed previously that super-resolution microscopy can significantly increase the optical space available in the cell for transcription profile imaging, but super-resolution microscopy experiments proved difficult to image in samples thicker than 1 μm and were experimentally cumbersome and time consuming to image (Lubeck and Cai, 2012). A recent development in expansion microscopy as well as correlation methods (Coskun and Cai, 2016), however, offers promise for multiplexing to levels of high transcript density (Chen et al., 2015a, 2016; Treweek et al., 2015; Yang et al., 2014). In addition, by labeling subcellular components (i.e., dendrites and axons) with antibodies, the local transcriptome (Cajigas et al., 2012) in compartments of the cell can be measured.

We observed that, because expression patterns among genes are highly correlated, the distinction between large
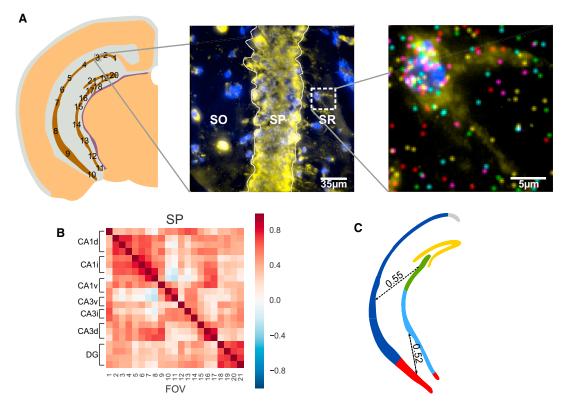
(legend on next page)

Figure 8. Correlations of the Transcription Profiles across the Pyramidal Layer

(A) mRNA counts in the cell bodies in the stratum pyramidale (SP) are grouped within each field of view. A single cell in the stratum radiatum (SR) is shown to illustrate individual mRNA localization. Stratum oriens (SO) is labeled for orientation.

(B) mRNAs in different subregions of the pyramidal layer show both long-distance spatial correlations as well as local correlations between neighboring fields. Both CA1 and dentate gyrus (DG) show high regional correlations. Correlation is calculated based on data from the 125 gene experiment.

(C) Illustration of regional and long-distance correlation patterns observed in (B). Correlated regions are colored, and long-distance correlations are shown as dotted lines with their median correlation coefficient written over the dotted line.

classes of cells can be determined from 10–20 genes, while a finer classification of cell clusters depends on the quantitative measurement of the combinatorial expression patterns of many genes (Figures 3C and 3D). This correlation among genes can be used to "stitch" our seqFISH data with single-cell RNA-seq data, similar to the approach explored with single-cell RNA-seq and ISH in Satija et al. (2015). By correlating seqFISH data to single-cell RNA-seq expression data, cells types identified based on RNA-seq can be "mapped" back into our seqFISH data.

As shown here, seqFISH with hundreds of genes in tissues can become a general and widely used tool to answer a wide range of fundamental questions in biology and medicine. For neuroscience, by combining the insights into the spatial organization of

Figure 7. Mapping of Cell Types to a Third Brain Slice Based on Expression Levels of 249 Genes

Upper right panel: cartoon of hippocampus with imaged regions labeled. Color key corresponds to the classes in Figure S6C.

(A–C) Similar to the slices shown in Figures 5 and 6, CA1d is relatively homogeneous in cell cluster composition. The dorsal-most region of the CA1 (A) shows predominantly class 15 cells. This homogeneity continues throughout the CA1d (B), ending in the last field imaged in the CA1d (C).

(D–G) Images from the CA1i region show that the cell class composition is different from that of the CA1d. Cell classes switch from predominately class 15 cells in the CA1d to predominately class 14 and 16 cells in the CA1i (D). This particular composition of cells is seen in the following field (E) and the field after that (F), terminating in the final field of the CA1i (G).

(H–K) Again, similar to Figures 5 and 6, images from the CA1 ventral regions show a much more complicated cellular composition and a high degree of cellular heterogeneity. The CA1v begins at field (H), and the increased heterogeneity in cell composition is seen in both adjacent field (I) and field (J) up to the point where the CA1 terminates (K).

(L–R) Images from the CA3 region show that the cellular compositions also create 3–4 subregions within the CA3. The cellular heterogeneity of the CA3 subregions mirrors that of the CA1, where the ventral region of the CA3 is very heterogenous (L), while the dorsal region of the CA3 is relatively homogeneous (Q). Field (M) in CA3v shows cellular composition heterogeneity similar to field (N). Field (O) also has a unique cellular composition in the CA3. The final transcriptional subregion within the CA3 begins in field (P) with a predominance of type 12 and 17 cells. This region continues until the end of the CA3 (R).

(S and T) The suprapyramidal (S) and infrapyramidal (T) blades of the DG region show the distinct SGZ versus GCL layering pattern seen in the previous two brains.

transcription provided by seqFISH with connectomics and electrophysiological measurements, we can obtain a comprehensive understanding of the molecular basis of the neuroanatomy of the brain.

EXPERIMENTAL PROCEDURES

Probe Design
Genes were selected from the Allen Brain Atlas database. We identified genes that are heterogeneously expressed in coronal sections containing the hippocampus at Bregma coordinates −2.68 mm anterior. We selected 100 genes that had high variances across these distinct regions and that also had low to medium expression levels. Probe sequences were designed using software developed in house. Full details are described in the Supplemental Experimental Procedures.

Probe Generation
All oligoarray pools were purchased as 92k synthesis from Customarray Inc. Probes were amplified from array-synthesized oligo pool as previously described (Beliveau et al., 2012; Chen et al., 2015b). Full details are described in the Supplemental Experimental Procedures.

Brain Extraction and Sample Mounting
C57BL/6 with Ai6 Cre-reporter (uncrossed) (Jackson Laboratories, SN: 007906) female mice aged 50–80 days were anesthetized with isoflurane according to institute protocols (protocol #1701-14) (Madisen et al., 2012). Mice were perfused with 4% PFA, and the brain was dissected out and placed in a 4% PFA buffer for 2 hr at room temperature. The brain was then immersed in 4°C 30% RNase-free Sucrose/1× PBS until the brain sank. Once sunk, the brain was embedded in OCT and sectioned. Full details are described in the Supplemental Experimental Procedures.

Sample Permeabilization, Hybridization, and Imaging
Sections were permeabilized in 4°C 70% EtOH for 12–18 hr. Brains were further permeabilized by the addition of RNase-free 8% SDS. A hybridization chamber was adhered around the brain section. RNA integrity test probes were hybridized overnight at 37°C in hybridization buffer (Table S3). Samples were washed in 30% wash buffer (WB) for 30 min. Probes were amplified. Following amplification, samples were washed in the same 30% WB for at least 10 min to remove excess hairpins. Samples were stained with DAPI and submerged in pyranose oxidase antibleaching buffer (Lubeck et al., 2014). If the RNA was deemed to be intact, DAPI data were collected in this hybridization. Samples were digested with DNase I for 4 hr at room temperature on the scope. Following DNase I, the sample was washed several times with 30% WB and the probes were hybridized overnight (Tables S4). Samples were again washed and amplified. Repeating this cycle with the appropriate probes for each hybridization barcoded the transcripts over multiple rounds of hybridization. Fluorescent Nissl stain was collected at the end of the experiment along with images of multispectral beads to aid chromatic aberration corrections. Full details are described in the Supplemental Experimental Procedures.

Image Processing
The images were first corrected to remove the uneven illumination profiles in each channel and the effects of chromatic aberration. The background intensity in the images was then subtracted. Full details are described in the Supplemental Experimental Procedures.

Image Registration
The processed images were then registered by first taking a maximum intensity projection along the z direction in each channel. All of the maximum projections of the channels of a single hybridization were collapsed, resulting in 4 composite images each containing all the points in a particular round of hybridization. Each of these composite images of hybridization 1–3 was then cross-correlated individually with the composite image of hybridization 4, and the position of the maxima of the cross-correlation was used as the translation factor to align hybridizations 1–3 to hybridization 4.

Cell Segmentation
For cells in the cortex, the cells were segmented manually using the DAPI images taken in the first round of hybridization and the fluorescent Nissl stain taken at the end of the experiment. Furthermore, the density of the point cloud surrounding a cell was taken into account when forming cell boundaries, especially in cells that did not stain with the Nissl stain. For the hippocampus, the cells were segmented by first manually selecting the centroid in 3D of each DAPI signal of every cell. Transcripts were first assigned based on nearest centroids. These point clouds were then used to refine the centroid estimate and create a 3D voronoi tessellation with a 10% boundary-shrinking factor to eliminate ambiguous mRNA assignments from neighboring cells. Regional segmentation was performed manually using the ImageJ ROI tool.

Barcode Calling
The potential mRNA signals were then found by LOG filtering the registered images and finding points of local maxima above a specified threshold value. Once all potential points in all channels of all hybridizations were obtained, dots were matched to potential barcode partners in all other channels of all other hybridizations using a 1-pixel search radius to find symmetric nearest neighbors. This procedure was repeated using each hybridization as a seed for barcode finding, and only barcodes that were called similarly in at least 3 out of 4 rounds were used in the analysis. The number of each barcode was then counted in each of the assigned cell volumes. Full details are described in the Supplemental Experimental Procedures.

Clustering
To cluster the dataset with two brains probed with 125 genes, we first Z score normalized each of the slices based on gene expression (Tables S5 and S6). Once the single-cell gene expression data were converted into Z scores, we computed a matrix of cell-to-cell correlations using Pearson correlation coefficients for all of the cells in the two brains. Then hierarchical clustering with Ward linkage was performed on the cell-to-cell correlation data using cells taken from the center of the field of view. To analyze the robustness of individual clusters, a random forest model was trained using varying subsets of the data and used to predict the cluster assignment of the remaining cells (Breiman, 2001). For Figures 4, 5, and 6, the entire field of cells was classified using the clustered cells as the training set. A bootstrap analysis by dropping different sets of cells was performed in increments (Figure S4). To determine the effect of dropping out genes on the accuracy of the clustering analysis, we used a random forest decision tree to learn the cluster definition based on the 125 gene data. Then we asked the decision tree to re-compute the cluster assignment on cell-to-cell correlation matrices with fewer and fewer genes (Figures 3C and 3D, green line). Bootstrap resampling was also performed with this analysis (Figures 3C and 3D, blue lines). The PCA and stone analyses were performed using the same cell-to-cell Z scored Pearson correlation matrix. The cell-to-cell correlation in Figure S3I was calculated with an increasing number of principal components dropped (have their eigenvalues set to zero). The cluster assignment accuracy was again computed through the random forest decision tree. The 249 gene experiment was clustered independently with Z score normalized data.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, eight figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2016.10.001.

REFERENCES

Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., Li, J.B., Senaratne, T.N., Williams, B.R., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proc. Natl. Acad. Sci. USA 109, 21301–21306.

Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. Science 313, 1642–1645.

Breiman, L. (2001). Random Forests. Mach. Learn. 45, 5–32.

Cajigas, I.J., Tushev, G., Will, T.J., tom Dieck, S., Fuerst, N., and Schuman, E.M. (2012). The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron 74, 453–466.

Cembrowski, M.S., Bachman, J.L., Wang, L., Sugino, K., Shields, B.C., and Spruston, N. (2016). Spatial Gene-Expression Gradients Underlie Prominent Heterogeneity of CA1 Pyramidal Neurons. Neuron 89, 351–368.

Cenquizca, L.A., and Swanson, L.W. (2007). Spatial organization of direct hippocampal field CA1 axonal projections to the rest of the cerebral cortex. Brain Res. Brain Res. Rev. 56, 1–26.

Chen, F., Tillberg, P.W., and Boyden, E.S. (2015a). Optical imaging. Expansion microscopy. Science 347, 543–548.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015b). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, aaa6090.

Chen, F., Wassie, A.T., Cote, A.J., Sinha, A., Alon, S., Asano, S., Daugharthy, E.R., Chang, J.-B., Marblestone, A., Church, G.M., et al. (2016). Nanoscale imaging of RNA with expansion microscopy. Nat. Methods 13, 679–684, http://dx.doi.org/10.1038/nmeth.3899, advance online publication.

Choi, H.M.T., Beck, V.A., and Pierce, N.A. (2014). Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. ACS Nano 8, 4284–4294.

Coskun, A.F., and Cai, L. (2016). Dense transcript profiling in single cells by image correlation decoding. Nat. Methods 13, 657–660.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. USA 112, 7285–7290.

Dong, H.-W., Swanson, L.W., Chen, L., Fanselow, M.S., and Toga, A.W. (2009). Genomic-anatomic evidence for distinct functional domains in hippocampal field CA1. Proc. Natl. Acad. Sci. USA 106, 11794–11799.

Fan, Y., Braut, S.A., Lin, Q., Singer, R.H., and Skoultchi, A.I. (2001). Determination of transgenic loci by expression FISH. Genomics 71, 66–69.

Fanselow, M.S., and Dong, H.-W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? Neuron 65, 7–19.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science 280, 585–590.

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F., and Regev, A. (2016). Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science 353, 925–928.

Jung, M.W., Wiener, S.I., and McNaughton, B.L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. J. Neurosci. 14, 7347–7356.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods 10, 857–860.

Kishi, T., Tsumori, T., Yokota, S., and Yasui, Y. (2006). Topographical projection from the hippocampal formation to the amygdala: a combined anterograde and retrograde tracing study in the rat. J. Comp. Neurol. 496, 349–368.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187–1201.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science 343, 1360–1363.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168–176.

Lubeck, E., and Cai, L. (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat. Methods 9, 743–748.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods 11, 360–361.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214.

Madisen, L., Mao, T., Koch, H., Zhuo, J.M., Berenyi, A., Fujisawa, S., Hsu, Y.-W.A., Garcia, A.J., 3rd, Gu, X., Zanella, S., et al. (2012). A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. Nat. Neurosci. 15, 793–802.

Miller, J.A., Nathanson, J., Franjic, D., Shim, S., Dalley, R.A., Shapouri, S., Smith, K.A., Sunkin, S.M., Bernard, A., Bennett, J.L., et al. (2013). Conserved molecular signatures of neurogenesis in the hippocampal subgranular zone of rodents and primates. Development 140, 4633–4644.

Muller, R.U., Stead, M., and Pach, J. (1996). The hippocampus as a cognitive graph. J. Gen. Physiol. 107, 663–694.

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 34, 171–175.

Petrovich, G.D., Canteras, N.S., and Swanson, L.W. (2001). Combinatorial amygdalar inputs to hippocampal domains and hypothalamic behavior systems. Brain Res. Brain Res. Rev. 38, 247–289.

Pitkänen, A., Pikkarainen, M., Nurminen, N., and Ylinen, A. (2000). Reciprocal connections between the amygdala and the hippocampal formation, perirhinal cortex, and postrhinal cortex in rat. A review. Ann. N Y Acad. Sci. 911, 369–391.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 4, e309.

Risold, P.Y., and Swanson, L.W. (1996). Structural evidence for functional domains in the rat hippocampus. Science 272, 1484–1486.

Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat. Methods 3, 793–795.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502, http://dx.doi.org/10.1038/nbt.3192.

Saunders, R.C., Rosene, D.L., and Van Hoesen, G.W. (1988). Comparison of the efferents of the amygdala and the hippocampal formation in the rhesus monkey: II. Reciprocal and non-reciprocal connections. J. Comp. Neurol. 271, 185–207.

Shah, S., Lubeck, E., Schwarzkopf, M., He, T.F., Greenbaum, A., Sohn, C.H., Lignell, A., Choi, H.M., Gradinaru, V., Pierce, N.A., and Cai, L. (2016). Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. Development *143*, 2862–2867.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82, http://dx.doi.org/10.1126/science.aaf2403.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *19*, 335–346, advance online publication.

Thompson, C.L., Pathak, S.D., Jeromin, A., Ng, L.L., MacPherson, C.R., Mortrud, M.T., Cusick, A., Riley, Z.L., Sunkin, S.M., Bernard, A., et al. (2008). Genomic anatomy of the hippocampus. Neuron *60*, 1010–1021.

Treweek, J.B., Chan, K.Y., Flytzanis, N.C., Yang, B., Deverman, B.E., Greenbaum, A., Lignell, A., Xiao, C., Cai, L., Ladinsky, M.S., et al. (2015). Whole-body tissue stabilization and selective extractions via tissue-hydrogel hybrids for high-resolution intact circuit mapping and phenotyping. Nat. Protoc. *10*, 1860–1896.

Witter, M.P., and Amaral, D.G. (1991). Entorhinal cortex of the monkey: V. Projections to the dentate gyrus, hippocampus, and subicular complex. J. Comp. Neurol. *307*, 437–459.

Yang, B., Treweek, J.B., Kulkarni, R.P., Deverman, B.E., Chen, C.-K., Lubeck, E., Shah, S., Cai, L., and Gradinaru, V. (2014). Single-cell phenotyping within transparent intact tissue through whole-body clearing. Cell *158*, 945–958.

Yang, S.M., Alvarez, D.D., and Schinder, A.F. (2015). Reliable Genetic Labeling of Adult-Born Dentate Granule Cells Using Ascl1 CreERT2 and Glast CreERT2 Murine Lines. J. Neurosci. *35*, 15379–15390.

Yi, F., Catudio-Garrett, E., Gábriel, R., Wilhelm, M., Erdelyi, F., Szabo, G., Deisseroth, K., and Lawrence, J. (2015). Hippocampal "cholinergic interneurons" visualized with the choline acetyltransferase promoter: anatomical distribution, intrinsic membrane properties, neurochemical characteristics, and capacity for cholinergic modulation. Front. Synaptic Neurosci. *7*, 4.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.