Single-cell transcriptomic analysis of CA1 inhibitory neurons

Kenneth D. Harris1,2, Carolina Bengtsson Gonzales3, Hannah Hochgerner2, Nathan G. Skene2, Lorenza Magno4, Linda Katona5, Peter Somogyi5, Nicolette Kessaris4, Sten Linnarsson2, Jens Hjerling-Leffler3

1 – UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
2 – UCL Department of Neuroscience, Physiology and Pharmacology, 21 University Street, London WC1E 6DE, UK
3 - Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden
4 – Wolfson Institute for Biomedical Research, University College London, Gower Street, London WC1E 6BT, UK
5 – Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3TH, UK

Single-cell transcriptomics provides a powerful tool to understand cell classes in neural circuits. In cortex, however, the identification of distinct biological cell types based on transcriptomic data is complicated by the existence of many incompletely-understood and closely-related neural classes, which can also show continuous variation in gene expression. The inhibitory cells of hippocampal area CA1 have been extensively characterized, leading to 23 currently defined cell types that provide a “Rosetta Stone” for transcriptomic analysis. Here, we studied the transcriptomes of 3638 CA1 inhibitory cells. Novel clustering methods identified all 23 previously described CA1 inhibitory types, while also suggesting 6 new inhibitory classes. Latent-factor analysis revealed a common continuum of expression of many genes within and between classes, which we hypothesized correlates with a continuum from faster-spiking cells that proximally target pyramidal cells, to slower active cells targeting pyramidal distal dendrites or interneurons. In vitro patch-sequ of Pvalb interneurons supported hypothesis.

Introduction

Understanding of the cell types comprising neural circuits is critical for an understanding their function. Single-cell RNA sequencing (scRNA-seq) provides a powerful opportunity to categorizing these types, as major cell classes can have substantially different patterns of gene expression (Habib et al., 2016; Macosko et al., 2015; Tasic et al., 2016; Usoskin et al., 2015; Zeisel et al., 2015). Nevertheless, relating gene expression patterns to underlying cell types is not always straightforward. This is particularly true in circuits such as the cortex, where many closely-related excitatory and inhibitory cell types exist. Indeed, while transcriptomics can divide cells into clusters with similar gene expression profiles, identifying them with known cell types will require connecting these molecular classes to morphological and functional categories.

The hippocampus provides a promising opportunity to relate transcriptomic analyses to cortical cell types. The hippocampus has a simpler architecture than isocortex, and intensive work over the last 30 years has built up a relatively complete picture of hippocampal inhibitory cells types, relating their morphological characteristics, connectivity, physiology and in vitro firing patterns to several tens of molecular markers (Freund and Buzsaki, 1996; Klausberger and Somogyi, 2008). This analysis has been most fully completed in area CA1, where 23 distinct inhibitory types have been defined so far (Somogyi, 2010).

Identification of transcriptomic clusters with cell types is complicated by the fact that cells can show continuous variation in gene expression. Previous work has shown apparently continuous variation within single cell types, as well as fuzzy boundaries between clusters corresponding to putative single classes (Marques et al., 2016; Tasic et al., 2016; Zeisel et al., 2015). Continuous variation in gene expression could reflect continuous variation of cellular properties, such as laminar location, physiology, or connectivity profiles. Furthermore, gene expression levels in interneurons are dynamically regulated, for example by activity or by learning (Cohen et al., 2016; Dehorer et al., 2015; Donato et al., 2013; Mardinly et al., 2016; Spiegel et al., 2014). Understanding how such continua might relate to both fixed features such as axonal and dendritic connectivity and to activity will be easier in systems such as CA1, whose cell types have been extensively characterized by previous work.

Here we describe a transcriptomic analysis of 3638 inhibitory neurons from mouse hippocampal area CA1, revealing 49 clusters out of which we could identify 43 as previously described cell types. All 23 previously described CA1 GABAergic classes could be identified in our database, with the larger number of clusters reflecting six hypothesized novel cell types, several previously unappreciated subtypes of existing types, and tiling of continuous gradation within single types by multiple clusters. A common continuous pattern of gradation between and within classes appeared to reflect a continuum from fast-firing cells targeting principal cell
CA1 inhibitory neurons

somata and proximal dendrites, to slower-firing cells targeting distal dendrites or interneurons. Patch-seq recordings of hippocampal Pvalb interneurons confirmed that this genetic latent factor correlates with the strength of a neuron’s fast-spiking phenotype.

Results

Data collection and identification of inhibitory cells

We collected cells from six Slc32a1-Cre;R26R-tdTomato mice, three of age p60 and three of age p27. Cells were procured using enzymatic digestion and manual dissociation of CA1 tissue after dissection as previously described (Zeisel 2015). The great majority of cells (4572/6971 cells total; 3283/3638 high-quality interneurons) came from the older animals. Because we observed no major difference in interneuron classes between ages, data was pooled between them (Figure S1). FACS sorting yielded an enriched, but not completely pure population of GABAergic neurons. A first-round clustering (using the method described below) was therefore run on the 5940 cells passing quality control, identifying 3638 inhibitory and 357 excitatory neurons (as judged by the expression of genes Gad1/Slc32a1 and Vglut1 respectively) as well as 1779 non-neuronal, and 166 unclassified cells.

Cluster analysis

We analyzed the data using a suite of three algorithms, derived from a probabilistic model of RNA distributions. All three methods were based on the observation that RNA counts within a homogeneous population can be approximated using a negative binomial distribution (See Experimental Procedures; Lu et al., 2005; Robinson and Smyth, 2008). The negative binomial distribution accurately models the high variance of transcriptomic read counts (Figure S2A,B). As a consequence, algorithms based on this distribution weight the presence or absence of a gene more than its numerical expression.

Figure 1. Two-dimensional visualization of expression patterns using nbtsNE algorithm, which places cells closely together according to a negative binomial criterion. Each symbol represents a cell, with different color/glyph combinations representing different cell classes (legend, right). Grey boxes and numbers refer to the “continents” referred to in the text and subsequent figures.
level – for example, this distribution treats read counts of 0 and 10 as more dissimilar than read counts of 500 and 1000 (Figure S2C).

The algorithm we used for clustering was termed ProMMT (Probabilistic Mixture Modeling for Transcriptomics). This algorithm fits gene expression in each cluster \( k \) by a multivariate negative binomial distribution with cluster-specific mean \( \mu_k \). The mean expression levels of only a small subset of genes are allowed to vary between clusters (150 for the current analysis); these genes are selected automatically by the algorithm by maximum likelihood methods. The use of such “sparse” methods is essential for probabilistic classification of high dimensional data (Bouveyron and Brunet-Saumard, 2014), and the genes selected represent those most informative for cluster differentiation. The number of clusters was chosen automatically using the Bayesian Information Criterion (BIC) (Schwarz, 1978). The ProMMT algorithm also provides a natural measure of the distinctness of each cluster, which we term the isolation metric (see Experimental Procedures).

The ProMMT algorithm divided CA1 interneurons into 49 clusters (Figure S3). We named the clusters using a multilevel scheme, after genes that are strongly expressed at different hierarchical levels; for example, the cluster Calb2, Vip, Nos1 belongs to a first level group characterized by strong expression of Calb2 (indicating interneuron selective interneurons); a second level group Calb2, Vip; and a third level group distinguished from other Calb2, Vip cells by stronger expression of Nos1. This naming scheme was based on the results of hierarchical cluster analysis of cluster means, using a distance metric.

**Figure 2.** Expression levels of 25 selected genes, that together allow identification of major cell classes. Each subplot shows an nbISNE map of all cells, with marker size indicating log-expression level of the gene named above the plot.
Data Visualization

To visualize cell classes in two dimensions, we modified the t-stochastic neighbor embedding algorithm (Maaten and Hinton, 2008) for data with negative binomial variability, terming this approach nbtSNE. In conventional tSNE, the similarity between data points is defined by their Euclidean distance, which corresponds to conditional probabilities under a Gaussian distribution. We obtained greater separation of clusters and a closer correspondence to known cell types, by replacing the Gaussian distribution with the same negative binomial distribution used in our clustering algorithm (see Experimental Procedures; Figure S5).

The nbtSNE maps revealed that cells were arranged in 9 major “continents” (Figure 1). The way expression of a single gene differed between classes could be conveniently visualized on these maps by adjusting the symbol size for each cell according to that gene’s expression level. Thanks to the extensive literature on CA1 interneurons, 25 genes together sufficed to identify the main continents with known cell classes (Figure 2), and it was also possible to identify nearly all the finer subclasses using additional genes specific to each class.

Identification of cell types

Explaining how we identified these transcriptomic clusters with known cell types requires an extensive discussion of the previous literature, which is presented in full in the Supplementary Material. Here, we briefly summarize the major subtypes identified (summarized in Figure 3).

Continent 1 was identified with the Sst positive hippocampopetal and O-LM cells of stratum oriens (so). These cells all expressed Sst and Grn1, and were further divided into two Npy+/Ngf+ clusters identified as...
hippocamposeptal neurons (Acády et al., 2000), and three Pnoc+/Reln+/Npy- clusters identified with O-LM cells (Katona et al., 2014). A highly distinct cluster located in an “island” off this continent contained cells strongly positive for Sst and Nos1 (Jinno and Kosaka, 2004), whose expression pattern matches those of both backprojection cells (Sik et al., 1994) and PENK-positive projection cells (Fuentesalba et al., 2008a), suggesting that these may reflect a single cell type. In addition, Continent 1 contains a previously undescribed subclass positive for Sst, Npy, and Reln.

Continent 2 was identified as basket and bistriated cells. These were all positive for Tac1 (the precursor to the neuropeptide Substance P), as well as Sath1 and Erbb4, but were negative for Gm1. They were divided into two Pvalb+/Sst- clusters identified with basket cells, two Pvalb+/Sst+/Npy+ clusters identified with bistriated cells (Klausberger et al., 2004), and three Pvalb- clusters identified with Oriens-Bistratified (o-Bi) cells (Losonczy et al., 2002).

Continent 3 was identified as axo-axonic cells, due to their expression of Pvalb but not Sath1 (Viney et al., 2013) This continent’s three clusters were Tac1-negative but positive for other markers including Snca, Phlh and C1ql1, which have also been associated with axo-axonic cells in isocortex (Tasic et al., 2016). We note that this dichotomy of Pvalb interneurons to Tac1 positive and negative subclasses is likely homologous to previous observations in isocortex (Vruwink et al., 2001).

Continent 4 was identified as Ivy cells and MGE-derived neurogliaform cells. These cells expressed Cacna2d1, which we propose as a unique identifier of hippocampal neurogliaform/ivy cells, as well as Lhx6 and Nos1 (Tricoire et al., 2010). They were divided into a Reln+ cluster identified with MGE-derived neurogliaform cells, and a Reln-/Vwa5a+ cluster identified with Ivy cells (Fuentesalba et al., 2008b). This continent is homologous to the isocortical Igtp class defined by Tasic et al (2016), which we hypothesize may represent isocortical neurogliaform cells of MGE origin; this hypothesis could be confirmed using fate-mapping.

Continent 5 was identified as CGE-derived neurogliaform cells. Its three clusters contained Cacna2d1 and many other genes in common with those of continent 4, but lacked Lhx6 and Nos1 (Tricoire et al., 2010). Similar to isocortical putative neurogliaform cells, this continent expressed Ndnp and contained a distinct subtype positive for Cxcl14 (Tasic et al., 2016). As with continent 4, continent 5 mainly expressed Reln but also contained a small Reln-negative cluster, which we suggest forms a rare and novel class of CGE-derived Ivy cell.

Continent 6 was identified with Sst-negative long-range projection interneurons. It divided into two distinct clusters, both of which were strongly positive for Ntrng1. The first strongly expressed Chrm2 but lacked Sst and Pvalb, identifying them as trilaminar cells (Jinno et al., 2007). The second subgroup lacked most classical molecular markers; this fact, together with their inferred laminar location at the sr/slm border, identified them as putative radiatum-retrohippocampal neurons that project to retrosplenial cortex (Jinno et al., 2007; Miyashita and Rockland, 2007).

Continents 7 and 8 were identified as what are traditionally called Cck interneurons. This term is somewhat unfortunate: while these cells indeed strongly express Cck, many other inhibitory classes express Cck at lower levels, including even Pvalb+ basket cells (Tricoire et al., 2011). Continents 7 and 8 cells comprised thirteen highly diverse clusters, but shared strong expression of Cnr1, Sncg, Trp53i11 and several other novel genes. Continent 8 is distinguished by expression of Cxcl14, which localizes these cells to the border of stratum lacunosum and stratum lacunosum-moleculare (sr/slm). This continent comprised a continuum ranging from soma-targeting basket cells identified by their Snc17a8 (vGlut3) expression, to dentricle targeting cells identified by expression of Calb1 or Reln (Klausberger et al., 2005; Somogyi et al., 2004). Continent 7, lacking Cxcl14, was identified as Cck cells of other layers, and contained multiple subtypes characterized by the familiar markers Calb1, Vip, Snc17a8 (Somogyi et al., 2004), as well novel markers such as Sem3a and Calca. Associated with continent 8 were several apparently novel subtypes: a rare and distinct group positive for both Snc17a8 and Calb1 and marked nearly exclusively by Lypld1; a Ntrng1+/Ndnp+ subgroup related to cells of continent 6; and a group strongly expressing both Vip and Cxcl14, which therefore likely corresponds to a novel Vip+/Cck+ interneuron at the sr/slm border.

Continent 9 was identified as interneuron-selective interneurons. Its eight clusters fell into three groups: Calb2+/Vip- neurons identified as IS-1 cells; Calb2+/Vip+ neurons identified as IS-2 cells; and Calb2+/Vip+ neurons identified as IS-3 cells (Acády et al., 1996; Freund and Buzsaki, 1996; Gulyás et al., 1996; Tyan et al., 2014). All expressed Penk (Blasco-Ibanez et al., 1998). These cells contained at least two novel subgroups: an IS-3 subtype positive for Nos1 and My11, homologous to the Vip Mybpc2 class defined in isocortex (Tasic et al., 2016); and a rare subclass of IS-1 cell positive for Igf1rp6.
Continuous variation between and within cell classes

Although the major continents of the expression map were clearly separated, clusters within these continents often appeared to blend into each other continuously. This indicates continuous gradation in gene expression patterns; while our probabilistic mixture model will group cells from a common negative binomial distribution into a single cluster, it will tile cells of continuously graded mean expression into multiple clusters.

These continua reflect both positive and negative correlations between genes. For example, in Pvalb-positive putative basket and bistratified cells, multiple genes associated with fast-spiking phenotype (such as Kcnc1, Kcna1, and Pvalb) were positively correlated with expression of the Na+/K+ pump Atp1b1, but negatively correlated with expression the ATP1B1 modulator Fxyd6 (mutations in which have been associated with schizophrenia; Choudhury et al., 2007). This correlated variability was evident both between subtypes of Pvalb neurons, and between different individual cells within Pvalb classes (Figure 4A). We conclude that Pvalb-positive interneurons exhibit continuous correlated variation in the expression of multiple genes; that the mean expression in different Pvalb subclasses lies at different locations along this continuum; but that variability also exists within the cells of a single subclass.

Patterns of correlated variability were similar across multiple interneuronal classes, in a manner consistent with their known spiking phenotypes (Figure 4A,B). For example, most CCK interneurons show a regular-spiking phenotype, but a minority of fast-spiking CCK interneurons have also been reported (Cope et al., 2002; Pawelzik et al., 2002). Consistent with this, we found that expression of fast-spiking associated genes such as Kcnc1 was on average lower in clusters corresponding to Cck interneurons, but that some cells still expressed these genes. Expression of Kcnc1 showed similar correlations with Atp1b1 and Fxyd6 within Cck interneurons as within Pvalb-positive neurons (Figure 4B). We suggest that the preserved correlations between these example genes reflect common phenotypic modulation between these cells, specifically reflecting altered ion exchange requirements in faster-spiking cells.

In addition to these commonalities, however, there were differences in the correlated variability across cell types. These mostly reflected differences in which genes are expressed by which classes. For example, because Pvalb is not expressed in Cck interneurons, it cannot exhibit a correlation; conversely, other genes expressed in Cck but not Pvalb interneurons (such as the cannabinoid receptor Cnr1) showed similar patterns of correlation with other genes (Figure 4B). Nevertheless, the patterns of correlation in continent 9 (identified with I-S cells) were largely unrelated to those of previous cell groups, even when the same genes were expressed. For example, Fxyd6 and Atp1b1 were positively correlated in this group (not shown).

To characterize this common mode of continuous variability across cell classes beyond these examples, we developed a probabilistic latent factor model (Experimental Procedures). In this model, gene expression vectors follow a negative binomial distribution whose mean depends on a single latent factor (i.e. a hidden variable), with a gene-specific weight. We applied this latent factor model to the entire dataset, and studied the weighting of genes with known function to infer the biological significance of the continuum (Figure 4C). Genes associated with fast-spiking phenotype (e.g. Kcnc1, Kcna1, Pvalb) had positive weights, as did mitochondrial genes; genes associated with ion exchange and metabolism (e.g. Atp1b1; Slc24a2); genes associated with GABA synthesis, transport, and vesicular release (e.g. Gad1, Slc6a1, Syp, Sc2a, Cplx2, Vamp1); and genes associated with fast ionotropic glutamate and GABA receptors (e.g. Gria1, Gabra1), as well as GABA receptors (e.g. Gabbr1, Gabbr2, Kcnj3, Kctd12). The genes correlating negatively with the latent factor were less familiar, but included Atp1b2, a second isoform of the Na+/K+ pump; Fxyd6, which modulates its activity; Nrsn1, whose translation is suppressed after learning (Cho et al., 2015), as well as most neuropeptides (e.g. Sst, Vip, Cartpt, Tac2, Penk, Crh; exceptional neuropeptides such as Cck showed positive correlation). Genes associated with neurofilaments and intermediate filaments (e.g. Nefh, Neff, Krt73) tended to show positive weights, while genes associated with actin processing (e.g. Gap43, Stnms1, Tmsb10) tended to show negative weights. Many other genes of as yet unknown function correlated positively and negatively with this latent factor (for example 6330403K07Rik, which showed strong negative correlations).

We therefore hypothesize that cells with large values of the latent factor have a fast-spiking firing pattern; possess more synaptic vesicles and release larger amounts of GABA; receive stronger excitatory and inhibitory inputs; and exhibit faster metabolism. These are all characteristics of Pvalb-expressing fast-spiking interneurons (Hu et al., 2014), but a similar continuum was observed also within Pvalb-negative neurons.
CA1 inhibitory neurons

The latent factor also differed between cells inhibiting the soma and dendrites of pyramidal cells. For example, consider the Cck.Ccx14 cells of continent 8 (Figure 4D), identified with Cck-positive neurons at the sr/slm border.

Immunohistochemistry has demonstrated that in these cells, gene expression correlates with axon target: SLC17A8-positive basket cells project to the pyramidal layer (Somogyi et al., 2004), while CALB1 is expressed in

Figure 4. Continuous variation of expression between and within cell classes. A, In Pvalb-positive neurons, genes associated with fast-spiking phenotype (e.g. Kcnc1 and Pvalb) are positively correlated with the Na+/K+ pump Atp1b1 but negatively correlated with its modulator Fxyd6. B, In Cck/Vip-positive neurons, expression of Kcnc1 is lower than in Pvalb neurons, but correlations with Atp1b1 and Fxyd6 persist. Pvalb is not expressed in these neurons, but a similar correlation is observed for Cnr1. C, Latent factor analysis reveals a common mode of variation across cell types. Y-axis shows weighting of each gene onto the latent factor; x-axis shows relative expression of gene in MGE- and CGE-derived cell classes. Genes associated with fast-spiking phenotype, GABA synthesis, metabolism, excitatory and inhibitory inputs, and vesicle release are positively weighted on the latent factor. Black text highlights selected genes of particular interest. D, In Cck interneurons, the latent factor correlates with what part of pyramidal cells they target. Top: zoom of nbISNE map for region of continent 8 identified with Cck interneurons at the sr/slm border, with symbol size representing value of hidden factor. Middle: symbol size represents expression of Slc17a8, a marker of soma-targeting Cck cells. Bottom: symbol size represents expression of Calb1, a marker of dendrite-targeting Cck cells.
neurons targeting pyramidal cell dendrites (Cope et al., 2002; Gulyás and Freund, 1996; Klausberger and Somogyi, 2008). Furthermore, in vitro recordings indicate that Cck interneurons targeting pyramidal somata have a faster-spiking phenotype: lower input resistance, narrower spikes, and less adaptation than those targeting pyramidal dendrites (Cope et al., 2002). The latent factor was higher in soma-targeting neurons: it was high to the west of continent 8, where Slc17a8 is expressed, but lower in the east where Calb1 is expressed (Figure 4D). Consistent with this result, the cannabinoid receptor Cnr1, which is more strongly expressed in soma-targeting neurons (Dudok et al., 2015; Lee et al., 2010) was more strongly expressed in the west. This therefore indicates a gradation from faster-spiking soma-targeting cells in the west of continent 8, to more regular-spiking dendrite-targeting cells in the east. Similar behavior was seen for many other cell classes. For example, Sst.Pnoc cells, identified with distal-targeting O-LM cells had lower latent factor values than Pvalb.Tac1 cells identified with proximal-targeting basket and bistratified cells. The smallest latent factor values belonged to the Calb2 cluster, identified with interneuron-specific interneurons.

In summary, we suggest that most interneuron classes exhibit a common continuum of gene expression patterns, that ranges from expression patterns typical of soma-targeting neurons that fire faster, synthesize and release more GABA, to slower-firing dendrite-targeting or interneuron-specific neurons whose action might rely more heavily on neuropeptides. Similar variability was seen within cells of a single type, with the exception of interneuron-selective cells, whose patterns of continuous variation differ to those of other classes.

Patch-seq analysis confirms that genetic continuum correlates with fast-spiking phenotype

To test the hypothesis that the latent factor correlates with a neuron’s electrophysiological firing pattern, we turned to Patch-seq analysis (Cadwell et al., 2016; Fuzik et al., 2016). We recorded from 25 hippocampal interneurons of area CA1 of Pvalb-Cre;R26R-TdTomato mice using in vitro whole-cell patch-clamp, and measured a set of standard parameters using square current steps. Principal component analysis of these parameters (Figure 5A) revealed a dominant mode of variability (PC1) corresponding to variations in the strength of fast-spiking phenotype, characterized by narrow spikes, fast membrane timeconstant, low adaptation and rheobase, and rapid firing, sometimes with delayed onset (Hu et al., 2014). We sequenced mRNA from the patch pipette using previously described methods (Fuzik et al., 2016) setting

![Figure 5](https://example.com/figure5.png)
CA1 inhibitory neurons

![Image](https://example.com/image.png)

**Figure 6.** Immunohistochemical characterization of intensely NOS1-positive neurons. A. A large multipolar neuron in stratum pyramidale is strongly SST and NPY positive in the somatic Golgi apparatus and weakly positive for CHRM2 in the somato-dendritic plasma membrane (maximum intensity projection, z stack, height 11 µm; inset, maximum intensity projection of 3 optical slices, z stack height 2 µm). A smaller more weakly NOS1-positive cell (double arrow) in lower left is immunonegative for the other molecules; a second NPY positive cell (arrow) adjoining the NOS1+ neuron is immunonegative for the other three molecules. B. A NOS1-positive cell and another NOS1-immunonegative cell (asterisk) at the border of stratum radiatum and lacunosum-moleculare are both positive for GRM1 in the plasma membrane and PENK in the Golgi apparatus and in granules, but only the NOS1+ cell is immunopositive for CHRM2 (maximum intensity projection, z stack, height 10 µm). C. An intensly NOS1-positive cell in stratum radiatum is also positive for PCP4 in the cytoplasm and nucleus, and for PENK in the Golgi apparatus and in granules (maximum intensity projection, z stack, height 15 µm).

To test the hypothesis, we computed the latent factor value for each patched neuron, using the gene weights obtained from the full scRNA-seq database (Figure 4C). Confirming the hypothesis that neurons of high latent factor values have faster-spiking phenotypes, we found that latent factor values were strongly correlated with the first principal component obtained from electrophysiological measurements (Spearman rank correlation, r=0.79; p=0.006). Analysis of the relationship between this genetic latent factor and individual firing properties (Figure S6) confirmed that cells with high latent factor values were more likely to exhibit fast-spiking and delayed-spiking phenotypes.

**Genetic continuum partially matches that predicted from activity-regulated gene expression**

Gene expression depends not only on differences between cell types, but also dynamic modulation by factors such as by activity levels (Cohen et al., 2016; Dehorter et al., 2015; Donato et al., 2013; Mardinly et al., 2016; Spiegel et al., 2014). To test whether the continuum of gene expression revealed by our latent factor analysis was similar to that expected from activity-dependence, we correlated each gene’s latent factor score with that gene’s modulation by *in vivo* light exposure after dark housing, using data from 3 classes of visual cortical interneurons (made available by Mardinly et al., 2016). We observed a relatively strong positive correlation in *Sst* neurons (r=0.26; p<0.01; Figure S7), suggesting that at least activity dependent modulation of *Sst* cells may cause them to move along this same continuum. A weaker but still significant correlation was observed for *Pvalb* neurons (r=0.11; p<0.002), whereas no significant relationship was found for *Vip* neurons (p=0.17). These data therefore suggest that at least some of the continuous variability of gene expression observed in CA1 interneurons may arise from activity-dependent modulation, but that activity-dependent modulation in *Vip* cells does not match the common continuum seen in other cell types. This latter hypothesis is consistent both with the different correlation patterns seen within continent 9 of our CA1 database, and with the exceptional...

**Histological confirmation of transcriptomic predictions**

The transcriptomic classification we derived makes a large number of predictions for the combinatorial expression patterns of familiar and novel molecular markers in distinct CA1 interneuron types. To verify our transcriptomic classification, we set out to test some of these predictions using traditional methods of molecular histology.

Our first tests regarded the very distinct Sst.Nos1 cluster. This cluster’s expression pattern matched three previously reported rare hippocampal inhibitory cell types: large SST-immunopositive cells that are intensely immunoreactive for NOS1 throughout the cytoplasm revealing their full dendrites (Jinno and Kosaka, 2004 Neuroscience 124:797-808); PENK-positive projection cells (Fuentealba et al., 2008a); and strongly NADPH diaphorase-labelled (i.e. NOS1 positive) backprojection cells (Sik et al., 1994). We therefore hypothesized that these cell types, previously regarded as separate, may in fact be identical or at least closely related. To test this hypothesis, we performed a series of triple and quadruple immunoreactions, focusing on the intensely NOS1-positive neurons (n=3 mice, n=70 cells: 39% in so/alueus; 10% in sp; 27% in sr; 24% at sr/slm border). Similar to previously reported PENK-projection, backprojection, and SST/NOS1 cells (Fuentealba et al., 2008a; Jinno and Kosaka, 2004; Sik et al., 1994) – but unlike SST-positive OLM and bistratified cells (Katona et al., 2014) – these neurons all showed aspiny or sparsely spiny dendrites. As expected from the Sst.Nos1 cluster, we found that they were all SST/NPY double positive (n=20/20) and were virtually all weakly positive for CHRM2 (n=36/38) and GRM1 (n=17/17) in the somato-dendritic plasma membrane, strongly positive for PCP4 (n=19/21) in the cytoplasm and nucleus, and for PENK (n=35/42) in the Golgi apparatus and granules in the soma and proximal dendrites (Figure 6). By contrast, the more numerous moderately NOS1 positive cells (which include many interneuron types such as ivy, MGE-neurogliaform and a subset of IS-3 neurons) were mostly immunonegative for CHRM2, PCP4 and PENK, although some were positive for GRM1. Our results are therefore consistent with the hypothesis that all three previously reported classes correspond to the Sst1.Nos1 cluster.

A second prediction of our classification was the expression of Npy in multiple subclasses of Cck cell, most notably the Slc17a8 and Calb1 expressing clusters of continent 8. This was unexpected, as NPY (at least at the protein level) has instead been traditionally associated with SST-expressing neurons and ivy/neurogliaform cells (Fuentealba et al., 2008a, Katona et al., 2014). Nevertheless, no studies to our knowledge have yet examined immunohistochemically whether the neuropeptides NPY and CCK can be colocalised in the same interneurons. We therefore tested this by double immunohistochemistry in sr and slm (Figure S8A,B, n=3 mice). Consistent with our predictions, 119 out of 162 (74±6%) of the cells immunopositive for pro-CCK were also positive for NPY (an additional 73 cells were positive for NPY only, which according to our identifications should represent neurogliaform and radiatum-retrohippocampal cells). A subset (176 cells) of NPY and/or pro-CCK immunopositive neurons were further tested for CALB1 in triple immunoreactions. As expected, nearly all CALB1-positive neurons were pro-CCK-positive (89±2%), and CALB1 immunoreactivity was seen in a subset of the cells containing both pro-CCK and NPY (27±3%). Additional triple immunohistochemistry for NPY, pro-CCK and SLC17A8 (VGLUT3) revealed triple positive cells in sr and particularly at the sr/slm border (Figure S8B). Due to the low level of somatic immunoreactivity for SLC17A8 (which as a vesicular transporter is primarily trafficked to axon terminals), we could not count these cells reliably; however of the cells that were unambiguously immunopositive for SLC17A8, in a majority we detected NPY. Additional analysis combining double in situ hybridization for Slc17a8 and Npy with immunohistochemistry for pro-CCK (Figure S8B, n=3 mice) confirmed that the great majority of Slc17a8-expressing cells were also positive for Npy and pro-CCK (84±3%). As predicted by our identifications, the converse was not true: a substantial population of Npy/pro-CCK double-positive cells (57±7% of the total) did not show detectable Slc17a8, which we identify with dendrite-targeting neurons in the east of continent 8.

Several cell types in our classification expressed Cxcl14, a gene whose expression pattern in the Allen Atlas shows localization largely at the sr/slm border. The Cxcl14-positive population includes all clusters of continent 8, which express Cck and contain subclusters expressing Npy, Calb1, Reln, and Vip; a subtype of CGE-derived neurogliaform cell that expresses Reln and Npy but lacks Nos1 and expresses Kit at most weakly; as well as IS-1, IS-2, and radiatum-retrohippocampal cells. However, our classification predicts that the Cxcl14-positive population should be distinct from all MGE-derived neurons, including MGE-derived neurogliaform cells.

To test these predictions, we performed in situ hybridization for Cxcl14 simultaneously with in situ hybridization or immunohistochemistry to detect Reln,
Figure 7. Analysis of Cxcl14 co-expression patterns confirms predicted properties Cck/Cxcl14 cells. A, Cxcl14-expressing cells are CGE-derived: in situ hybridization for Cxcl14 combined with immunohistochemistry for YFP in the Lhx6-Cre/R26R-YFP mouse yields no double labelling. B, Double in situ hybridization for Cxcl14 and Reln marks a population of neurons located primarily at the sr/slm border. Note Reln expression without Cxcl14 in so and slm, likely reflecting O-LM and neurogliaform cells. C-E, Subsets of the Cxcl14-positive neurons are positive for pro-CCK or CALB1 (in situ hybridization plus immunohistochemistry), or Npy (double in situ hybridization). (f, g) No overlap was seen of Cxcl14 with Nos1 or Kit. In all panels, arrowheads indicate double-expressing neurons. Layer abbreviations: so, stratum oriens; sp, stratum pyramidale; sr, stratum radiatum; b, sr/slm border region; slm, stratum lacunosum-moleculare; sm, stratum moleculare of the dentate gyrus. Scale bars: 200 µm (a), 100 µm (b-g).

Npy, CALB1, CCK, PV-ALB, Sst, Nos1 and Kit (n=3 mice; Figure 7). In addition, we combined fluorescent in situ hybridization for Cxcl14 with immunohistochemistry for YFP in Lhx6-Cre/R26R-YFP mice, which allows identification of developmental origin by marking MGE-derived interneurons (Fogarty et al., 2007). The results of these experiments were consistent with our hypotheses. We found that within CA1, Cxcl14-expressing cells were primarily located at the R-LM border (71±3%), although a subpopulation of cells were also found other layers. We also found overlap of Cxcl14 with YFP in the Lhx6-Cre/R26R-YFP mouse, confirming the CGE origin of Cxcl14 expressing neurons (Figure 7A); consistent with this finding, no overlap was seen with Cxcl14 and Sst or Pvalb (data not shown). The majority of Cxcl14-positive cells expressed Reln (72±4%), accounting for 42±9% of Reln-expressing neurons (Substantial populations of Reln+/Cxcl14- cells were located in strata oriens and lacunosum-moleculare and likely represent O-LM and MGE-neurogliaform cells, respectively (Figure 7B). Indeed, although less than half of Reln cells were located at the R-LM border (44±1%), the great majority of Reln+/Cxcl14+ cells were found there (88±6%). Consistent with the expected properties of continent 8 cells, a large fraction of the Cxcl14 population were immunoreactive for pro-CCK (62±6%; Figure 7C), while substantial minorities were positive for CALB1 (29±2%; Figure 7D) or Npy (25±5%; Figure 7E). However, as expected from the lack of Cxcl14 in MGE-derived neurogliaform and IS-3 cells, we observed no overlap of Cxcl14 with Nos1 (0 out of 209 cells; Figure 7F), and very weak overlap with Kit, which is primarily expressed in the Cxcl14-negative CGE-neurogliaform population (1 of 264 cells respectively, from all mice; Figure 7G).

The cluster Cck/Cxcl14/Vip presented a puzzle, since Cxcl14 is located primarily at the sr/slm border, whereas...
immunohistochemistry in rat has localized CCK/VIP basket cells to sp (Acsady et al., 1996). Because Cxcl14 expression can sometimes also be found in sp, we tested whether this cluster reflects sp cells, by combining in situ hybridization for Cxcl14 with immunohistochemistry against VIP in mouse CA1 (Figure 8). This revealed frequent co-expression at the sr/slm border (8 ±1% Cxcl14 cells positive for Vip; 23 ±1% Vip cells positive for Cxcl14), but very few Cxcl14 cells in sp, and essentially no double labelling (1 of 147 Vip cells in sp was weakly labelled for Cxcl14). We therefore conclude that this cluster indeed represents a novel cell type located at the sr/slm border, expressing Cck, Vip, and Cxcl14.

Discussion

The molecular architecture of CA1 interneurons has been intensively studied over the last decades, leading to the identification of 23 inhibitory classes. Our transcriptomic data showed a remarkable correspondence to this previous work, with all previously-described classes identified in our database. Our analysis also revealed a continuous mode of variability common across multiple cell types, si hypothesized novel molecular subdivisions of previously described cell types.

Our results suggest that three of the previous described CA1 cell classes may in fact represent a single cell type. The Sst.Nos1 class is strongly positive for Nos1, and also expresses Sst, Npy, Chrm2, Pep4 and Penk, but unlike Penk-positive interneuron-selective cells of continent 9 lacks Vip. This class is homologous to the Int1 and Sst Chodl classes defined in isocortex, which have been identified with long-range projecting sleep-active neurons (Geraschenko et al., 2008; Magno et al., 2012; Tasic et al., 2016; Zeisel et al., 2015). Because previous studies tested different subsets of these molecules, the equivalence of these types may have been overlooked. For example, the PENK-immunopositive neurons with projections to subiculum reported in rat were shown to be VIP-negative, but not tested for SST or NOS1 (Fuentealba et
al., 2008b); the NADPH diaphorase-labelled (i.e. strongly NOS1 positive) axons reported by Sik et al. (1994) as projecting to CA3 and dentate were not labelled for SST or PENK; and the SST/NOS1 cells identified by Jinno and Kosaka (2004) in mouse were not tested for long-range projections or for PENK. While it remains possible that a larger transcriptomic sample of these rare neurons would reveal subclasses, our present data suggest that Sst.Nos1 cells are a homogeneous population. We therefore suggest that they constitute a population of inhibitory neurons with diverse long-range projection targets. Interestingly, the targets of PENK-positive projection cells are most commonly PVALB-positive interneurons, unlike conventional IS cells, which preferentially target SST cells (Fuentealba et al., 2008a). If these do indeed correspond to sleep-active neurons, this fact may provide an important clue to the mechanisms underlying sleep in cortical circuits.

The match between our transcriptomic analysis and previous immunohistochemical work (primarily in rat) is so close that it is simpler to describe the few areas of disagreement than the many areas of agreement. First, ACTN2 has been used as a neurogliaform marker in rat (Price et al., 2005), but was almost completely absent from any cell type of our database. We suggest this reflects a species difference, as previous attempts with multiple ACTN2 antibodies have been unsuccessful in mouse (JHL, unpublished observations), and Actn2 labelling is not detectable in the Allen atlas (Lein et al., 2007). A second inconsistency regards NCALD, which in rat was reported not to overlap with PVALB, SST, or NPY (Martinez-Guijarro et al., 1998), but did so in our data. Finally, it has previously been reported that a subset of O-LM cells show Htr3a expression (Chittajallu et al., 2013). In our data, we observed at most weak expression of Htr3a in Sst cells, and the cells showing it belonged to clusters identified as hippocamposeptal rather than O-LM cells.

Our analysis revealed several rare and presumably novel cell groups, although we cannot exclude that some of these were inadvertently included from areas bordering CA1. Sst.Npy.Serpine2 and Sst.Npy.Mgat4c, which simultaneously expressed Sst, Npy, and Rhl fit the expected expression pattern of neither O-LM nor hippocamposeptal cells; Sst.Erbb4.Rgs10 is a distinct group related to Pvalb basket and bistratified cells; Cck.Lypd1 formed a rare and highly distinct class expressing Cck, Scl17a8, and Calb1 that may be localized near the CA3 border; Ntn5.I Synpr showed an expression pattern with features of both sr/slm Cck neurons and projection cells; and Cck.Cxcl14.Vip represents a cell class strongly positive for both Cck and Vip located at the sr/slm border that appears to represent a pyramidal- rather than interneuron-targeting class. The analysis also revealed subdivisions of known types, such as the division of IS-3 cells into Nos1 positive and negative groups, and the division of CGE-NGF cells into Car4- and Cxcl14-expressing subtypes. Finally, our data suggested that with further data, even rarer types are likely to be found, such as a small group of cells with features of both basket and axo-axonic cells located off the coast of continent 3. Indeed, such cells have been encountered by quantitative electron microscopic analysis of synaptic targets in the rat (P. S., unpublished observations).

Latent factor analysis revealed a common continuum of gene expression across the database, with differences both between clusters, and within the cells of a single cluster. Many of the genes positively correlated with the latent factor are associated with fast-spiking phenotype, and this relationship was confirmed using patch-seq in Pvalb interneurons. Several other classes of genes were correlated with the latent factor including genes related to presynaptic function, GABA release, metabolism, or excitatory and inhibitory inputs (including GABAs). Clusters with high mean values of the factor were identified with interneuron types targeting pyramidal cell somas or proximal dendrites (such as Pvalb or Cck/Scl17a8 expressing basket cells), while those with low mean values were identified with interneurons targeting pyramidal distal dendrites (such as Sst or Cck/Calb1 expressing dendrite-targeting cells) or other interneurons. This correlation makes functional sense: because distal dendritic inputs will be subject to passive dendritic filtering, their vesicle release does not need to be so accurately timed, as reflected in lower expression of fast-timing genes such as Kcncl1 (Hu et al., 2014). I-S cells had the lowest mean values of the latent factor, consistent with their small axonal trees and metabolic machinery (Gulyás et al., 2006). The stronger expression of most neuropeptides in cells of low latent factor suggests that these slower, distal-targeting interneurons may also rely more heavily on neuropeptide signaling, for which slow firing rates support outputs transduced by slower G-protein coupled receptors.

The precise genes modulated by the latent factor varied between cell classes. For example, while Pvalb and Cnr1 both correlated positively with the factor, they were expressed in different sets of neurons. The latent factor analysis suggested several genes as interesting candidates for future research, such as Trp53r11, Yjefn3 and Rgs10, associated with faster spiking Cck cells; Zcchc12 and 6330403K07Rik, both associated with slower-firing cells of all classes; and Fxyd6, associated with slow-spiking that may modulate ion exchange. Intriguingly, we found that genes for neurofilaments and other
intermediate filaments were positively correlated with the latent factor, while genes involved in actin processing were negatively correlated; we speculate that this might reflect a different cytoskeletal organization required for somatic- and dendritic-targeting neurons.

In addition to differences in mean latent factor values between cell types, we also observed continuous gradation within cell types, along the same continuum. This may at least partially reflect activity-dependent modulation of gene expression. Indeed, analysis of the data of Mardinly et al (2016) showed that genes positively related to the latent factor were also more likely to be upregulated by light exposure in visual cortical interneurons. We therefore hypothesize that the latent factor represents a program of gene expression that differs between cell types (presumably due to early developmental factors), but can also be fine-tuned by activity dependence in adulthood, with higher levels of network activity leading to a homeostatic increase in expression of factor-related genes such as those relating to GABA release and interneuron metabolism. Interestingly, the pattern of continuous variation within continent 9 was different to other classes, perhaps indicating a different role for activity-dependent homeostasis in interneuron-selective cells.

The relatively simple cytoarchitecture of CA1 compared to other cortical regions suggests that the organization revealed by these data represent a “lower bound” on the level of complexity that can exist in cortical circuits. For example, it is likely that isocortex has considerably more complex cell-type organization due to its lamination and arealization. An understanding of the multifarious classes of such cells is an essential step toward deciphering their function.

Acknowledgements

We thank T. Viney, A. Joshi, G. Unal, and B. Bekkouche for discussions and comments on the manuscript. This work was supported by the Wellcome Trust (108726 to KDH, NK, PS, SL, and JH-L), Medical Research Council (PS) European Research Council (261063 to SL), Swedish Research Council (TARGET to SL, and 2014-3863 to JH-L), StratNeuro (JH-L), Knut and Alice Wallenberg Foundation (SL) and European Union FP7/Marie Curie Actions (322304 to JH-L).

Author Contributions

KDH initiated collaboration (with JH-L), devised and implemented transcriptomic data analyses (Figs. 1-5, S1-S7), identified cell classes, wrote manuscript.

CBG bred animals, performed Patch-Seq experiments, and analyzed data (Fig. 5).

HG and NGS optimized protocols for adult mice and performed single cell sequencing.

LM performed FISH and IHC experiments and cell counts (Figs. 7,8 and S8C).

LK performed IHC experiments and cell counts (Figs. 6, S8A,B).

PS performed IHC experiments and cell counts (Figs. 6, S8A,B) contributed to cell class identification and manuscript writing.

NK supervised FIH and IHC experiments (Figs. 7,8 and S8C).

SL supervised single-cell sequencing.

JH-L initiated collaboration (with KDH), supervised single-cell and Patch-Seq experiments and analysis, contributed to manuscript writing.
CA1 inhibitory neurons

REFERENCES


Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munuga, H., He,