Check for updates

Circuit and molecular architecture of a ventral hippocampal network

Mark M. Gergues [1,2,8], Kasey J. Han [3,4,8], Hye Sun Choi [3,8], Brandon Brown [4], Kelsey J. Clausing [3,7], Victoria S. Turner [1,2], Ilia D. Vainchtein [2,3], Anna V. Molofsky [1,2,3,5] and Mazen A. Kheirbek [1,2,3,5,6] ✉

The ventral hippocampus (vHPC) is a critical hub in networks that process emotional information. While recent studies have indicated that ventral CA1 (vCA1) projection neurons are functionally dissociable, the basic principles of how the inputs and outputs of vCA1 are organized remain unclear. Here, we used viral and sequencing approaches to define the logic of the extended vCA1 circuit. Using high-throughput sequencing of genetically barcoded neurons (MAPseq) to map the axonal projections of thousands of vCA1 neurons, we identify a population of neurons that simultaneously broadcast information to multiple areas known to regulate the stress axis and approach–avoidance behavior. Through molecular profiling and viral input–output tracing of vCA1 projection neurons, we show how neurons with distinct projection targets may differ in their inputs and transcriptional signatures. These studies reveal new organizational principles of vCA1 that may underlie its functional heterogeneity.

Over the last decade, the vHPC has been increasingly recognized as a critical node in the extended limbic circuitry that controls motivated and emotional behavior[1]. The nature of the region's functions, for example, controlling anxiety-related behaviors[2–7], influencing reward-seeking behavior[2,8–10] and modulating the neuroendocrine response to stress[11], have stoked considerable interest in the potential contribution of the vHPC to mood and anxiety disorders.

The vCA1 is highly functionally and anatomically heterogenous and sends outputs to numerous target regions. Studies using retrograde-tracing approaches indicate that vCA1 projection neurons are largely segregated, with only a small fraction of cells projecting to multiple downstream targets[4,12–15]. This has led to the hypothesis that divergent outputs from vCA1 encode different kinds of stimuli and differentially contribute to distinct aspects of behavior[2,4,13,15]. For example, it has recently been shown that vCA1–lateral hypothalamus (LH) and vCA1–medial prefrontal cortex (mPFC) projecting neurons modulate anxiety-related behavior[4,6,13,16], whereas vCA1–amygdala projections contribute to contextual fear encoding and renewal[15], and vCA1–nucleus accumbens (NAc) neurons encode reward-related information[2,8,9]. In addition, cells that project to more than one region have also been suggested to have specialized roles. In rats, cells with collateral projections to the NAc, mPFC and amygdala are, for example, most active during sharp wave ripples in a task-dependent manner, and those projecting to both infralimbic and prelimbic regions of the mPFC and those projecting to the amygdala and mPFC are implicated in fear renewal[2,14,17].

But while heterogeneity has been well documented at the functional level, the basic principles of how inputs and outputs of the vCA1 are organized remain unclear. Here, we use a combination of genetic and anatomical tools to address the heterogeneity of vCA1 by asking three questions: (1) what proportion of individual vCA1 neurons project to single versus multiple downstream targets? (2) do neurons that project to different downstream targets have distinct presynaptic inputs? and (3) are vCA1 neurons that project to different downstream areas transcriptionally distinct?

Results

MAPseq for high-throughput axonal tracing of vCA1 neurons.
We first sought to understand the diversity of projection patterns of individual vCA1 neurons, to extend earlier work that relied on either retrograde axonal tracing or single-neuron tracing/reconstruction to map vHPC outputs[4,12–15,18,19]. These studies were limited by the number of fluorescent marker colors available to retrogradely label projection neurons and the low-throughput nature of single-neuron tracing/reconstruction, respectively. We instead performed high-throughput single-neuron tracing with multiplexed analysis of projections by high-throughput sequencing (MAPseq)[20,21].

We used viral vectors to label vCA1 neurons with a library of random RNA sequences (barcodes) and allowed them to transport into the axonal processes, such that each neuron was expected to take up a unique but random barcode (Methods; refs. [20,21]). Then, we dissected seven known target regions[19,22,23] and sequenced the barcodes present to determine which neurons projected to which region (Fig. 1a). The relative abundance of each individual barcode was used as a proxy for projection strength (density of axons in the tissue) to that target[20,21]. Target regions were chosen for their putative roles in vHPC-mediated control of emotional behavior: mPFC[2,6], NAc[8,9], LH[4], lateral septum (LS)[13], bed nucleus of stria terminalis (BNST)[24], basal amygdala (BA)[4,12,15] and central amygdala (CeA)[15] (Fig. 1a,b).

We analyzed the axonal projection patterns of 2,494 vCA1 neurons from 12 mice to the seven downstream target regions. We found that ~77% (1,920/2,494) of barcoded vCA1 neurons projected to only one of the assayed target regions, ~18.2% (453 of 2,494) sent axons to two targets, ~3.7% (93 of 2,494) to three

[1]Neuroscience Graduate Program, University of California, San Francisco, San Francisco, CA, USA. [2]Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA. [3]Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, USA. [4]School of Medicine, University of California, San Francisco, San Francisco, CA, USA. [5]Kavli Institute for Fundamental Neuroscience, University of California, San Francisco, San Francisco, CA, USA. [6]Center for Integrative Neuroscience, University of California, San Francisco, San Francisco, CA, USA. [7]Present address: Program in Neuroscience, Harvard Medical School, Cambridge, MA, USA. [8]These authors contributed equally: Mark M. Gergues, Kasey J. Han, Hye Sun Choi. ✉e-mail: mazen.kheirbek@ucsf.edu
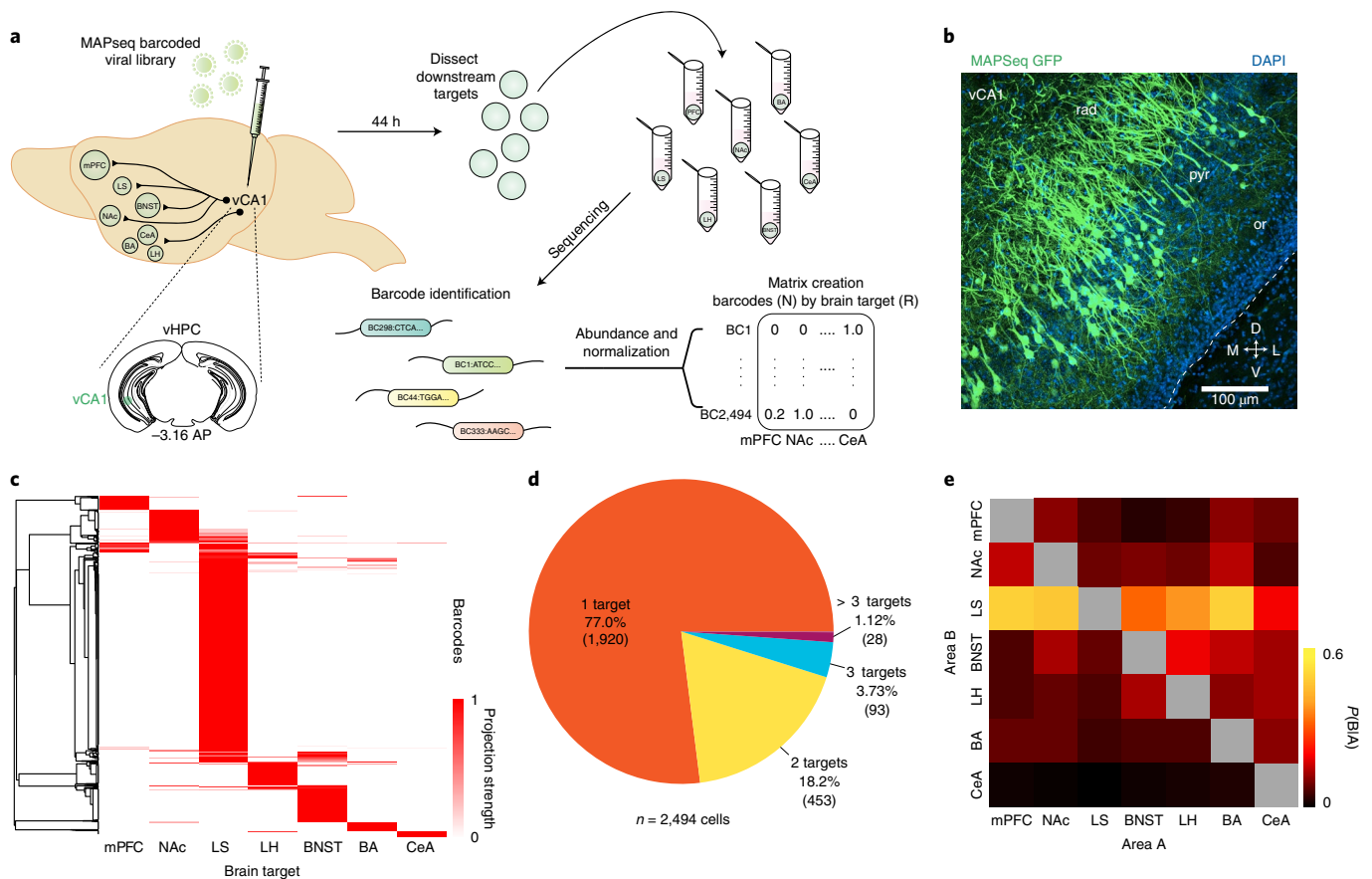
Fig. 1 | High-throughput mapping of vCA1 projections using MAPseq. a, Experimental design. We mapped the projections from vCA1 to seven downstream targets using MAPseq. Briefly, we injected a virus encoding random RNA sequences (barcodes) into vCA1 and allowed them to transport into the axonal processes. Then, we dissected seven known target regions and sequenced the barcodes present to determine which neurons projected to which region. The relative abundance of a barcode present in each area was used as a proxy for the relative projection strength, and an $N \times R$ matrix was generated for analysis (N, barcodes; R, brain regions). See Methods for details on each step of the experimental pipeline. **b**, Expression of MAPseq green fluorescent protein (GFP) at the injection site in vCA1. Images are representative of $n = 4$ mice. **c**, Heat map of projection strengths (from $N \times R$ matrix) of all 2,494 vCA1 neurons (from 12 mice) to seven targets mapped with MAPseq. Each row represents an individual neuron's projection strength to each of the seven target regions, normalized to the maximal value in that row, resulting in a projection strength scale from 0 to 1 (Methods). **d**, Distribution of the number of projection targets of mapped vCA1 neurons. **e**, Heat map of conditional probability for two regions (Methods) indicating the proportion of cells projecting to area B within the subset of cells that project to area A.

targets and ~1.1% (28 of 2,494) to greater than three of the assayed targets (Fig. 1c,d). Analysis of the conditional probability for two regions receiving input from the same neurons suggested that BNST and LS were most likely to receive input from cells targeting multiple regions (Fig. 1e). To further explore the targeting patterns of neurons projecting to multiple downstream areas, we determined whether any projection motifs were overrepresented or underrepresented within our population of barcoded neurons. Here, we defined a projection motif as a projection pattern shared by group of cells, regardless of the projection strength of each cell to any given target (all projection motifs are available in Supplementary Table 1). For example, a BNST–BA projection motif includes cells that only project to both the BNST and the BA and not to any other assayed targets. We focused our analysis on those cells that projected to more than one target (Fig. 2a). We compared our experimental data to a null model that assumes each neuron has no preferences among the possible motifs, and that its probability of projecting to one region is independent of its probability of projecting to another region (that is, its projections are binomially distributed; Fig. 2b,c and see Supplementary Table 1 and Methods). We found 9 motifs that were overrepresented and 15 motifs were underrepresented (Fig. 2b,c). The most overrepresented projection motifs were those

cells that bifurcated to the LS and NAc and to the LS and BNST. The most underrepresented motifs were those cell that bifurcated to the LH and BA and to the BA and CeA, which supports previous fluorescent retrograde-tracing studies[4,15]. For each projection motif, we plotted the projection strength (abundance of barcode sequences in a downstream target, which infers the axon density in the tissue[21]) for each cell, normalized by the maximum projection strength for each cell (Fig. 2d,e and Extended Data Fig. 1b). As in the heat map in Fig. 1c, each line (different colors) is the projection strength of an individual neuron to each of the seven target regions and is normalized to the maximal value in that row, resulting in a projection strength scale from 0 to 1. Within these motifs, neurons showed heterogeneity; cells projected to each area with differing strengths (Fig. 2d,e and Extended Data Fig. 1b). These MAPseq data reveal the heterogeneity of vCA1 projection neurons, as many vCA1 cells projected to only one of the seven targets assayed and a notable proportion projected to multiple targets in a non-random fashion.

Input–output viral tracing of vCA1 neurons that project to NAc, mPFC, LH, BA, adBNST and LS. We next determined whether vCA1 output neurons previously implicated in approach/avoidance behavior and fear learning receive long-range, extrahippocampal
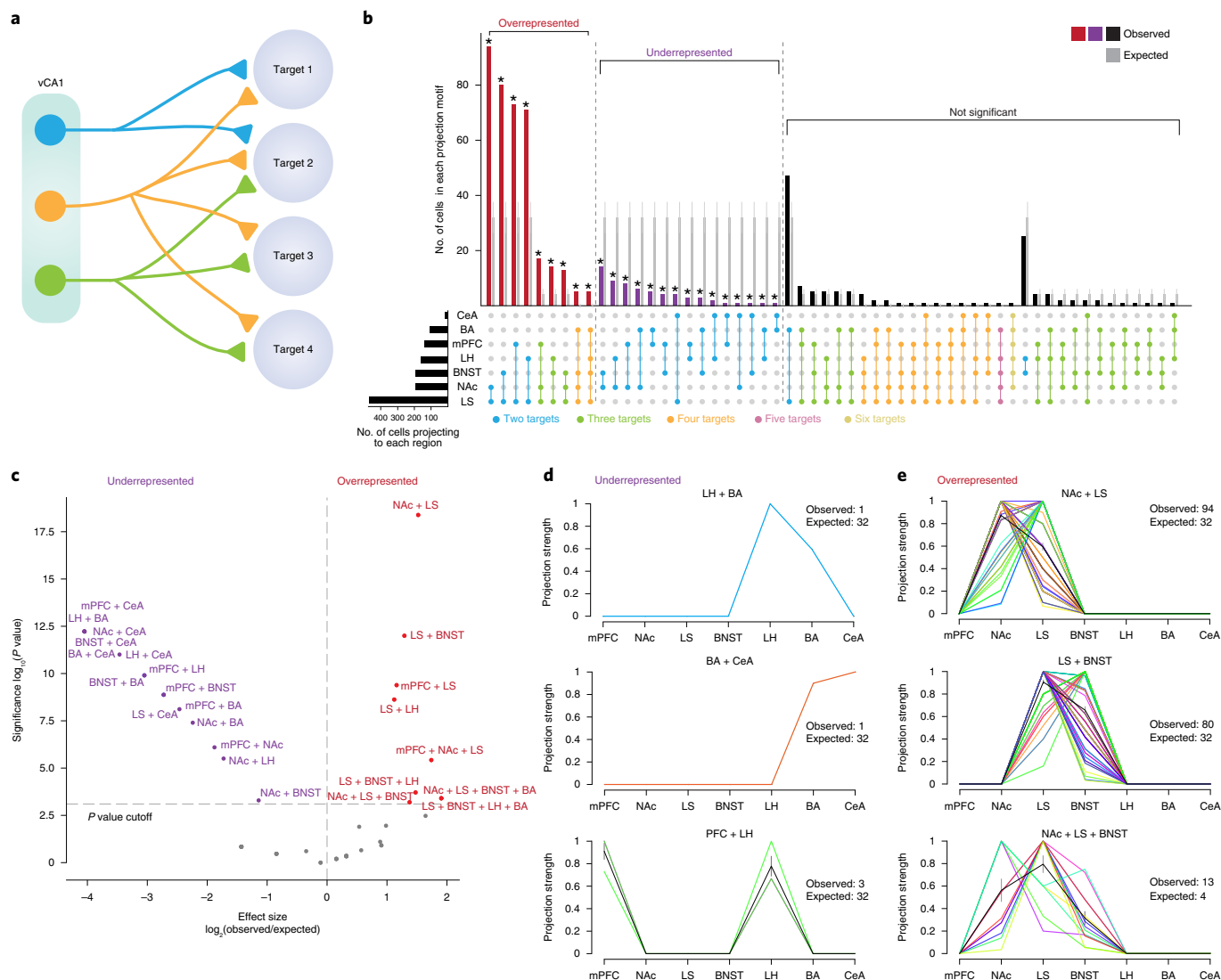
Fig. 2 | vCA1 projection motifs. a, Schematic of different kinds of projection motifs with multiple targets. A motif is defined as a projection pattern shared by group of cells. **b**, UpSet plot of the number of cells within each observed projection motif. Colored bars represent observed motif numbers from $n = 574$ total cells with multiple targets (number of cells in each motif is shown in Supplementary Table 1). We compared the observed motif number with expected from the null model. Gray bars represent expected counts, with the mean and standard deviation of the null binomial models (Methods). P values were calculated from two-sided binomial tests, adjusted for multiple comparisons using the Bonferroni method; $*P_{adj} < 0.05$ (P values and effect sizes are provided in Supplementary Table 1). **c**, Plot of statistical significance and effect size of over and underrepresented projection motifs; see Methods for generation of the null model ($n = 574$ cells, cell number in each motif; P values and effect sizes provided in Supplementary Table 1). **d,e**, Projection strengths of individual neurons within the underrepresented group (**d**) and overrepresented group (**e**). Each colored line is an individual neuron's projection strength to each of the seven target regions, normalized to the maximal value in that row, resulting in a projection strength scale from 0 to 1 (see Methods for normalization). The black line represents the mean projection strength for all neurons in that motif, and the gray line represents the s.e.m. Each motif is labeled above each graph. Number of biologically independent cells: LH + BA ($n = 1$), BA + CeA ($n = 1$), PFC + LH ($n = 3$), NAc + LS ($n = 94$), LS + BNST ($n = 80$) and NAc + LS + BNST ($n = 13$). Insets show observed and expected cell numbers. All other motifs are available in Extended Data Fig. 1.

inputs from similar or diverse upstream regions. Identifying vCA1 cells that extend axons to mPFC, NAc, LS, anterodorsal BNST (adBNST), BA and LH, we cataloged the neuronal patterns of long-range inputs using the 'tracing the relationship between input and output' (TRIO) method[25].

Briefly, we injected a retrograde traveling Cre-recombinase (AAV2-retro-CAG-Cre[26]) into one of each of the six downstream target areas and rabies helper virus (RV) into vCA1 (AAV 1-synP-FLEX-TVA-P2A-EGFP-P2A-B19G[27]), which was followed 3 weeks later by RV-EnvA-ΔG-mCherry injection into the same region of vCA1 (Fig. 3a–c and Extended Data Fig. 2f,g). We compared the sources of inputs to these six vCA1 output streams by

brain-wide analysis of long-range, extrahippocampal input neurons (Fig. 4a and Methods). Overall, these six groups of output neurons did not differ in their input patterns, as they received input from the same upstream areas (Fig. 4a,b, Extended Data Fig. 3a–e and Supplementary Table 2). However, there were a few differences regarding the proportion of inputs from some areas to specific output neurons. First, while all projection neurons received thalamic input, most notably input from the paraventricular nucleus of the thalamus (PVT), TRIO suggested that vCA1–LH projection neurons receive proportionally more input from PVT than the projections to BA and mPFC (Fig. 4c, Extended Data Fig. 2a and Supplementary Table 2). Second, TRIO predicted that vCA1–BA projection
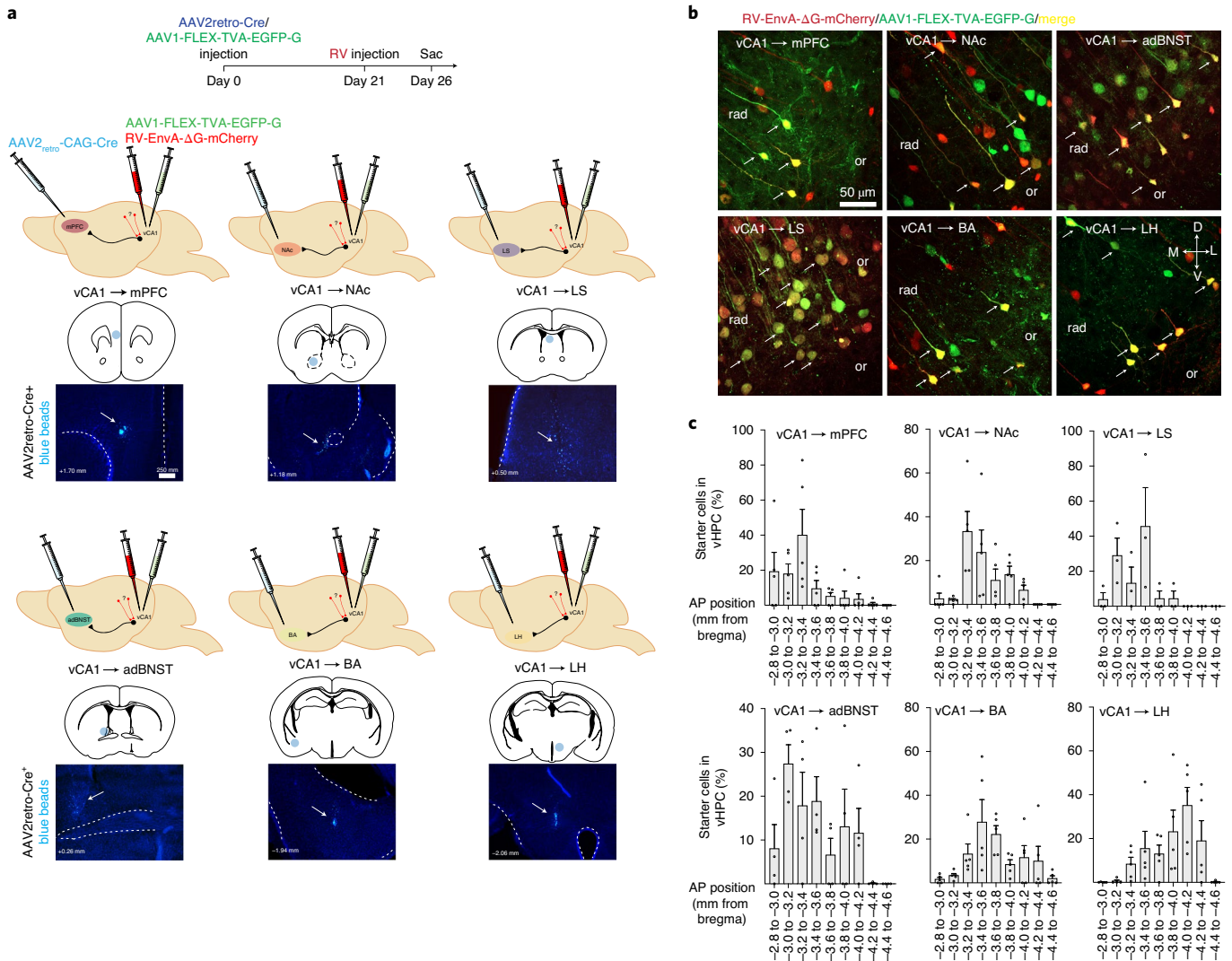
Fig. 3 | Viral input–output tracing of vCA1. a, Schematics and timeline for input–output rabies tracing. We targeted six vCA1 outputs, the NAc, LH, BA, mPFC, adBNST and LS, with Cre-dependent rabies helper virus and EnvA-g deleted rabies virus expressing mCherry injected into the targetted regions and with injection of AAV2retro-Cre into vCA1 (Methods). Brain schematics show the injection sites, labeled with blue beads, representative from $n = 6$ (vCA1–mPFC), 8 (vCA1–NAc), 9 (vCA1–LH), 9 (vCA1–BA), 4 (vCA1–adBNST) and 3 (vCA1–LS) mice. **b**, Starter cells defined by coexpression of GFP and mCherry in vCA1 in the six targeted output populations, representative from $n = 5$ (vCA1–BA), 4 (vCA1–adBNST), 3 (vCA1–LS), 5 (vCA1–mPFC), 5 (vCA1–NAc) and 5 (vCA1–LH) mice. **c**, Position of starter neurons counted along the AP axis of the vHPC for each projection population ($n = 5$ vCA1–BA, 4 vCA1–adBNST, 3 vCA1–LS, 5 vCA1–mPFC, 5 vCA1–NAc and 5 vCA1–LH mice). Error bars represent the s.e.m.

neurons receive proportionally more input from the basal forebrain, most notably from the nucleus of the diagonal band, than vCA1–LS and vCA1–BNST projections (Fig. 4c, Extended Data Fig. 2e and Supplementary Table 2). And third, TRIO predicted that vCA1–BNST projection neurons get proportionally more input from the lateral amygdala than the other projections assayed (Fig. 4c, Extended Data Fig. 2b and Supplementary Table 2). These results suggest that these six vCA1 projection streams receive, in general, similar upstream inputs, but that there are a number of small biases in the strength of some of these inputs to some outputs.

Molecular profiling of vCA1 neurons that project to the NAc, mPFC, LH and BA. Finally, we determined whether vCA1 neurons projecting to four target areas—LH, NAc, BA and mPFC—differ in terms of the genes they express. To do so, we used transcriptional profiling methods that accessed translating mRNAs in vCA1 neurons defined by their projection to one of these four targets.

We labeled vCA1 neurons that projected to the mPFC, NAc, BA and LH in different cohorts of Rosa26[FsTRAP] mice (using the same AAV2retro-Cre approach as with TRIO above) and then used a translating ribosome affinity purification (TRAP) approach to profile translation in each of these populations (Fig. 5a)[28]. RNA sequencing (RNA-seq) was then performed to identify differentially enriched mRNAs in the distinct vCA1 projection neurons (Fig. 5a and Supplementary Tables 3–5).

Pairwise comparisons of vCA1 projections between each vCA1 projection neuron type revealed 24 genes that were significantly differentially expressed (Extended Data Fig. 3a and Supplementary Table 4). Interestingly however, these comparisons revealed a unique transcriptional profile in vCA1–mPFC projections relative to other subcortical projections (Fig. 5b, Extended Data Fig. 3a,b and Supplementary Table 4). Thus, we performed analysis of vCA1–mPFC versus vCA1–subcortical regions and found 653 genes that were differentially expressed ($P_{adj} < 0.05$; Fig. 5c and Supplementary

Table 4). Gene ontology analysis of 481 mPFC-enriched genes revealed strong enrichment for metabolic genes, particularly those involved in oxidative phosphorylation, that have also been implicated in neurodegenerative diseases such as Alzheimer's, Huntington's and Parkinson's diseases (Fig. 5d,e and Supplementary Table 5).

To further validate the enriched genes identified in our dataset, we performed in situ hybridization to examine cerebellar degeneration-related antigen 1 (*Cdr1*), which was one of the top genes enriched in vCA1–mPFC neurons (Extended Data Fig. 3c). Quantification of *Cdr1* transcript in vCA1 of mice injected with a retrograde tracer, cholera toxin subunit B (Alexa fluor 555 conjugate: CTB-555), in the mPFC revealed the majority of vCA1–mPFC neurons expressed *Cdr1*. Notably, unsupervised clustering analysis suggested that this dataset was not able to resolve unique gene expression signatures for all of the vCA1 projection populations examined; thus, we cannot rule out potential transcriptomic differences among the subcortically projecting neurons (Extended Data Fig. 3d,e). In summary, our transcriptomic data highlight a clear gene expression signature that distinguishes mPFC-projecting vCA1 neurons from subcortical-projecting neurons.

Discussion

Together, these data enrich our understanding of the organization of the vCA1—how neurons there differentially connect to upstream inputs, how activity there disseminates to downstream areas and how neurons display subtly different molecular identities. We find that vCA1 projections show not only show a one-to-one connectivity with downstream areas, but that a notable fraction of neurons project to multiple downstream areas in a non-random fashion. Using rabies-mediated input–output tracing of six vCA1 projection streams, we find that, overall, these six groups of output neurons received a similar distribution of inputs from upstream subcortical areas, with a few differences in the proportion of those inputs. Finally, analysis of translating mRNAs revealed a gene expression signature that differentiates vCA1–mPFC projections from those neurons that project to the LH, BA and NAc.

Our MAPseq data describing the organization of vCA1outputs expands on previous studies that used bulk axonal tracing, retrograde fluorescent tracers and single-cell reconstruction to understand vCA1 (refs. [4,14,15,17–19,22,23]). While this work has provided considerable insight into the organization of vCA1, these techniques lack single-cell specificity (namely bulk anterograde tracing), the number of colors available (retrograde tracing), and single-neuron reconstruction is labor intensive, low throughput and limited in the number of neurons reconstructed. With MAPseq, we

could scale up analysis of vCA1 in all these respects by increasing the number of neurons and downstream targets that could be assayed simultaneously (seven targets and ~2,500 neurons in this study). Statistical analysis on the fraction of neurons that project to more than one downstream target allowed us to find some order in the organization of single-neuron projection patterns. We discovered two primary organizational principles: one that had been documented before, and one that is new to this report. In line with previous studies of vHPC[4,12,14,15,17,29,30], we found a large fraction of cells project to one assayed area, while a smaller proportion project to multiple areas. In addition, our data presents a new finding; those neurons that project to multiple areas do so in a non-random way (these projections are discussed in detail below). We found that some projection patterns of single neurons were significantly more abundant than expected, while other targeting patterns were less abundant. This would indicate that there is some organization in vCA1 projection patterns, with some downstream areas receiving input from the same cells, while others receive input from largely distinct neurons. This non-random organization in broadcasting neurons has recently been described in the primary visual cortical outputs to higher-order areas[20]. However, it is important to note the limitations of MAPseq, namely the spatial resolution of the target dissection and the lack of cell-type-specific genetic information of the cells traced. Future studies using higher-resolution dissection techniques and newer variants of sequencing-based tracing techniques (such as BARseq[31]) that provide cell-type specificity will help in these regards.

Comparing our results to those using retrograde approaches in the vHPC reveals a number of similarities, allowing us to both confirm and expand upon previous results. In the majority of these studies, two projection targets were assayed. For example, a recent study from our group analyzed vCA1 neurons that projected to the LH and BA and found that, using retrograde tracers, 3.4% of neurons sent projections to both targets[4]. In line with this, our MAPseq data found 4.03% of neurons projecting to the BA and LH target both areas (see Supplementary Table 1 for all motifs). Another recent study using retrograde tracing to analyze projections from the vHPC (in the ventral subiculum) to the mPFC, LH and NAc found that 2% of neurons projected to both the mPFC and LH, 5% to the mPFC and NAc and 6% to the NAc and LH[30]. Again, this is strikingly similar to our MAPseq vCA1 data, in which we found 2.6% of neurons projected to the mPFC/LH, 6.42% to the mPFC/NAc and 4.22% to the NAc/LH. We again found similarities comparing our amygdala results to those of a recent study using retrograde viral approaches; their findings showed that 3% of vHPC neurons projected to both the BA and CeA[15], compared to 2.3% in our MAPseq

Fig. 4 | RV-labeled inputs to vCA1 output neurons. a, Representative images of extrahippocampal long-range input neurons in the thalamus, amygdala and basal forebrain (from $n = 6$ vCA1-mPFC, 8 vCA1-NAc, 9 vCA1-LH, 9 vCA1-BA, 4 vCA1-adBNST and 3 vCA1-LS mice). **b**, Fraction of total extrahippocampal input neurons found in the thalamus, amygdala, basal forebrain, midbrain/hindbrain and hypothalamus. Significance was assessed using one-way ANOVA followed by post hoc comparisons. All tests were two-sided. *$P_{adj} < 0.05$ for post hoc comparisons after Holm–Sidak correction for multiple comparisons. Thalamic nuclei: vCA1-LH versus vCA1-BA, $t_{33} = 3.652$, $P_{adj} = 0.0133$; vCA1-mPFC versus vCA1-LH, $t_{33} = 3.008$, $P_{adj} = 0.0679$. Basal forebrain/septal nuclei: NAc versus adBNST, $t_{33} = 3.649$, $P_{adj} = 0.0134$; BA versus adBNST, $t_{33} = 3.551$, $P_{adj} = 0.0164$. All statistical tests and P values are provided in Supplementary Table 2. **c**, Of the five areas in **b**, we analyzed the proportion of input to each projection population in 15 input areas using one-way ANOVA, followed by post hoc comparisons. All tests are two-sided. *$P < 0.05$ for post hoc comparisons after Holm–Sidak correction for multiple comparisons. PVT: CA1-LH versus vCA1-BA, $t_{33} = 3.421$, $P_{adj} = 0.0249$; vCA1-mPFC versus vCA1-LH, $t_{33} = 2.966$, $P_{adj} = 0.0752$. LA: mPFC versus adBNST, $t_{33} = 3.243$, $P_{adj} = 0.0294$; NAc versus adBNST, $t_{33} = 4.358$, $P_{adj} = 0.0018$; LH versus adBNST, $t_{33} = 4.013$, $P_{adj} = 0.0045$; BA versus adBNST, $t_{33} = 3.612$, $P_{adj} = 0.0129$; adBNST versus LS, $t_{33} = 3.320$, $P_{adj} = 0.0261$. NDB: BA versus adBNST, $t_{33} = 3.339$, $P_{adj} = 0.0309$; BA versus LS, $t_{33} = 3.133$, $P_{adj} = 0.0495$. See Supplementary Table 2 for proportion counts (mean ± s.e.m.) for assayed input regions, statistical tests and P values. Error bars represent the s.e.m. **d**. Schematic of input–output connection patterns in vCA1. vCA1 largely integrates input from upstream areas, with some biases in the proportion of input (darker lines) from the thalamus, amygdala and basal forebrain. Then, this information is output to the six targets by cells that either form one-to-one connections (black) with a given target, or those that branch to multiple downstream areas (light green). See Fig. 2 and Supplementary Table 1 for different branching motifs. RE, nucleus of reuniens; PT, parataenial nucleus; Xi, xiphoid thalamic nucleus; BLA, basolateral amygdala; BMA, basomedial amygdala; LA, lateral amygdala; CeA, central amygdala; MS, medial septum; NDB, diagonal band nucleus; DR, dorsal raphe nucleus; MR, median raphe; LHA, lateral hypothalamic area, PVH, paraventricular nucleus of hypothalamus; mPOA, medial preoptic area.

data. Finally, comparing our MAPseq data to those of recent studies that used the herpes simplex virus or CTB to map vHPC outputs to the mPFC and BA revealed ~9% of neurons project to both areas[12,17], whereas we found slightly less, with 4.75% of cells projecting to both areas. In addition, a study using antidromic optogenetic stimulation for tagging vCA1 neurons in rats found a small population that targeted the mPFC, NAc and BA[2], and again these projection cells were found in our dataset. One advantage of MAPseq is that it allowed us to assay targeting of multiple downstream areas simultaneously. This allowed us to find that, in the neurons that projected to the BA,
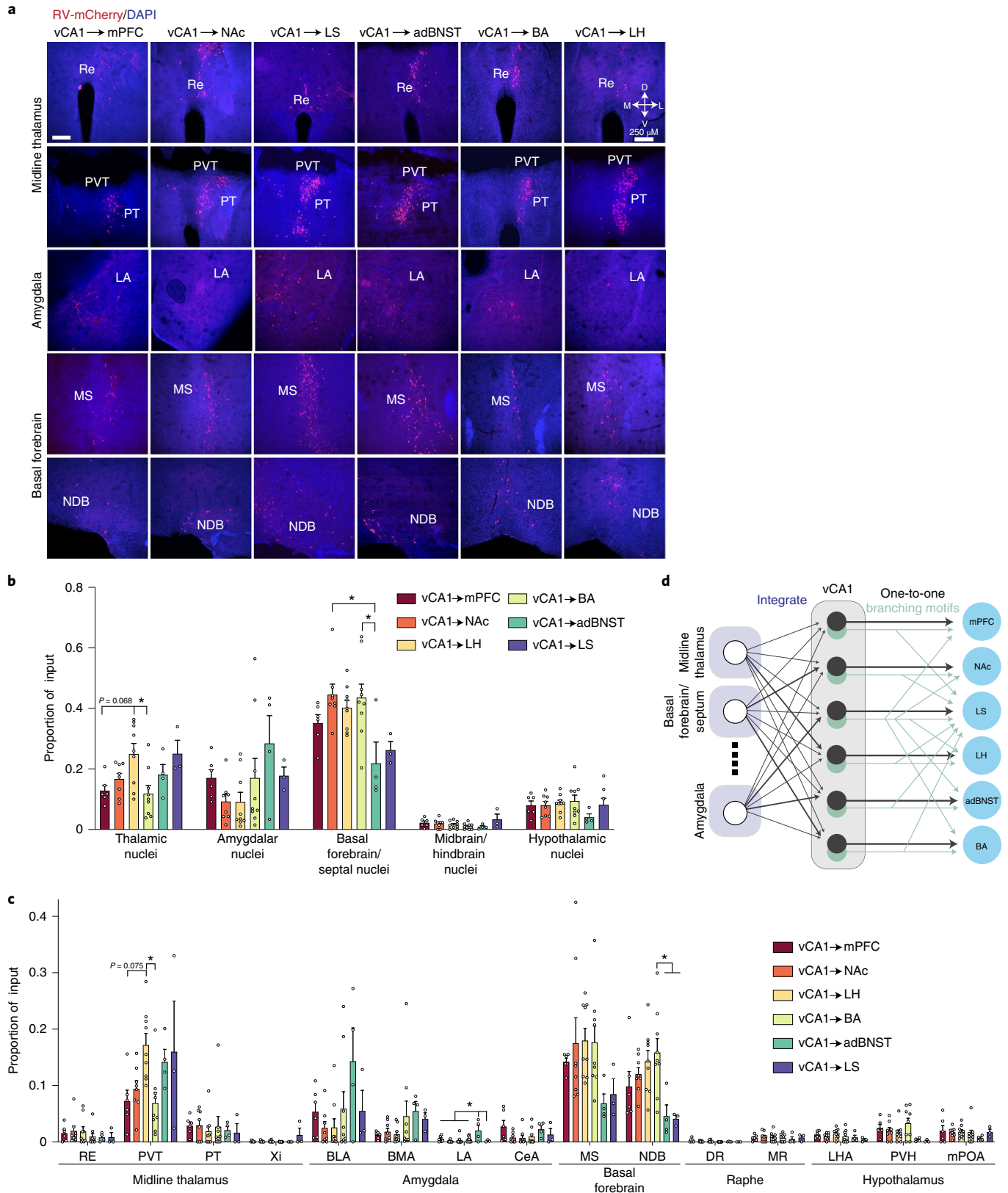
mPFC, NAc and LH, the majority of cells that projected to two of these areas (as assayed by retrograde labeling in the studies above) also sent collaterals to at least one other area (see Supplementary Table 1 for counts of all projection motifs and relative enrichment of distinct motifs). Thus, understanding the heterogeneity of vCA1 neuronal projection patterns and the logic of how they may be organized can aid in the design of future functional experiments to further elucidate the function of the vHPC.

Our dataset supported previous bulk anterograde tracing studies and single-cell reconstruction studies indicating a dense projection from the vCA1 to the LS[18,22,23]. These studies suggest that one of the main output routes of vCA1 neurons connects anterodorsally into the LS[23], and one study traced neurons that simultaneously projected to the LS, NAc and association cortices[18]. Indeed, these projection motifs that included LS and at least one other target were found to be overrepresented in our vCA1 dataset. Analysis of our MAPseq data found an overrepresentation of those vCA1 neurons that sent bifurcating projections to the BNST and LS, LH and LS and trifurcating projections to all three of these downstream areas. From a functional perspective, this organization is intriguing, as modulation of distinct cell types in the vCA1, LS, BNST and LH have all been shown to control distinct aspects of anxiety-related behavior. For example, stimulation of the LS itself or cells in vHPC that project to the LS is anxiolytic, while inhibition of vHPC cells projecting to the LS is anxiogenic[13,32]. In addition, subpopulations of LS neurons are recruited by the open arms of the EPM and during periods of mobility in contexts previously associated with a footshock[32,33]. In the anterodorsal BNST (adBNST), an area long known for its role in anxiety-related behavior[34], neurons have been found to be preferentially activated by safe areas, such as the closed arm of the elevated plus maze (EPM), while stimulation of distinct adBNST projections can reduce distinct anxiety-related features in mice (such as open arm time in the EPM, respiration rate and approach behaviors)[35]. In addition, the BNST, LS and LH are all identified as potential relays between the vHPC and the PVH, the entry point to the hypothalamic–pituitary–adrenal (HPA) axis through which the vHPC may modulate the neuroendocrine responses to stressful stimuli[36]. Thus, studies aimed at understanding how information is encoded by these bifurcating and trifurcating vCA1 neurons, how they may modulate distinct features in anxiety-related behavior and how they may affect the HPA axis to control neuroendocrine responses to stressors will provide new insight into the role of the vCA1 in the control of emotional behavior.

In addition, we found an overrepresentation of vCA1 neurons projecting to the LS that bifurcate into either the mPFC or NAc. Recent studies have shown that the vCA1 projection to mPFC modulates anxiety-related behavior and fear recall, while the vCA1 projection to the NAc modulates reward seeking[2,6,9,16,37]. However, here we show a substantial portion of these vCA1 neurons that project to NAc and mPFC also send a collateral to the LS. Targeting these

bifurcating vCA1 neurons will provide insight into whether they have specialized functional roles that differ from vCA1 neurons with other targets, how they coordinate the activity in the LS, mPFC and NAc to regulate approach/avoidance decisions, and more generally, how stimuli of differing valences may be encoded within subpopulations of vCA1 neurons defined by their connectivity.

Our input–output tracing found that overall, vCA1 projection neurons receive similar upstream input, suggesting that vCA1 acts as a hub to integrate information and broadcast to multiple downstream areas (although in a non-random fashion as indicated by our MAPseq results; Fig. 4d). However, at the integration level, we found some subtle biases in the proportion of inputs to anatomically defined outputs. For example, vCA1–LH projections, when compared to vCA1–mPFC and vCA1–BA neurons, are differentially innervated by inputs from the PVT, while vCA1–BNST neurons receive more input from subregions of the amygdala. The PVT has been shown to be activated by numerous stimuli, including stressors, appetitive and aversive stimuli and changes in internal state[38]. However, what information the PVT transmits to the vCA1 and how it may modulate the output of vCA1 generally, or vCA1–LH specifically, remains poorly understood. Recent studies have shown that the population of vCA1–LH neurons, when compared to vCA1–BA neurons, are enriched in neurons that encode features of anxiety-provoking environments such as the open arms of the EPM, or the center of an open field arena[4]. Targeting the PVT–vCA1 connection will provide insight into how the PVT signal may be integrated at the level of the vCA1 to generate an output signal to drive avoidance-related behaviors. In addition, future studies will elucidate how distinct vCA1 projection populations integrate short-range input (not assayed here), either from local inhibitory circuits or intrahippocampal excitatory afferents from the CA3 hippocampal region.

Finally, analysis of translating mRNAs revealed a number of genes that are differentially expressed in some vCA1 neuronal subpopulations, with the vCA1–mPFC projection having a unique transcriptional profile. The dataset generated with the TRAP method used here did not resolve a unique gene expression signature for each of the vCA1 projection populations; thus, future studies using alternative methods, such as single-nucleus sequencing, may provide the resolution needed to distinguish between these subpopulations of pyramidal neurons. However, our analysis did find that when compared to subcortically projecting vCA1 neurons, vCA1–mPFC neurons were enriched for genes involved in metabolic and oxidative pathways. Future studies aimed at understanding the developmental trajectories, wiring properties or differential sensitivities to stress and damage will lend greater insight into the differences between vCA1–mPFC versus subcortically projecting neurons. In addition to *Cdr1*, we found that neurogranin (*Nrgn*) was one of the top genes enriched in vCA1–mPFC neurons. *Nrgn* has been associated with schizophrenia[39], and its expression has been localized to

Fig. 5 | Transcriptional profiling of vCA1 neurons defined by connectivity. a, Experimental design. Rosa26[fsTRAP] mice were injected with AAV2retro-Cre into either the NAc, LH, BA or mPFC. Expression of EGFP in vCA1 in mice injected in each pathway, and images are representative of *n* = 3 mice per pathway. vCA1 was dissected, and GFP immunoprecipitations (IPs) were performed for analysis of differential expression using RNA-seq (*n* = 6, pooled ventral hippocampi for LH performed in duplicate; *n* = 6, ventral hippocampi for BA performed in triplicate; *n* = 6 pooled ventral hippocampi for NAc performed in triplicate; and *n* = 6 ventral hippocampi for mPFC performed in triplicate). **b**, Heat map of the differentially expressed genes from vCA1–mPFC versus vCA1–subcortical regions. Significant genes passed the threshold of adjusted *P* value < 0.05 as determined by DESeq2. *P* values were acquired by two-sided Wald test corrected for multiple testing with the Benjamini–Hochberg method (Methods and Supplementary Table 4). **c**, Volcano plot depicting differentially expressed genes between vCA1–mPFC versus vCA1–subcortical regions (red dots, genes passing threshold of adjusted *P* < 0.05 as in **b**; NS, not significant). All genes and *P* values are available in Supplementary Table 4. **d**, Metascape enrichment network plot of genes enriched in the mPFC relative to subcortical regions (genes passing threshold of adjusted *P* < 0.05 as in **b**; enriched ontology terms were determined with a two-sided, hypergeometric test, and Benjamini–Hochberg multiple testing correction in Metascape (Supplementary Data Table 5). **e**, Ingenuity pathway analysis (IPA) of canonical pathways for vCA1–mPFC versus pooled vCA1–subcortical targets (genes passing the threshold of adjusted *P* < 0.05 as in **b**). The top five enriched pathways as determined by IPA with *P* < 0.05 (Supplementary Table 5) are shown. Colors indicate pathway activity. eIF2, eukaryotic initiation factor 2; mTOR, mammalian target of rapamycin).

the hippocampus, where it modulates synaptic plasticity, contextual and spatial learning and anxiety-related behavior[40–42]. Future studies will shed light into how Nrgn levels in vCA1–mPFC neurons are regulated by experience, and how Nrgn may differentially modulate information transfer between vCA1 to mPFC over other subcortical

targets. These data not only provide future opportunities for targeting these classes of cells for visualization and manipulation, but also by providing unique molecular signatures of cell types with distinct functions, they may reveal new avenues for testing therapeutic targets.
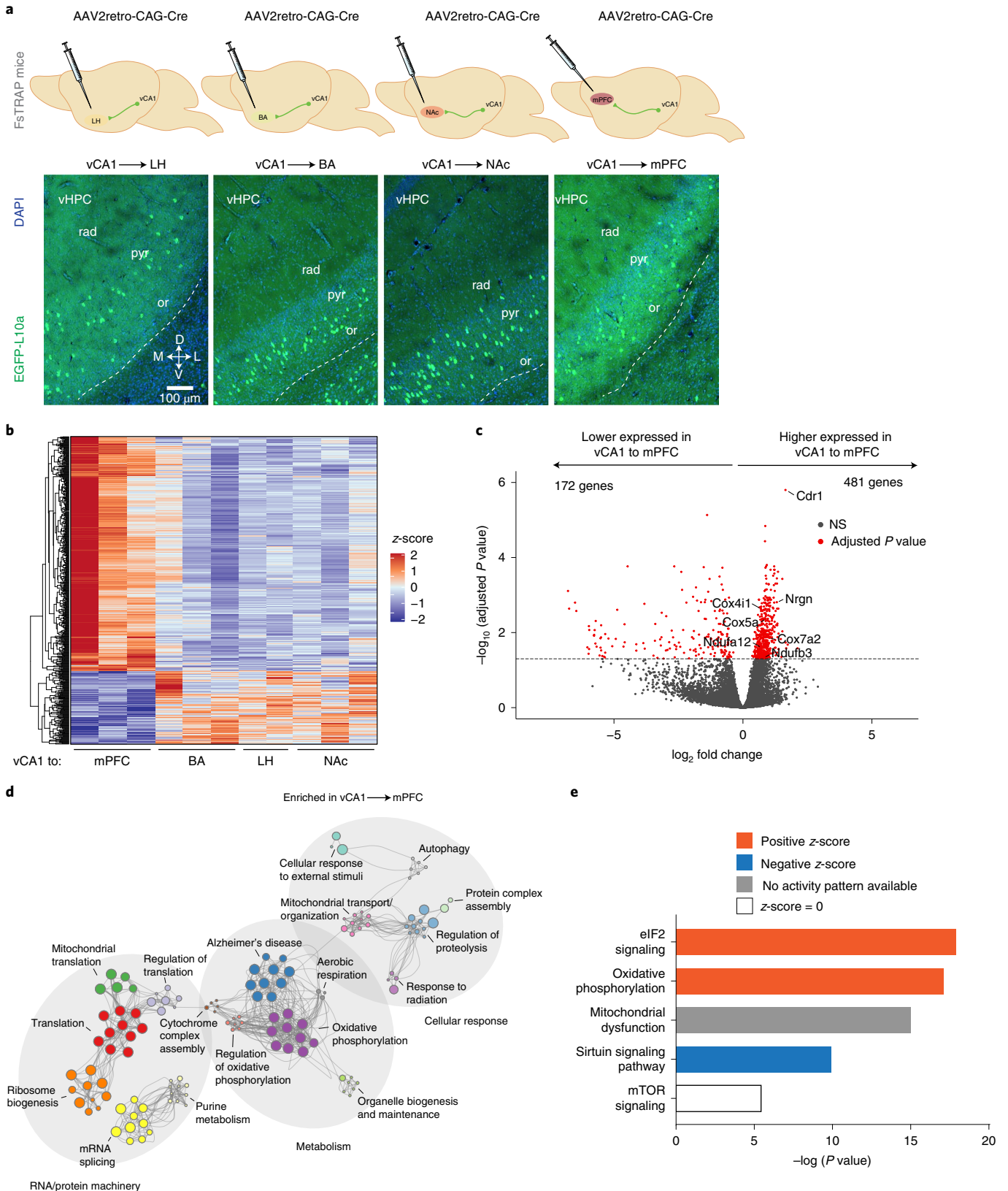
Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-020-0705-8.

References

1. Fanselow, M. S. & Dong, H. W. Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* **65**, 7–19 (2010).
2. Ciocchi, S., Passecker, J., Malagon-Vina, H., Mikus, N. & Klausberger, T. Brain computation. Selective information routing by ventral hippocampal CA1 projection neurons. *Science* **348**, 560–563 (2015).
3. Felix-Ortiz, A. C. et al. BLA to vHPC inputs modulate anxiety-related behaviors. *Neuron* **79**, 658–664 (2013).
4. Jimenez, J. C. et al. Anxiety cells in a hippocampal–hypothalamic circuit. *Neuron* **97**, 670–683 (2018).
5. Kjelstrup, K. G. et al. Reduced fear expression after lesions of the ventral hippocampus. *Proc. Natl Acad. Sci. USA* **99**, 10825–10830 (2002).
6. Padilla-Coreano, N. et al. Direct ventral hippocampal–prefrontal input is required for anxiety-related neural activity and behavior. *Neuron* **89**, 857–866 (2016).
7. Kheirbek, M. A. et al. Differential control of learning and anxiety along the dorsoventral axis of the dentate gyrus. *Neuron* **77**, 955–968 (2013).
8. Britt, J. P. et al. Synaptic and behavioral profile of multiple glutamatergic inputs to the nucleus accumbens. *Neuron* **76**, 790–803 (2012).
9. LeGates, T. A. et al. Reward behaviour is regulated by the strength of hippocampus–nucleus accumbens synapses. *Nature* **564**, 258–262 (2018).
10. Reed, S. J. et al. Coordinated reductions in excitatory input to the nucleus accumbens underlie food consumption. *Neuron* **99**, 1260–1273 (2018).
11. Jacobson, L. & Sapolsky, R. The role of the hippocampus in feedback regulation of the hypothalamic–pituitary–adrenocortical axis. *Endocr. Rev.* **12**, 118–134 (1991).
12. Kim, W. B. & Cho, J. H. Synaptic targeting of double-projecting ventral CA1 hippocampal neurons to the medial prefrontal cortex and basal amygdala. *J. Neurosci.* **37**, 4868–4882 (2017).
13. Parfitt, G. M. et al. Bidirectional control of anxiety-related behaviors in mice: role of inputs arising from the ventral hippocampus to the lateral septum and medial prefrontal cortex. *Neuropsychopharmacology* **42**, 1715–1728 (2017).
14. Wang, Q., Jin, J. & Maren, S. Renewal of extinguished fear activates ventral hippocampal neurons projecting to the prelimbic and infralimbic cortices in rats. *Neurobiol. Learn Mem.* **134**, 38–43 (2016).
15. Xu, C. et al. Distinct hippocampal pathways mediate dissociable roles of context in memory retrieval. *Cell* **167**, 961–972 (2016).
16. Padilla-Coreano, N. et al. Hippocampal–prefrontal theta transmission regulates avoidance behavior. *Neuron* **104**, 601–610 (2019).
17. Jin, J. & Maren, S. Fear renewal preferentially activates ventral hippocampal neurons projecting to both amygdala and prefrontal cortex in rats. *Sci. Rep.* **5**, 8388 (2015).
18. Arszovszki, A., Borhegyi, Z. & Klausberger, T. Three axonal projection routes of individual pyramidal cells in the ventral CA1 hippocampus. *Front Neuroanat.* **8**, 53 (2014).
19. Bienkowski, M. S. et al. Integration of gene expression and brain-wide connectivity reveals the multiscale organization of mouse hippocampal networks. *Nat. Neurosci.* **21**, 1628–1643 (2018).
20. Han, Y. et al. The logic of single-cell projections from visual cortex. *Nature* **556**, 51–56 (2018).
21. Kebschull, J. M. et al. High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron* **91**, 975–987 (2016).
22. Cenquizca, L. A. & Swanson, L. W. Analysis of direct hippocampal cortical field CA1 axonal projections to diencephalon in the rat. *J. Comp. Neurol.* **497**, 101–114 (2006).
23. Cenquizca, L. A. & Swanson, L. W. Spatial organization of direct hippocampal field CA1 axonal projections to the rest of the cerebral cortex. *Brain Res. Rev.* **56**, 1–26 (2007).
24. Glangetas, C. et al. NMDA-receptor-dependent plasticity in the bed nucleus of the stria terminalis triggers long-term anxiolysis. *Nat. Commun.* **8**, 14456 (2017).
25. Schwarz, L. A. et al. Viral-genetic tracing of the input–output organization of a central noradrenaline circuit. *Nature* **524**, 88–92 (2015).
26. Tervo, D. G. et al. A designer AAV variant permits efficient retrograde access to projection neurons. *Neuron* **92**, 372–382 (2016).
27. Kohara, K. et al. Cell-type-specific genetic and optogenetic tools reveal hippocampal CA2 circuits. *Nat. Neurosci.* **17**, 269–279 (2014).
28. Ekstrand, M. I. et al. Molecular profiling of neurons based on connectivity. *Cell* **157**, 1230–1242 (2014).
29. Cembrowski, M. S. & Spruston, N. Heterogeneity within classical cell types is the rule: lessons from hippocampal pyramidal neurons. *Nat. Rev. Neurosci.* **20**, 193–204 (2019).
30. Wee, R. W. S. & MacAskill, A. F. Biased connectivity of brain-wide inputs to ventral subiculum output neurons. *Cell Rep.* **30**, 3644–3654 (2020).
31. Chen, X. et al. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179**, 772–786 (2019).
32. Thomas, E., Burock, D., Knudsen, K., Deterding, E. & Yadin, E. Single-unit activity in the lateral septum and central nucleus of the amygdala in the elevated plus maze: a model of exposure therapy? *Neurosci. Lett.* **548**, 269–274 (2013).
33. Besnard, A. et al. Dorsolateral septum somatostatin interneurons gate mobility to calibrate context-specific behavioral fear responses. *Nat. Neurosci.* **22**, 436–446 (2019).
34. Lebow, M. A. & Chen, A. Overshadowed by the amygdala: the bed nucleus of the stria terminalis emerges as key to psychiatric disorders. *Mol. Psychiatry* **21**, 450–463 (2016).
35. Kim, S. Y. et al. Diverging neural pathways assemble a behavioural state from separable features in anxiety. *Nature* **496**, 219–223 (2013).
36. Cullinan, W. E., Herman, J. P. & Watson, S. J. Ventral subicular interaction with the hypothalamic paraventricular nucleus: evidence for a relay in the bed nucleus of the stria terminalis. *J. Comp. Neurol.* **332**, 1–20 (1993).
37. Hallock, H. L. et al. Molecularly defined hippocampal inputs regulate population dynamics in the prelimbic cortex to suppress context fear memory retrieval. *Biol. Psychiatry* https://doi.org/10.1016/j.biopsych.2020.04.014 (2020).
38. Hsu, D. T., Kirouac, G. J., Zubieta, J. K. & Bhatnagar, S. Contributions of the paraventricular thalamic nucleus in the regulation of stress, motivation and mood. *Front. Behav. Neurosci.* **8**, 73 (2014).
39. Ruano, D. et al. Association of the gene encoding neurogranin with schizophrenia in males. *J. Psychiatr. Res.* **42**, 125–133 (2008).
40. Huang, K. P. et al. Neurogranin/RC3 enhances long-term potentiation and learning by promoting calcium-mediated signaling. *J. Neurosci.* **24**, 10660–10669 (2004).
41. Jones, K. J. et al. Rapid, experience-dependent translation of neurogranin enables memory encoding. *Proc. Natl Acad. Sci. USA* **115**, E5805–E5814 (2018).
42. Miyakawa, T. et al. Neurogranin null mutant mice display performance deficits on spatial learning tasks with anxiety-related components. *Hippocampus* **11**, 763–775 (2001).

Methods

Mice. All procedures were conducted in accordance with the U.S. National Institutes of Health Guide for the Care and Use of Laboratory Animals and the institutional Animal Care and Use Committees at University of California, San Francisco (UCSF). For rabies tracing and MAPseq experiments, male and female adult C57BL/6J mice were supplied by the Jackson Laboratory and were used from 8–12 weeks of age. Rosa26$^{\&TRAP}$ mice (JAX: 022367) were supplied by the Jackson Laboratory and bred in house, and genotyping was performed as recommended by the Jackson Laboratory via standard PCR of tail-derived genomic DNA. Mice were co-housed with littermates (2–5 per cage) in a temperature (22–24 °C) and humidity (40–60%) controlled environment. Mice were maintained with unrestricted access to food and water on a 12-h light/dark cycle, with tissue processed during the light phase. All mice were randomly assigned to experimental conditions, with approximately equal numbers of male and female mice.

General stereotaxic surgical procedures. Animals were 8–12 weeks of age at time of viral injection surgery. Mice were anesthetized with 1.5% isoflurane with an oxygen flow rate of ~11 min⁻¹, and head-fixed in a stereotactic frame (David Kopf). Eyes were lubricated with an ophthalmic ointment, and body temperature was maintained at 34–37 °C with a warm water recirculator (Stryker). Fur was shaved and the incision site was sterilized with isopropyl alcohol ×3 and betadine solution ×3 before beginning surgical procedures. Lidocaine HCl 2% solution was injected subcutaneously, local to the incision, and postsurgical analgesia was provided by meloxicam and slow-release buprenorphine. A craniotomy was made at the injection site with a round 0.5-mm drill bit (David Kopf). A Nanoject II syringe (Drummond Scientific) was used with a pulled glass pipette (tip width of 20–30 μm) to inject viruses as described below.

MAPseq. *Sample generation.* Experiments were conducted as previously described[20,21]. Adult male and female mice ($n = 12$) were anesthetized and injected with MAPseq Sindbis viral barcode library (3×10^{10} genome copies (GC)/ml, diversity of 2×10^7 different barcode sequences; Cold Spring Harbor Laboratories) into the vCA1 region of the vHPC at −3.20 mm ML, −3.16 mm AP and −3.70 and −3.80 mm DV from the brain surface (with 96.6 nl per DV site) for a total of 193.2 nl virus.

The titer and volume of Sindbis virus used was similar to that previously published, in which single-cell analysis indicated that, on average, each neuron expressed one barcode, with a small fraction expressing more than one[21]. However, as discussed at length previously[20,21,43], if neurons express more than one barcode, this would not change the distribution of projection patterns of assayed neurons. While the total number of traced neurons would be overestimated, the relative abundance of each projection motif type and bulk connection strength would not be changed. In addition, using our known barcode diversity (2×10^7) and a range of estimated values for infected cells from a lower bound of 4,000 (see "Overrepresented and underrepresented motifs") to an upper bound of 200,000, which assumes that 50% of all pyramidal cells in the entire hemispheres were labeled[44] (which is an overestimation, as our injections were localized to a small portion of the vCA1 region), we predicted that >99% of cells would be labeled uniquely if the fraction of uniquely labeled cells $F = (1 - (1/N))^{k-1}$ where N is the barcode diversity and k is the number of infected neurons[21].

Approximately 44–46 h after surgery, mice were rapidly decapitated, and brains were flash frozen in a slurry of 2-methylbutane (Fisher Scientific) on dry ice and stored at −80 °C until processing. Brains were then embedded in O.C.T. compound (Fisher Scientific) and sectioned coronally on an HM525 cryostat (Fisher Scientific) with 100–200-μm thickness. Target areas (PFC, NAc, LS, BNST, LH, BA and CeA), negative control area (dorsolateral striatum, DLS) and the source region of interest (vCA1) were mounted on SuperFrost Plus (Fisher Scientific) slides and kept on dry ice. Brain punches of target areas were collected (while avoiding fiber tracts) on dry ice with a chilled 500-μm puncher (Electron Microscopy Instruments), which was cleaned between each brain area with 100% ethanol, and samples were stored in 1.5-ml microcentrifuge tubes from each mouse and kept on dry ice until processing. Each sample was then homogenized in 400 μl of TRIzol (Thermo Fisher) and vortexed followed by a quick spin, and then kept on dry ice before shipping to MAPseq core facility for barcode extraction and sequencing (Cold Spring Harbor Laboratories)[20,21].

Barcode extraction and sequencing was completed as previously described[20]. In brief, total RNA was extracted from each sample using TRIzol (Thermo Fisher). Sample RNA and spike-in RNA (obtained by in vitro transcription of a double-stranded ultramer with sequence 5′-GTCATGATCATAATACGACTC ACTATAGGGGACGAGCTGTACAAGTAAACGCGTAATGATACGGCGACC ACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNN NNNNNNNNNNNNNNNNNNNATCAGTCATCGGAGCGGCCGCTACCTA ATTGCCGTCGTGAGGTACGACCACCGCTAGCTGTACA-3′ (IDT)30) were mixed and reverse transcribed with SuperscriptIV reverse transcriptase (Thermo Fisher; following the manufacturer's instructions) using the gene-specific primer 5′-CTTGGCACCCGAGAATTCCANNNNNNNNNNNNXXXXXXXXXTGT ACAGCTAGCGGTGGTCG-3′, where X is one of >300 true-seq-like sample-specific identifiers and N12 is the unique molecular identifier (UMI). All first-strand cDNAs were pooled and purified using SPRI beads (Beckman

Coulter) to produce double-stranded cDNA. Samples were treated with ExonucleaseI (NEB) and two rounds of nested PCR were performed using primers 5′-CTGTACAAGTAAACGCGTAATG-3′ and 5′-CAAGCAGAAGACGGCATAC GAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3′ for the first PCR and primers 5′-AATGATACGGCGACCACCGA-3′ and 5′-CAAGCAGAAGACGGCATACGA-3′ for the second PCR with Accuprime Pfx polymerase (Thermo Fisher). Finally, PCR amplicons were gel extracted using Qiagen MinElute Gel extraction kit and library sequenced on a Illumina NextSeq 500, set at high-output run and paired-end 36, using the SBS3T sequencing primer for paired-end 1 and the Illumina small RNA-seq primer 2 for paired-end 2.

All preprocessing of sequencing data was performed at the MAPseq Core Facility at Cold Spring Harbor Laboratories exactly as described previously (see Supplementary Note 4 in ref. [21]). All sequencing was performed blinded to sample identity. Briefly, Illumina sequencing results (.fastq files) were merged into one file that contained paired-end 1 (barcode sequence) and paired-end 2 (the 12-nucleotide UMI and 8-nucleotide slice-specific identifier (SSI)), so that each line contained corresponded to a single read containing the 30-nucleotide barcode, the 2-nucleotide pyrimidine anchor (YY), the 12-nucleotide UMI and the 8-nucleotide SSI. Reads were demultiplexed based on SSI using, fastx_barcode_ splitter tool (https://hannonlab.cshl.edu/fastx_toolkit/commandline.html#fastx_ barcode_splitter_usage/) and filtered to remove ambiguous bases and collapsed to unique sequences and sorted. Next, a threshold for the number of reads was selected for sequencing. As in Kebshull et al.[21], a minimum read threshold was manually selected to remove the long tail of the sequence rank profile of the Illumina results to avoid contamination with PCR and sequencing errors. The remaining reads were collapsed after removal of 12 nucleotides of the UMI to convert reads into counts. Spike-in molecules, 24-nucleotide barcodes followed by the sequence ATCAGTCA, were split from the virally expressed barcodes and processed separately. Error correction was performed using the short-read aligner bowtie, which generated all possible alignments of barcode sequences (>1 counts allowing up to 3 mismatches). All barcode sequences that mapped to each other were found, and low complexity sequences were removed by filtering barcodes with stretches of more than six identical nucleotides.

Raw barcodes reads were then normalized by relative number of spike-in RNAs and organized into an N × R matrix with N barcodes detected in R possible regions. The N barcodes is a proxy for the number of cells sending projections in the source region; in the text, we refer to each row as a cell. Each value in a row indicates the number (non-negative integer values) of detected barcodes in that region, which corresponds to the strength of the projection (density of axons) to that region. Matrices were concatenated for all mice to obtain a single larger matrix. To limit analysis only to cells that project to at least one region, we removed rows that contained all zeros. We threshold-filtered data based on presence of tenfold enrichment of source (vCA1) barcode reads compared to at least one target area and removal of any barcode with a read count in the negative target area (DLS) before further analyses. Lastly, to compare projection patterns between neurons on the same scale, we normalized each row in the N × R matrix to the maximum value for that row, such that raw barcode counts were rescaled between zero and one.

MAPseq data analysis. *Conditional probability.* To calculate the conditional probability $P(B|A)$ for a pair of regions A and B, we first found the number of cells that project to region A, including cells that project to A alone or to A with projections to other regions as well), denoted N_A. We then found the subset of cells in this group that project to region B, inclusive of cells that project to A and B, and may or may not have projections to additional regions), denoted $N_{B|A}$. The conditional probability is the proportion of cells projecting to B within the subset of cells that project to A, that is $P(B|A) = \frac{N_{B|A}}{N_A}$

Overrepresented and underrepresented motifs. To quantify the significantly overrepresented and underrepresented projection motifs in our data, we developed a null model with which to compare motifs. First, we note that our data is a matrix of neurons by target regions, and when analyzing projection motifs, we first binarized this matrix to only consider whether a neuron projects to a particular target, regardless of its projection strength. This binarized matrix is then interpreted as a bipartite graph where the neurons form one node set and the target regions form the other node set, and values of 1 in the matrix indicate an edge between nodes[45].

We then constructed our null model as an Erdős–Rényi random (bipartite) graph where edge formation is determined by a constant probability following a binomial distribution[43,46]. This results in a null model where neurons are assumed to not have any intrinsic preference for projecting to particular target regions and where the probability of projecting to one region is conditionally independent from that of another region.

The probability of edge formation, denoted p_e, determines the edge density of the null model, which we assumed to be equal to our empirical data. In generating a random graph using p_e, some simulated neurons will have zero projections to any of the target regions; however, due to the nature of MAPseq, the empirical data only includes neurons that have at least one projection to the target regions sampled. To properly model our data, we need to know the total number of neurons, including those that had no projections to one of our target regions.

We can estimate this by noting that $N_f = N_0 \, p_e$, where N_f denotes the number of neurons in our empirical data and N_0 refers to the total number of neurons, such that $N_0 - N_f$ would be the number of neurons with zero projections[20].

First, we infer N_0 from the empirical data by assuming a binomial model and recalling that the probability of at least one projection is one minus the probability of no projections:

$$N_f \approx N_0 \cdot \left(1 - \prod_i^R (1 - p_i)\right)$$

Where p_i refers the probability of a neuron projecting to region i among the total R regions. We define $p_i = \frac{s_i}{N_0}$, where s_i is the number of neurons in the empirical data that project to region i. By substitution we get:

$$N_f \approx N_0 \cdot \left(1 - \prod_i^R \left(1 - \frac{s_i}{N_0}\right)\right)$$

We now have a polynomial with one unknown, N_0, and we can solve this using any root-solving algorithm. By knowing N_0 and N_f, we can then solve for p_e:

$$N_f \approx N_0 \cdot (1 - (1 - p_e))^R$$

This is because the number of neurons with at least one projection using the null model with p_e must reproduce the number of neurons in the empirical data, which are also filtered to only include those with at least one projection. This equation is also a polynomial with one unknown, p_e, and again we can solve it numerically using a root-solving algorithm. In our data, $N_f = 2,494$, and we computed $N_0 \approx 4,211$, $p_e \approx 0.120$. Using p_e and N_0 we can compute the expected counts for each motif, for example, the expected counts for each single-projector motif under the null model is computed as $N_0 \, p_e \, (1 - p_e)^6$, since there are seven target regions. We then compute the two-sided P values for our observed counts under the null binomial model using the binomial test.

k-means clustering. We performed k-means clustering on the normalized $N \times R$ data matrix. We chose the k parameter (number of clusters) using the 'elbow method'[47], in which we plot the cluster inertia versus increasing trials of k from 1 to 15, and choose the k value at which the inertia first begins to plateau. The inertia is the within-cluster sum of squared distances between each point in the cluster and its centroid. For the full set of seven regions, we found $k = 6$ to be near optimal. Each centroid's coordinates denote the representative projection strengths (of the cluster) for the corresponding region.

We plotted the centroids as a dot plot (that is, grid-structured scatter plot) where each row is a centroid (cluster) and the color of the dots along each column corresponds to the coordinate value of the centroid. Each coordinate value corresponds to a projection strength to a region, for example, a bright yellow dot in row 1 and column 2 represents a strong projection to that region for cluster 1. The k-means results demonstrate that the cells are well separated in the original high-dimensional space by the region to which they maximally project. All MAPseq data were analyzed using custom scripts written in Python.

Input–output rabies tracing. *Surgeries.* Male and female mice were unilaterally injected with AAV2retro-CAG-Cre (2.8×10^{12} vg ml^{-1}, UNC Vector Core) with 1:50 Fluoromax beads (Thermo Fisher) targeting one of the six downstream projections from vHPC: mPFC, NAc, LH, adBNST, LS and BA. If the blue beads were not found in the target area, the animal was excluded from analysis. For the mPFC, 96.6 nl was injected at 1.7 mm AP, −0.3 mm ML and −2.75 mm DV, and another 64.4 nl was injected at −2.5 mm DV; NAc condition: 161 nl was injected at 1.2 mm AP, −1.1 mm ML, −4.6 mm DV; LH condition: 64.4 nl was injected at each DV coordinate, −2.0 mm AP, −0.75 mm ML, −5.25 mm DV, −5.0 mm DV and −4.75 mm DV; BA condition: 96.6 nl was injected at −1.7 mm and −2.0 mm AP, −3.0 mm ML, −5.5 mm DV then another 64.4 nl was injected at −5.25 mm DV; adBNST condition: 128.8 nl was injected at 0.2 mm AP, −1 mm ML, −4.3 mm DV, −4.15 mm DV, −4.0 mm DV; and LS condition; 128.8 nl was injected at 0.5 mm AP, −0.35 mm ML, −3.24 mm DV, −3 mm DV and −2.75 mm DV. Followed by an injection of AAV1-synP-FLEX-TVA-P2A-EGFP-P2A-B19G (3.9×10^{12} vg ml^{-1}; UNC Vector Core) targeting the vCA1 region of vHPC with 64.4 nl at −3.25 mm ML, −3.14 mm AP, −3.85 mm DV, 64.4 nl at −3.55 mm DV and, finally, 32.2 nl at −3.55 mm DV. Two to three weeks later, mice were injected with 64.4 nl of RV-EnvA-ΔG-mCherry (3.79×10^8 vg ml^{-1}; Salk) into vCA1 at 3.25 mm ML, −3.14 mm AP and −3.55 mm DV.

Tissue processing. Five days after RV-EnvA-ΔG-mCherry injection mice were deeply anesthetized with ketamine (80 mg kg^{-1}) intraperitoneally and transcardially perfused with 0.01 M PBS followed by 4% paraformaldehyde (PFA). Brains were extracted and post-fixed in 4% PFA overnight, followed by cryoprotectant with 30% sucrose for at least 24 h. Brains were then serially sectioned at 40-μm coronal sections with a Leica SM2000R freezing microtome then consecutive sections were mounted on SuperFrost Plus slides (Fisher Scientific) and coverslipped with DAPI-Fluoromount-G Clear mounting medium (Fisher Scientific).

Imaging. Whole-brain images were acquired from every section at ×10 magnification with an Aperio Versa slide scanner (Leica Biosystems). Individual sections on each slide were isolated and then registered to the Allen Brain Atlas using NeuroInfo software with Brainmaker module (MBF Bioscience), and all RV$^+$ neurons outside the hippocampus (with the exception of contralateral CA3) were mapped and counted. In vHPC sections, cells coexpressing mCherry and EGFP (starter cells) were imaged with a CSU-W1 spinning disk widefield confocal microscope (Nikon Imaging Center) at ×40 magnification and then registered to the Allen Brain Atlas using NeuroInfo (MBF Bioscience) and counted.

For each brain, the number of input neurons in a specific brain region was normalized to the total number of input neurons (RV$^+$) counted across the brain outside the hippocampus. All data collection and counts were made blinded to the experimental group. The mean and s.e.m. are listed in Supplementary Table 3 for each brain region. For the analysis, one-way ANOVAs were run and those regions with $P < 0.05$ were run with multiple t-tests with Holm–Sidak correction for multiple comparisons. All statistical results for significant effects are available in Supplementary Table 4. Data was analyzed using Prism 8 (GraphPad).

Projection TRAP-RNA-seq profiling. *Brain tissue generation.* Adult FsTRAP heterozygous (B6.129S4-Gt(ROSA)26Sortm1(CAG-EGFP/Rpl10a-birA) Wtp/J, Rosa26^{fsTRAP}; stock no. 022367, Jackson Laboratory) mice were injected with AAV2retro-CAG-Cre (2.8×10^{12} vg ml^{-1}, UNC Vector core) with the same coordinates and parameters in mPFC, NAc, LH and BA as described in Input–output rabies tracing. Three to four weeks later, mice were rapidly decapitated and bilateral ventral hippocampi were microdissected on ice (by peeling away the hippocampus from the cortex via the hippocampal fissure, then removing the dentate gyrus and CA3 and then dissecting vCA1 for analysis), and placed in ice-cold buffer B (1× HBSS, 4 mM NaHCO$_3$, 2.5 mM HEPES and 35 mM glucose) with 100 mg ml^{-1} cycloheximide (Sigma-Aldrich). Ventral hippocampi from male and female mice were pooled for each projection target (mPFC, NAc, LH or BA) in duplicate or triplicate. Processing of the samples was completed as described in Ekstrand et al. [28]. Briefly, samples were homogenized via glass homogenizer (Kimble Kots 20) in 1.5 ml of ice-cold buffer C (10 mM HEPES, 150 mM KCl and 5 mM MgCl$_2$) with 0.5 mM dithiothreitol (DTT; Sigma), 80 U ml^{-1} RNasin Plus (Promega), 40 U ml^{-1} Superase-In (Life Technologies), 100 mg ml^{-1} cycloheximide, protease inhibitor cocktail (Roche) and 100 ng ml^{-1} GFP-Trap Protein (ChromoTek), followed by homogenization with a variable-speed homogenizer (Glas-Col) at 4 °C, three times at 250 r.p.m. and ten times at 750 r.p.m. Samples were then transferred into new centrifuge tubes and centrifuged at 2,000g for 10 min at 4 °C, followed by addition of 140 μl of 10% IGEPAL CA-630 (NP-40; Sigma-Aldrich) and DHPC (100 mg/0.69 ml; Avanti Polar Lipids). Solutions were then mixed before centrifugation at 17,000g for 15 min at 4 °C, and supernatant solution was transferred to new microcentrifuge tubes. Beads and GFP antibody were prepared by washing 300 μl of Streptavidin MyOne T1 Dynabeads (Invitrogen) twice with PBS on a magnetic rack and loaded with biotinylated Protein L (Thermo Scientific) for 35 min at room temperature. Bead mixture was then blocked with five washes of 3% IgG-free and protease-free BSA (Jackson ImmunoResearch) in PBS and resuspended in buffer A (10 mM HEPES, 150 mM KCl, 5 mM MgCl$_2$ and 1% NP-40) with 100 mg monoclonal anti-GFP antibody[48]) at room temperature. Finally, before use, anti-GFP$^+$ bead mixture was washed twice in buffer A with 0.5 mM DTT, 80 U ml^{-1} RNasin Plus and 100 mg ml^{-1} cycloheximide, and then samples were added and incubated for 40 min at 4 °C. The sample mixture was washed three times with buffer D (10 mM HEPES, 350 mM KCl, 5 mM MgCl$_2$ and 1% NP-40), 0.5 mM DTT, 80 U ml^{-1} RNasin Plus and 100 mg ml^{-1} cycloheximide and then moved to a new tube before a fourth wash was performed. Next, 100 μl of lysis buffer was added to each sample and purified with RNeasy Micro kit (Qiagen) and stored at −80 °C before being sent for sequencing at the Functional Genomics Core at UCSF.

RNA-seq and analysis. RNA-seq library was generated from SMART-seq.v4 ultra low input RNA kit (Takara Bio) and Nextera XT DNA Library preparation kit with multiplexing primers (Illumina) according to the manufacturers' protocols. All sequencing was completed blinded to sample identity. Fragment size distribution and concentration were measured with 5200 Fragment Analyzer System and DNA high-sensitivity chip (Agilent). Libraries were multiplexed at six per flow-cell lane and sequenced on a HiSeq 4000 to generate single-end 50-bp reads following the manufacturer's protocol (Illumina). Unique transcript reads were aligned to the mouse genome (vGRCm38.78) with STAR v.2.5.2b[49] using quantMode GeneCounts and soft clipping the 3′ adaptor: GATCGGAAGAGCACACGTCTGAACTCCAGTCAC, with the following added settings: outFilterMismatchNoverLmax = 0.04, outFilterMismatchNmax = 999, alignSJDBoverhangMin = 1, outFilterMultimapNmax = 1, alignIntronMin = 20, alignIntronMax = 1000000 and alignMatesGapMax = 1000000. Aligned reads were counted using HTseq with default parameters[50,51].

Count normalization and differential gene expression analysis were performed using DESeq2 (ref. [52]) with default parameters as described in the vignette (https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html/). All raw counts were used as input; no additional filtering was applied. See Supplementary Table 4 for all genes, normalized counts, fold

changes and raw and corrected P values. Genes were deemed significant when the adjusted P value < 0.05. Volcano plots for the different comparisons were made using EnhancedVolcano package (R package version 1.4.0; https://github.com/kevinblighe/EnhancedVolcano/). The ComplexHeatmap package[49] was used to depict differentially expressed genes from mPFC versus subcortical regions (normalized counts from DESeq2 were scaled to −2 to +2). Gene ontology term enrichment analysis was run on genes that had higher expression levels in the mPFC versus subcortical regions using Metascape[52]. In addition, all differentially expressed genes between mPFC versus subcortical regions were run through IPA (Qiagen) to define activated/inhibited canonical pathways.

Clustering plots in Extended Data Fig. 3 were plotted using the following R packages: prcomp v3.6.3 for calculating PCA, ggplot2 v3.3.0 for plotting and hierarchical clustering, and pheatmap v1.0.12 for plotting and clustering.

RNAscope. Reagents were purchased from ACDBio. CDR1 probe was warmed for 10 min at 40 °C, and then cooled to room temperature. RNAScope fluorescent multiplex reagents (Amp 1–4-FL) were equilibrated to room temperature. Fixed tissues were transferred briefly to RNase-free dH20 and mounted onto slides. Hydrophobic barriers were drawn around each section. Slides were left to dry in the dark at room temperature for at least 30 min. Slides were then dipped rapidly in 100% EtOH and left to air dry for several minutes. Slides were then transferred to a humidity tray. Two to four drops of protease IV were added and sections were left to incubate for 30 min at room temperature. Afterwards, slides were dipped rapidly in a Coplin jar with RNase-free dH2O. Four drops of CDR1 probe were added to cover each section, and then slides were incubated for 2 h at 40 °C in a humidified tray. Slides were washed in 1× wash buffer twice for 2 min at room temperature. Four drops of Amp 1-FL were added to cover each section, and the slides were then incubated for 30 min at 40 °C in a humidified tray. Slides were washed in 1× wash buffer twice for 2 min at room temperature. This process was repeated for Amp 2–4 FL, except incubation periods were 15 min, 30 min and 15 min, respectively. The color module chosen for Amp 4-FL was Alt A-F(C2 in far red). After final washing in 1× wash buffer, slides were coverslipped with ProLong Gold Antifade Mountant. Slides were imaged using a Nikon Ti confocal, and counts were taken at ×20 in a 1.3 mm² field of view (nine FOVs from two mice).

Statistical analysis. No statistical methods were used to predetermine sample size a priori, but the sample sizes used were similar to those reported in previous studies[21,25,28]. Data distribution was assumed to be normal, but this was not formally tested. For identification of overrepresented and underrepresented motifs from MAPseq data, we compared the observed counts to those expected from the null model using a two-sided binomial test with P values corrected for multiple comparisons using the Bonferroni method. P values for all motifs are provided in Supplementary Table 1. For input–output tracing, data were analyzed using one-way ANOVAs, followed by post hoc comparisons, with Holm–Sidak correction for multiple comparisons. All tests were two-sided. In the dot plots, data are presented as the mean ± s.e.m. All statistical tests and P values for rabies input–output tracing are provided in Supplementary Table 2. For RNA-seq data, differential gene expression analysis was performed using DESeq2 (ref. [52]) using parameters as described in the vignette (https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html/). Briefly, P values were acquired by two-sided Wald test, corrected for multiple comparisons with the Benjamini–Hochberg method. Prism 8 (GraphPad), MATLAB (MathWorks) and Python (SciPy) were used for data analyses.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
RNA-seq data are available in the NCBI's Gene Expression Omnibus under accession number GSE150869, and MAPseq data are available at NLM Sequence Read Archive BioProject under accession number PRJNA633836. RNA-seq differential expression data is provided in Supplementary Table 4. All MAPseq source data can be downloaded at https://github.com/mkheirbek/. MAPseq motif

counts are provided in Supplementary Table 1. All counts for input–output rabies tracing (n, mean and s.e.m.) are provided in Supplementary Table 2.

Code availability
Code for analysis is posted on the Kheirbek Lab GitHub site (https://github.com/mkheirbek/).

References
43. Huang, L. et al. BRICseq bridges brain-wide interregional connectivity to neural activity and gene expression in single animals. *Cell* **182**, 177–188.e27 (2020).
44. Abusaad, I. et al. Stereological estimation of the total number of neurons in the murine hippocampus using the optical disector. *J. Comp. Neurol.* **408**, 560–566 (1999).
45. Pavlopoulos, G. A. et al. Bipartite graphs in systems biology and medicine: a survey of methods and applications. *Gigascience* **7**, 1–31 (2018).
46. Miele, V., Matias, C., Robin, S. & Dray, S. Nine quick tips for analyzing network data. *PLoS Comput. Biol.* **15**, e1007434 (2019).
47. Thorndike, R. L. Who belongs in the family? *Psychometrika* **18**, 267–276 (1953).
48. Doyle, J. P. et al. Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).
49. Gu, Z., Eils, R. & Schlesner, M. Complex heat maps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
50. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
51. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
52. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

Author contributions
Conceptualization: M.A.K.; methodology: M.A.K.; investigation: M.M.G., K.J.H., K.J.C., H.S.C. and V.S.T.; formal analysis: M.A.K., M.M.G., K.J.H., H.S.C., B.B., I.D.V. and A.V.M.; writing: M.A.K. (original draft) and all authors (review, editing and methods); resources: M.A.K.; and supervision: M.A.K.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41593-020-0705-8.

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-020-0705-8.

Correspondence and requests for materials should be addressed to M.A.K.

Peer review information *Nature Neuroscience* thanks Lucas Pozzo-Miller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at www.nature.com/reprints.
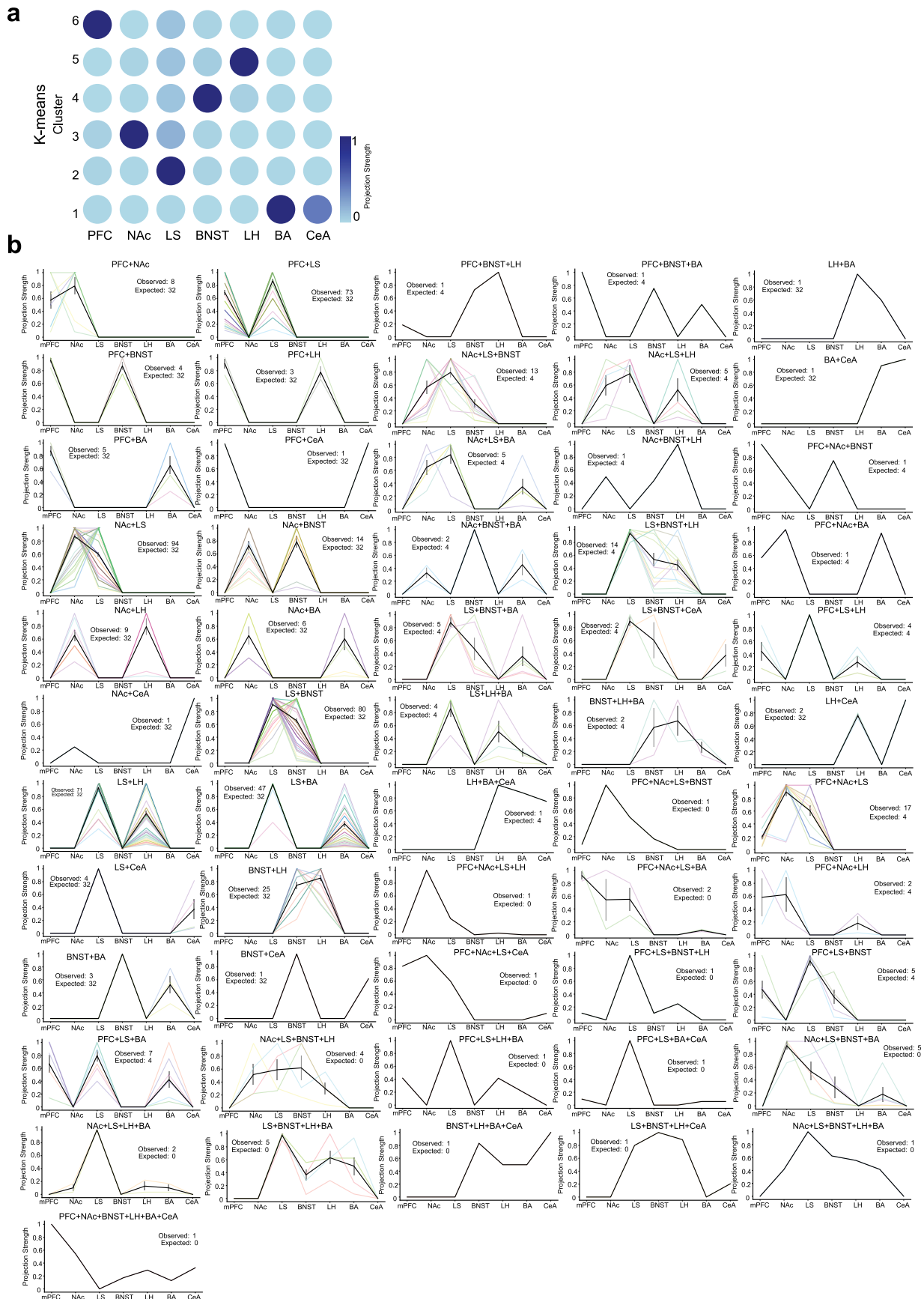
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Clustering MAPseq data and projection strengths of individual vCA1 neurons. a. k-means clustering of MAPseq data. Clusters are plotted as a grid-structured scatter plot where each row is a cluster and the color of the dots along each column correspond to the normalized projection strength to a region **b**. Plots of normalized projection strength of all multiple projection barcode motifs from MAPseq data. As in Fig. 2 each line (different color) is an individual neuron's projection strength to each of the 7 target regions, normalized to the maximal value in that row, resulting in a projection strength scale from 0 to 1 (see Methods). Black line is mean projection strength for all neurons in that motif, and grey is SEM. Exact cell number for each motif is provided in the inset (observed and expected from null model, see Methods), and in Supplementary Data Table 1.
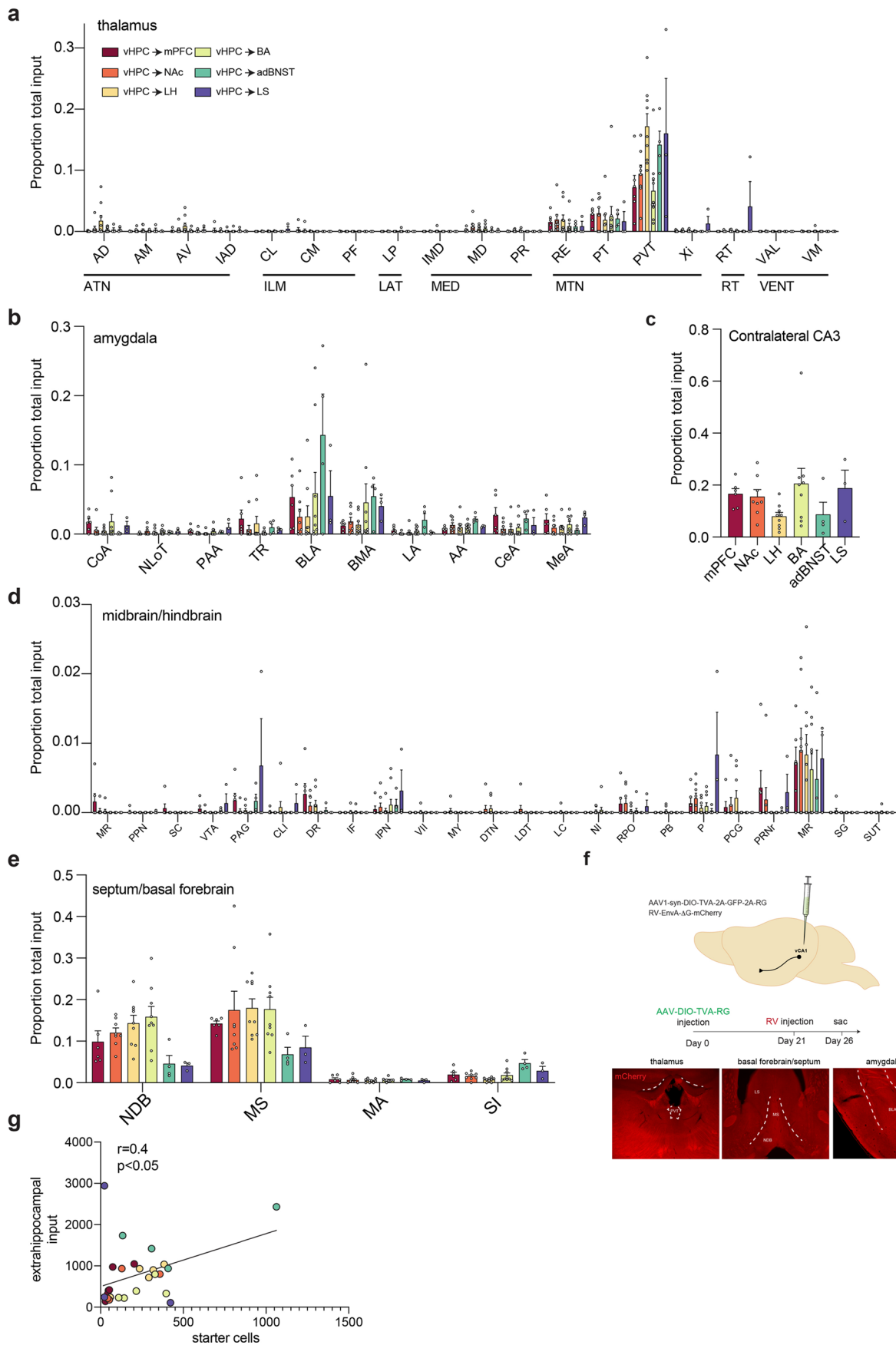
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Brainwide extra-hippocampal input to vCA1 neurons that project to LH, BA, NAc, adBNST, LS and mPFC. a–d. Fraction of extrahippocampal input from nuclei in the (a) thalamus, (b) amygdala, (c) contralateral CA3, (d) midbrain/hindbrain and (e) basal forebrain/septum. See Methods for all abbreviations. n= 6 vCA1-mPFC, 8 vCA1-NAc, 9 vCA1-LH, 9 vCA1-BA, 4 vCA1-adBNST and 3 vCA1-LS mice . Error bars represent standard error of the mean. See Extended data table 2 for proportion counts (mean $\pm$ SEM) for all assayed input regions and all statistical values. **f**. Controls for non-specific rabies infection. No long-range inputs were labeled in mice injected with Cre-dependent rabies helper virus and EnvA-G deleted rabies mCherry virus in the vCA1 subregion of vHPC in the absence of AAV2retro Cre injection (n=2 mice, no long-range input cells detected). *right*, image of red channel from thalamus, basal forebrain and amygdala of control mice. **g**. Number of extrahippocampal inputs correlates with the number of starter cells in RV samples (Pearson's r=0.4035, two-tailed, p=0.037). AA-anterior amygdala, AD-anterodorsal nucleus, AM-anteromedial nucleus, BLA-basolateral amygdala, BMA-basomedial amygdala, CeA-central amygdala, CL-central lateral nucleus of the thalamus, CLI-central linear nucleus raphe, CM-central medial nucleus of the thalamus, CoA-cortical amygdala, DR-dorsal raphe nucleus, DTN-dorsal tegmental nucleus, IAD-interanterodorsal nucleus, IF-interfascicular nucleus raphe, IMD-Intermediodorsal nucleus of the thalamus, IPN-interpeduncular nucleus, LA-lateral amygdala, LC-locus coeruleus, LDT-laterodorsal tegmental nucleus, LHA-lateral hypothalamic area, LP-lateral posterior nucleus of the thalamus, LS-lateral septum, MR-median raphe, MA-magnocellular nucleus, MD-intermediodorsal nucleus of the thalamus, mPOA-medial preoptic area, MS-medial septum, MY-medulla, NDB-diagonal band nucleus, NI-nucleus incertus, NLoT-nucleus of lateral olfactory tract, P-pons, PAA-piriform-amygdalar area, PAG-periaqueductal gray, PB-parabrachial nucleus, PCG-pontine central gray, PF-parafascicular nucleus, PPN-pedunculopontine nucleus, PR-perireunensis nucleus, PRNr-pontine reticular nucleus, PT-parataenial nucleus, PVH-paraventricular nucleus of hypothalamus, PVT, paraventricular nucleus of the thalamus, RE-nucleus of reuniens, RPO-nucleus raphe pontis, RT- reticular nucleus of the thalamus, SC- superior colliculus, SG-supragenual nucleus, SI- substantia innominate, SUT-supratrigeminal nucleus, TR-piriform transition area, VAL-ventral anterior-lateral complex of the thalamus, VII-facial motor nucleus, VM-ventral medial nucleus of the thalamus, VTA-ventral tegmental, Xi-xiphoid thalamic nucleus.
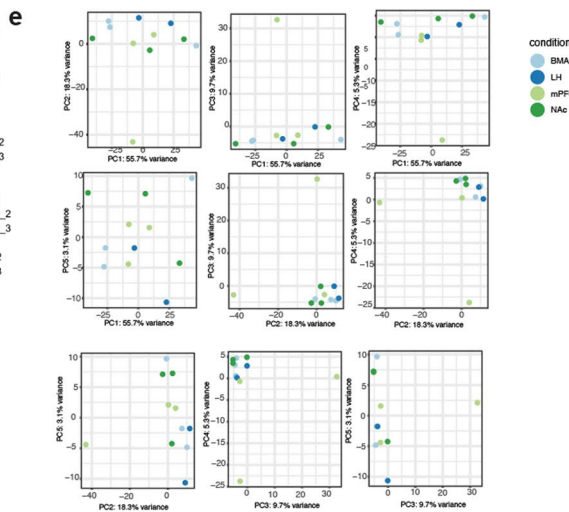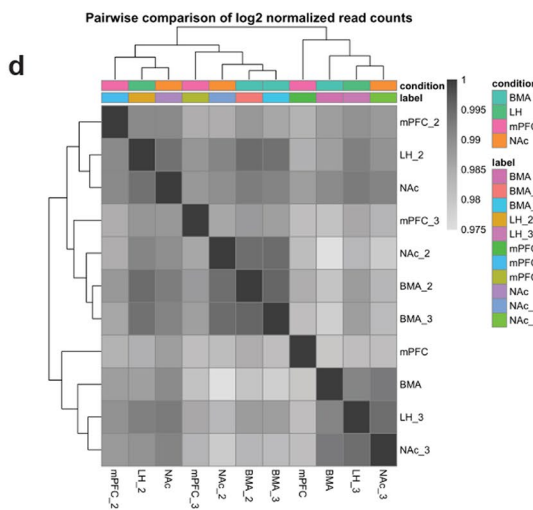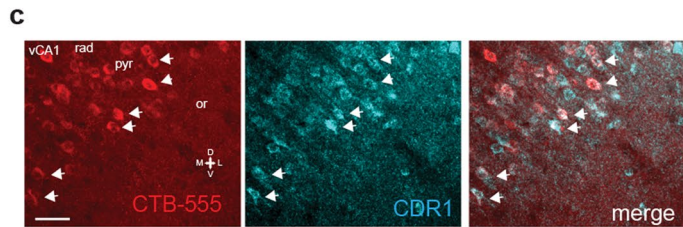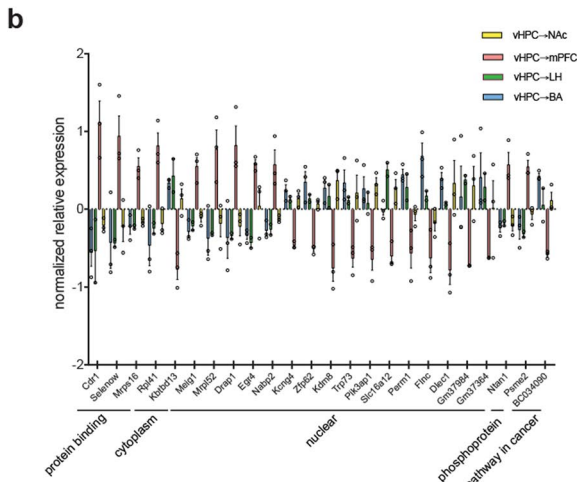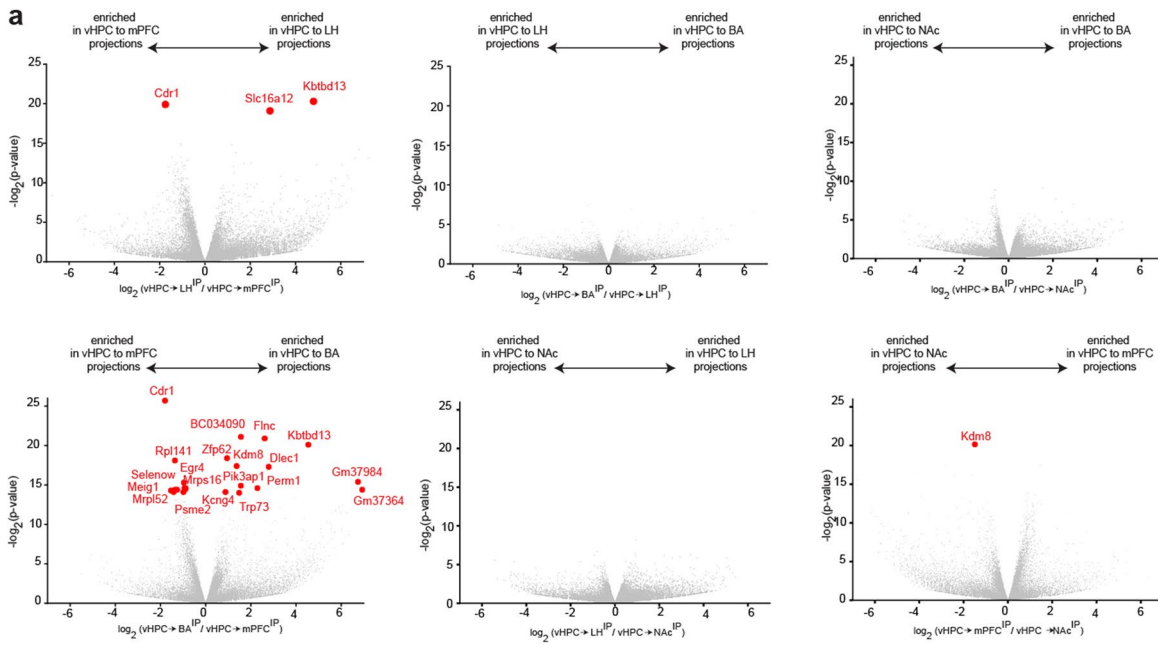
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Transcriptional profiling of vCA1 projection neurons. a. Volcano plots of differentially expressed genes for each set of pairwise comparisons. Red dots indicate differentially expressed genes that passed p<0.05 cutoff using Wald test, two-sided, and after correction for multiple comparisons with Benjamini Hochberg. All exact p-values provided in Extended Data Table 4. **b**. Normalized relative expression of the all 24 genes that passed the significance threshold for differential expression in pairwise comparisons . n=3 vCA1 to mPFC replicates, 2 vCA1 to LH replicates, 3 vCA1 to NAc replicates and 3 vCA1 to BA replicates. Error bars indicate SEM. **c**. RNAscope of CDR1 transcript in vCA1 neurons defined by their projection to the mPFC. Arrowheads indicate co-labeled neurons. Scale bar 40 μm. *Right*. Quantification of overlap in CTB labeled and CDR1 labeled cells, n=9 FOVs from 2 mice, error bars indicate SEM. **d**. Heatmap correlations of total log2 normalized read counts from individual sample sets used for profiling experiments. **e**. Principal components plots of 500 most variable genes in the dataset, showing first 5 PCs.

nature research

Corresponding author(s): Kheirbek

Last updated by author(s): Aug 4, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Aperio ImageScope (Leica v12.4.0.5043) NeuroInfo Software with Brainmaker module (MBF Biosciences v2020.1.1) DESeq2 (v1.29.8) STAR (v2.5.2b) HTSeq (v0.12.4) EnhancedVolcano package (v1.4.0) Metascape (v3.5) Ingenuity Pathway Analysis (Qiagen v. June 2020)
Data analysis	Graphpad Prism (v8.0) MathWorks Matlab (vR2018a) Python (v3.6)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the raw data for this manuscript has either been deposited online (RNA sequencing data is available in the NCBI's Gene Expression Omnibus, accession # GSE150869, and MAPseq data is available at NLM Sequence Read Archive BioProject: PRJNA633836.) Processed data and all code for MAPseq analysis at

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Statistical methods to predetermine sample size were not used. Sample sizes were based on previous studies using the techniques described (Ekstrand et al , Cell, 157, 1230-1242, 2014; Han et al, Nature, 556, 51-56, 2018; Swartz et al, Nature, 524, 88-92, 2015)
Data exclusions	All exclusion criteria were established before data collection. For rabies tracing, brains were excluded if location of AAV2retro injection was outside of the target area, or if helper virus injection was outside of vHPC in overlying cortex. For MapSeq data, we threshold-filtered data based on 10-fold enrichment of source (vCA1) barcode reads compared to at least 1 target area. Any barcodes that did not have a 10-fold enrichment in at least 1 target area were not analyzed. In addition, any barcodes with a read count in a negative target area (dorsolateral striatum) were removed (see Han et al, Nature, 556, 51-56)
Replication	The experiments were done with multiple animals to ensure reproducibility. The number of replications is presented in the figures. All protocols, replications and numbers are provided in the methods, and all raw data is provided in the supplemental tables.
Randomization	Mice were randomly assigned to projection groups, across litters. For MapSeq group allocation does not apply.
Blinding	All counts for rabies experiments were done by experimenter blinded to group All MapSeq and and TRAP profiling, sequencing was performed by experimenter blinded to sample identity.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	male and female adult (>8 weeks of age) C57BL/6 mice and B6.129S4-Gt(ROSA)26Sortm1(CAG-EGFP/Rpl10a,-birA)Wtp/J, Rosa26fsTRAP, stock #022367, Jackson Laboratory. Mice were co-housed with littermates (2-5 per cage). Mice were maintained with unrestricted access to food and water on a 12-hour light/dark cycle, 22-24 deg C, 40-60% humidity, and cages were changed twice a week by veterinary technicians.
Wild animals	No wild animals were used in this study
Field-collected samples	This study did not involve field collected samples
Ethics oversight	All procedures were conducted in accordance with the U.S. NIH Guide for the Care and Use of Laboratory Animals and the institutional Animal Care and Use Committees at UCSF.

Note that full information on the approval of the study protocol must also be provided in the manuscript.