

# Persistent transcriptional programmes are associated with remote memory

<https://doi.org/10.1038/s41586-020-2905-5>

Michelle B. Chen<sup>1,5</sup>, Xian Jiang<sup>2,3,5</sup>, Stephen R. Quake<sup>1,4</sup> & Thomas C. Südhof<sup>2,3</sup>

Received: 26 September 2019

Accepted: 17 August 2020

Published online: 11 November 2020

 Check for updates

The role of gene expression during learning and in short-term memories has been studied extensively<sup>1–3</sup>, but less is known about remote memories, which can persist for a lifetime<sup>4</sup>. Here we used long-term contextual fear memory as a paradigm to probe the single-cell gene expression landscape that underlies remote memory storage in the medial prefrontal cortex. We found persistent activity-specific transcriptional alterations in diverse populations of neurons that lasted for weeks after fear learning. Out of a vast plasticity-coding space, we identified genes associated with membrane fusion that could have important roles in the maintenance of remote memory. Unexpectedly, astrocytes and microglia also acquired persistent gene expression signatures that were associated with remote memory, suggesting that they actively contribute to memory circuits. The discovery of gene expression programmes associated with remote memory engrams adds an important dimension of activity-dependent cellular states to existing brain taxonomy atlases and sheds light on the elusive mechanisms of remote memory storage.

Long-term memories do not form immediately after learning, but are consolidated over time<sup>4</sup>. Previous studies have identified important contributions of molecular and cellular processes to learning and memory, such as gene expression changes, cAMP signalling and synaptic plasticity<sup>1</sup>, and identified a central role for RNA synthesis and protein translation in memory consolidation<sup>2</sup>. Despite these discoveries, the molecular underpinnings of memory consolidation remain elusive. For instance, while changes in gene expression are found in the first 24 h of learning, it is unclear whether these changes are maintained or whether new changes are acquired to consolidate a long-term memory trace<sup>3</sup>. Moreover, the dependence on the hippocampus for long-term memory is known to degrade over time, with cortical structures such as the medial prefrontal cortex (mPFC) becoming increasingly important<sup>5</sup>. Recently, the development of activity-dependent genetic labelling tools have allowed identification of sparsely activated neuronal ensembles, enabling access to the molecular mechanisms that underlie experience-dependent connectivity and plasticity<sup>6</sup>.

## Neuron subtypes in remote memory engrams

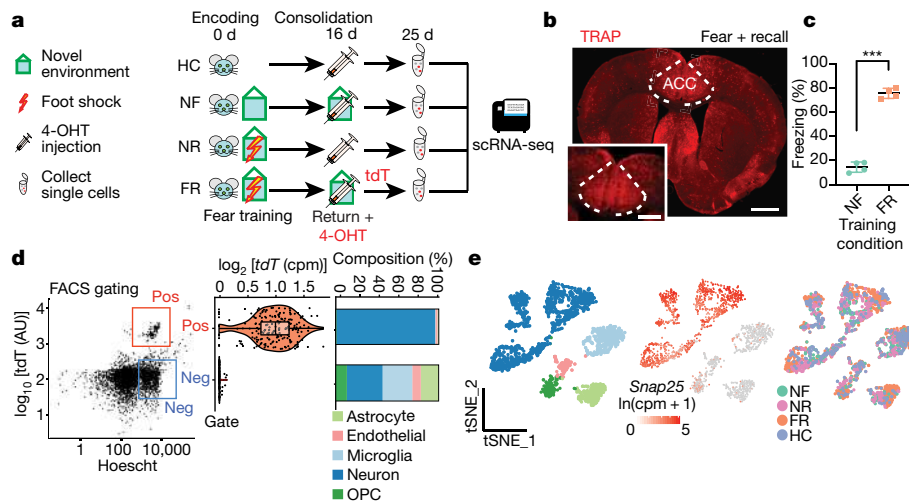
To identify and study the transcriptional programmes of neurons involved in remote memory, we used *TRAP2; Ai14* mice expressing iCre-ERT2 recombinase in an activity-dependent manner along with a *iCre*-dependent *tdTomato* (*tdT*) reporter allele (Extended Data Fig. 1a), enabling us to label memory recall-activated neurons. To genetically label the engrams that are associated with remote memory, we trained mice in a conditioning chamber with three pairs of tone–foot shocks on day 0, and induced fear memory recall (FR) on day 16 by returning mice to the conditioning chamber. Memory recall activates the consolidated memory engrams, thereby labelling the

neuronal ensemble that encodes the consolidated remote memory<sup>7</sup>, while also inducing memory reconsolidation<sup>8</sup>. Because the molecular and cellular substrates of consolidation and reconsolidation were indistinguishable in this paradigm, we denote them here collectively as memory consolidation. We used control mice that were not fear conditioned but exposed to the recall context (no fear (NF)), fear conditioned but not subjected to recall (no recall (NR)), and neither fear conditioned nor exposed to the recall condition (homeage (HC)) (Fig. 1a–c). All mice were injected with 4-hydroxytamoxifen (4-OHT) before the FR procedure (or at the equivalent time) to allow activity-dependent production of tdT, thus enabling the distinction between the molecular programmes that are specific to remote memory versus background activation.

Nine days after FR<sup>9</sup>, single neuronal and non-neuronal cells were collected from the mPFC (Extended Data Fig. 1b) via fluorescence-activated cell sorting (FACS) and gating on the tdT signal, followed by plate-based single-cell mRNA sequencing (Fig. 1d). The percentage of TRAPed cells collected via FACS was significantly higher in FR (about 1.5% of all cells) than in other conditions (Extended Data Fig. 2a), further confirming that the TRAP2 system captured increased neuronal activity during the FR process. We sequenced 3,691 neurons (*Snap25*<sup>+</sup>/*tdT*<sup>+</sup> or *Snap25*<sup>+</sup>/*tdT*<sup>−</sup> mRNA) and 2,672 non-neuronal cells with high quality and depth (Extended Data Fig. 1c, d). Unbiased transcriptome clustering of cells from all four training conditions allowed the identification of major cell types and confirmed the dominance of neurons among *tdT*<sup>+</sup> cells, whereas *tdT*<sup>−</sup> cells comprised both neurons and non-neuronal cells (*Cldn5*<sup>+</sup> endothelial, *Pdgfra*<sup>+</sup> oligodendrocyte progenitor cells (OPCs), *Cx3cr1*<sup>+</sup> microglia and *Aqp4*<sup>+</sup> astrocytes) (Fig. 1d, e, Extended Data Fig. 1e). Both *tdT*<sup>+</sup> and *tdT*<sup>−</sup> cells from FR and control groups were represented in all clusters, which suggests that neither the neuronal

<sup>1</sup>Department of Bioengineering, Stanford University, Stanford, CA, USA. <sup>2</sup>Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA. <sup>3</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. <sup>4</sup>Chan Zuckerberg Biohub, Stanford, CA, USA. <sup>5</sup>These authors contributed equally: Michelle B. Chen, Xian Jiang.

<sup>✉</sup>e-mail: [steve@quake-lab.org](mailto:steve@quake-lab.org); [tcs1@stanford.edu](mailto:tcs1@stanford.edu)



**Fig. 1 | Labelling and collection of single memory engram cells via the TRAP2;Ai14 line.** **a**, The experimental paradigm includes generating remote fear-memory traces via contextual fear conditioning, isolating TRAP<sup>+</sup>-activated neurons via flow cytometry and plate-based single-cell RNA sequencing (scRNA-seq). **b**, Representative image of tdT<sup>+</sup> TRAPed (red) cells in the anterior cingulate cortex (ACC) region of the mPFC 9 days after an injection of 4-OHT (at the time of remote memory recall). Scale bars, 1 mm (0.5 mm for zoomed in images). **c**, Degree of freezing upon return to the novel context 16 days after fear conditioning (FR) or no conditioning (NF) ( $n = 4$  mice per condition,

\*\*\* $P = 4.89 \times 10^{-9}$ ; mean  $\pm$  s.d.). **d**, Representative flow gating for tdT<sup>+</sup> TRAPed cells (about 1.5% of events in FR) in one mouse from the FR condition (left). Post-sequencing analysis of 722 example cells from a representative FR brain shows enrichment of *tdT* mRNA (pos) in the positive sort gate (orange violin) (middle). Enrichment of neuronal cells in the positive gate and neuronal and non-neuronal cell types in the negative gate in all brains (right). AU, arbitrary units; cpm, counts per million. **e**, All sorted cells, with neuronal cells identified via the expression of *Snap25* mRNA. All training and control conditions are represented in all cell clusters.

activation state nor the training paradigm significantly altered fundamental cell-type identities (Fig. 1e).

Sub-clustering of 3,691 *Snap25*<sup>+</sup> neurons using the top 2,000 highly variable genes revealed seven putative neuron sub-populations—four glutamatergic (C0, C1, C3 and C5) and three GABAergic neuron populations (C2, C4 and C6)—all of which were consistently observed throughout four biological replicates (Fig. 2a–c, Extended Data Fig. 2b). These are molecularly distinct populations (Fig. 2d), with each subtype expressing at least one distinctive marker gene (see Methods; Fig. 2e). All subtypes contained *tdT*<sup>+</sup> cells, which indicated activation of all neuron subtypes regardless of the training state (Extended Data Fig. 2c). A comparison of key layer-specific marker genes (C0-*Dkk1*, C1-*Rprm*, C2-*Calb2*, C3-*Tesc*, C4-*Tnfrsf8*, C5-*Tshz2* and C6-*Lhx6*) to existing cortical single-cell expression databases<sup>10</sup> confirmed their presence in the mPFC (Extended Data Fig. 2d).

Surprisingly, no significant differences were found in the neuron subtype composition of TRAPed populations between the FR and NF groups (Fig. 2f, Extended Data Fig. 2e), which suggests a lack of training-dependent recruitment of neuron types during consolidation compared to baseline active populations in a NF memory scenario. Both excitatory and inhibitory neuron types were found in active FR populations, with glutamatergic cells comprising about 60–70%. Within the same FR brains, active and inactive populations had roughly similar compositions of neuron subtypes, with the exception of C2-*Calb2* and C3-*Tesc*, suggesting only slight shifts in the recruitment or retirement of neuron subtypes due to activity.

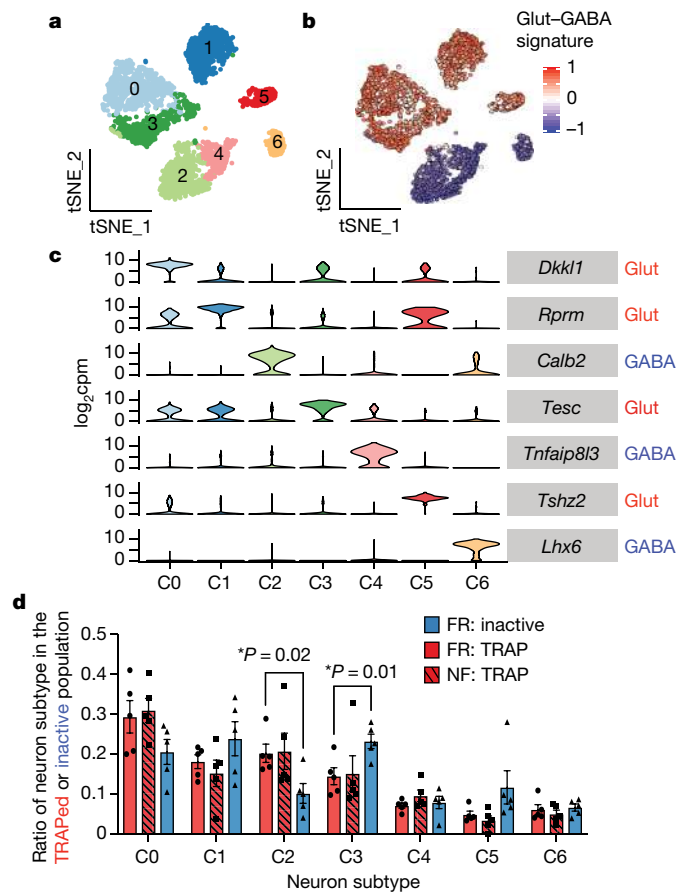
### Memory-associated gene expression

To determine whether remote-memory-associated transcriptional changes occur in recall-activated neurons, we looked for differentially expressed genes (DEGs;  $\log_2$  fold change ( $\log_2FC$ ) > 0.3 and false discovery rate (FDR) < 0.01) in TRAPed FR versus NF cells (Fig. 3a). Single-cell resolution enables a comparison of neurons within the same subtype and the identification of genes that are specifically associated with memory consolidation and recall. Of 23,355 genes, 1,292 were found

to be consolidation-dependent. Expression patterns indicated an overall transcriptional activation, with more genes upregulated than downregulated. Interestingly, DEGs were heterogenous across neuron subtypes, which suggests that remote memory consolidation involves subtype-specific transcriptional programmes (Fig. 3b).

We applied a set of strict criteria to identify possible effector genes. First, each DEG had to be differentially expressed in at least three-quarters of biological replicates, enforcing reproducibility. The removal of DEGs that are also differentially expressed between the inactive populations in FR versus NF mice allowed the identification of changes that were specific to active populations (Extended Data Fig. 3a). Next, DEGs must be differentially expressed when FR cells are compared to NR and HC controls, ensuring that DEGs are not just a consequence of a fear experience. Last, DEGs had to pass a permutation test with shuffled labels (Extended Data Fig. 3b). These criteria produced a set of 99 ‘remote-memory-associated DEGs’ (Fig. 3c; see Methods). Several genes encoded proteins with regulatory roles, including known regulators of transcription (*Hmg20a*, *Hnrnpk* and *Zfp706*) and translation (*Nck2*, *Alpl1* and *Eif2ak1*). Interestingly, even among the condensed list of remote-memory-associated DEGs, we found strong enrichments in genes encoding proteins involved in vesicle exocytosis (*Vamp2*, *Gdi2*, *Rab15*, *Rab5a*, *Rab24*, *Atp6v0c*, *Syt13*, *Stx1b* and *Nsf*), transmembrane transport (*Slc30a9*, *Slc25a46*, *Mfsd14a*, *Tmem50a*, *Gpm6a*, *Mfsd14b* and *Abcf3*), dendritic spine organization (*Strip1*, *Pls3* and *Gsk3b*) and long-range intracellular transport (*Timm29*, *Atad1*, *Pak1*, *Plehhb2*, *Sarnp*, *Rtn3*, *Dmnt1*, *Sar1a* and *Hid1*) (Fig. 3c, Extended Data Fig. 3b, c). More than half of the remote-memory-associated DEGs are associated with neuronal diseases, suggesting links between the functional role of these genes to various memory-affecting neuronal disorders, in addition to the regulation of remote memory.

We further investigated the specificity of our findings by analysing TRAPed neurons that were activated by a salient experience unrelated to fear memory (Extended Data Fig. 4a). Using food deprivation as a salience signal, we identified TRAPed neuronal ensembles that contained the same neuronal subtypes as in TRAPed FR ensembles, but with differences in the subtype composition ratios (Extended Data Fig. 4b–d).



**Fig. 2 | Molecular identification of active neurons during remote memory consolidation.** **a**, Dimensional reduction of all *Snap25*<sup>+</sup> neurons ( $n = 3,691$  cells) reveals seven distinct neuronal subtypes (C0–C6). **b**, Neuronal subtypes fall into two distinct categories: excitatory (glutamatergic) and inhibitory (GABAergic). The Glut–GABA signature is calculated based on the difference of the scaled expression level of *Gad1* and *Slc17a7*. **c**, Expression levels of the top marker genes for each neuron subtype (C0–C6). **d**, Differences in the composition of neuron subtypes of TRAPed populations in FR and NF conditions, as well as inactive populations in FR mice (1 mouse per condition per replicate,  $n = 5$  replicates, two-sided *t*-test) in the composition of TRAPed populations between FR and NF conditions are found. FR TRAPed populations are composed of significantly more C2 (GABAergic) neurons and less C3 (glutamatergic) neurons than the inactive population (error bars indicate s.e.m.).

While we found a total of 143 DEGs (FDR < 0.01) when comparing salience to no salience groups, there was almost no intersection of these DEGs with those found in FR mice (Extended Data Figs. 4e, 5). Thus, while new transcriptional programmes are activated in salience ensembles, the nature of these molecular changes is experience-specific and probably modulated by the particular valence of and/or functional requirements arising from the experience.

Hierarchical clustering of TRAPed FR neurons by the expression levels for each remote-memory-associated DEG allowed distinct populations of ‘highly activated’ and ‘lowly activated’ cells to emerge, suggesting that different transcriptional modules are concertedly regulated during memory consolidation in each neuronal subtype (Fig. 3d). To determine the subtype specificity of these modules, we found that the fraction of cells activated with the subtype-specific DEGs was generally highest in the corresponding subtype when compared to the activation levels in other subtypes or in the inactive populations (Fig. 3e, Extended Data Fig. 6a). Together, this could indicate the presence of subtype-specific common regulatory elements.

To address this possibility, we analysed our DEGs using hypergeometric optimization of motif enrichment (HOMER) to search for common regulatory motifs in an unbiased manner (see Methods; Extended Data Fig. 6b). We found 12 putative de novo and two known motifs enriched within our target DEG set ( $P < 0.01$ ). While we did not find significant enrichment of motifs within subtype-specific DEGs, the *Hif1b* binding motif was found in >40% of total DEGs, including the synaptic transmission and plasticity-related genes *Rab5a*, *Rab24*, *Vamp2*, *Gdi2*, *Gpm6a*, *Strip1*, *Ptp4a1*, *Trim32*, *Mfsd14a*, *Mfsd14b* and *Slc25a46*. Interestingly, these findings agree with recent studies indicating a potential dual role for HIF1 transcription factors during hippocampal-dependent spatial learning and consolidation under normoxic conditions<sup>11</sup>. Interestingly, motifs associated with *Creb*, *Nfkb*, *Chp* and *C/ebp*—canonical transcriptional regulators of neuronal activity and plasticity<sup>12,13</sup>—were absent near the transcription start site (–400 to +100 bp).

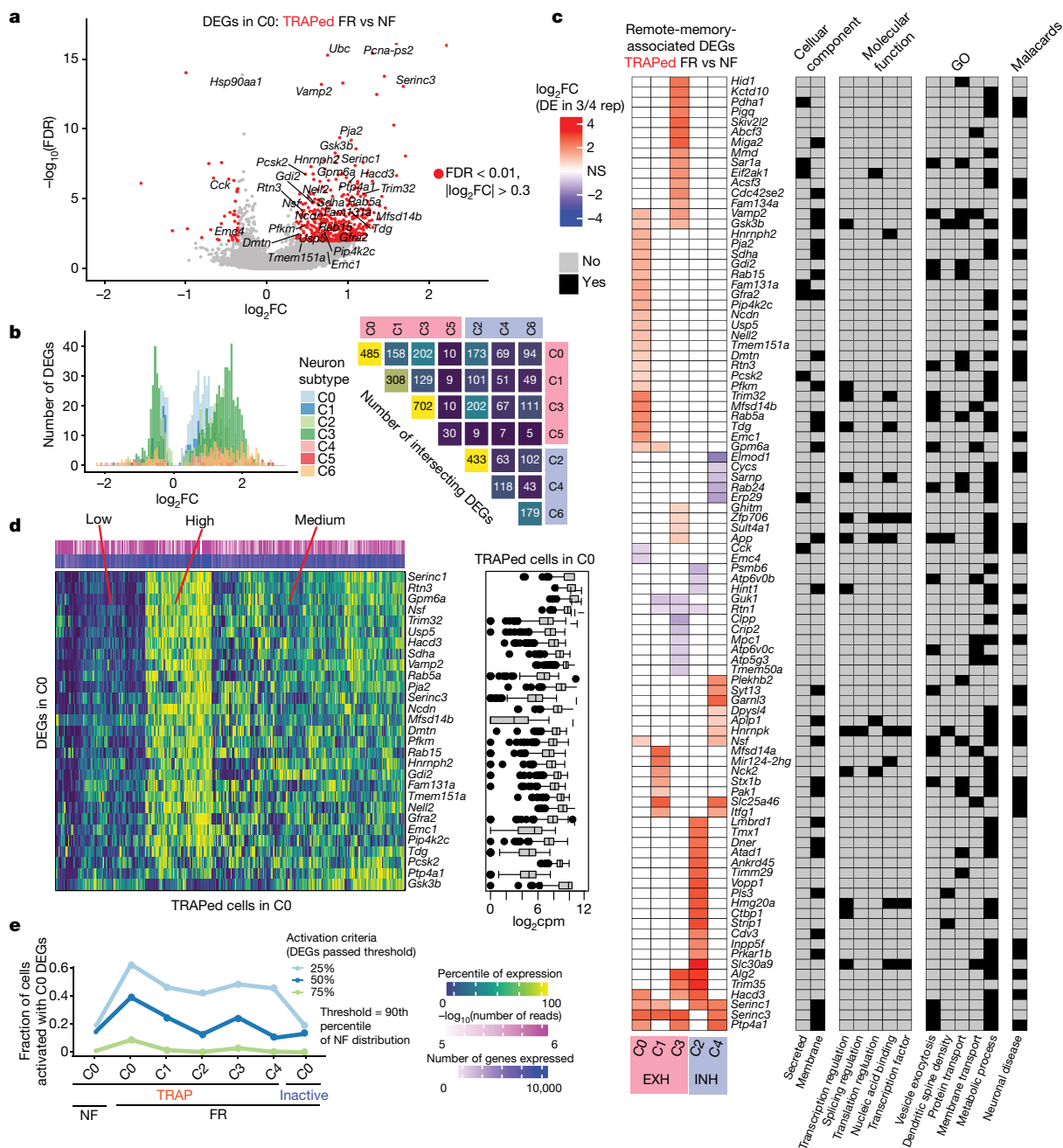
### Vesicle exocytosis signatures in memory

To further elucidate the significance of these remote-memory-dependent transcriptional programmes, we used STRING to look for known and predicted protein–protein interactions. *K*-means clustering of the gene nodes revealed a significantly connected network ( $P = 1.75 \times 10^{-6}$ ) that was centred around a large cluster of genes related to vesicle-mediated transport, exocytosis and neurotransmitter secretion, all of which were highly connected (confidence = 0.4; Extended Data Fig. 6c). Remarkably, 20 out of 99 remote-memory-associated DEGs fell within these functional categories, including *Stx1b*, *Syt13*, *Vamp2*, the SNAP receptor (SNARE) ATPase (*Nsf*) and the GTPase *Rab5a*, all of which are functionally linked to the SNARE complex and to vesicle exocytosis (Fig. 4a). Interestingly, the two most highly and ubiquitously upregulated genes across subtypes were *Serinc1* and *Serinc3*, which are thought to be serine incorporators<sup>14</sup>. Notably, phosphatidylserine phospholipids are calcium-dependent binding partners for synaptotagmins<sup>15</sup>, suggesting that *Serinc1* and *Serinc3* may have important roles in enhancing phosphatidylserine levels and vesicle membrane fusion during memory consolidation. Finally, in situ hybridization confirmed the endogenous proportions of neuronal subtypes in TRAPed populations (Extended Data Fig. 7a, b), as well as the upregulated expression of key remote-memory-associated DEGs, including *Serinc3*, *Syt13*, *Vamp2* and *Stx1b* in respective neuronal subtypes (Fig. 4b, c, Extended Data Fig. 7c).

### Non-neuronal gene expression changes

Remarkably, we discovered that non-neuronal cells also exhibited transcriptional changes associated with remote memory consolidation (FR compared to NF mice; Fig. 5a, b, Extended Data Fig. 8a, b). These signatures were distinct from those of neurons, indicating that non-neuronal programmes may support maintenance of the remote fear-memory trace. Surprisingly, >95% of these DEGs were upregulated, which suggests an overall transcriptional activation during consolidation. Not only was this response detectable weeks after the initial learning but it was observed even without enrichment of the non-neuronal cells directly associated with the TRAPed engram cells (that is, the TRAP method is neuron-specific).

Astrocytes and microglia showed the greatest number of transcriptional changes, with 181 and 308 genes perturbed, respectively ( $\log_2FC > 1$  and FDR < 0.01) (Fig. 5c). Most of these DEGs represent largely diverging pathways (Fig. 5d). In particular, upregulated astrocytic genes were enriched in lipid, cholesterol and steroid metabolic functions (*Gja1*, *Hmgcr*, *Dhcr7*, *Insig1*, *Acl3*, *Idi1*, *Acsbg1*, *10Asah1* and *Hacd3*) as well as glucose transport (*Abcc5*, *Slc39a1*, *Slc6a1*, *Slc27a1*, *Slco1c1*, *Gnb1* and *Ttyh1*). Enhanced metabolic support from astrocytes may be required during memory consolidation since astrocyte–neuron metabolic coupling is elevated during neuronal activity<sup>16</sup>. Moreover, 95 out of 181 astrocyte DEGs were reproduced when comparing



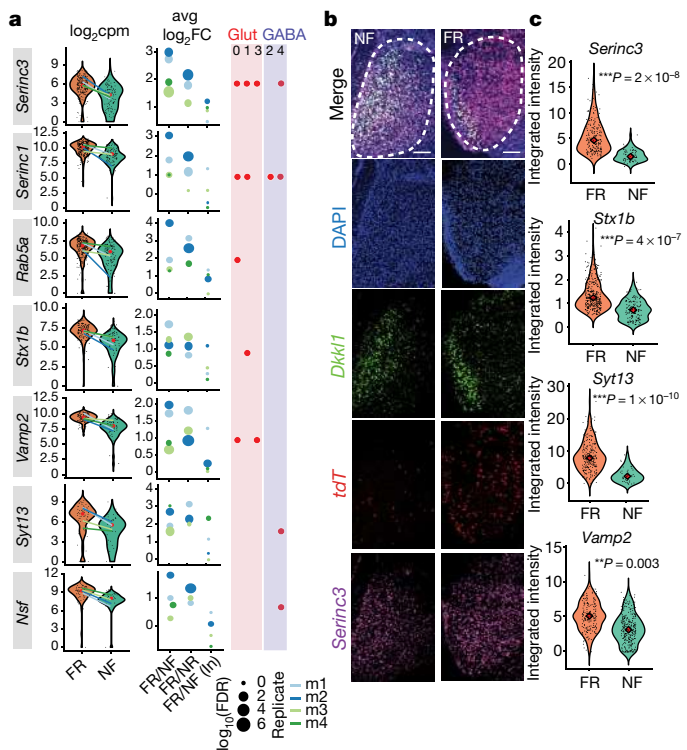
**Fig. 3 | Transcriptional programmes activated by consolidation of remote memories are distinct across neuron subtypes.** **a**, DEGs in FR versus NF in the C0 neuron subtype ( $n = 126$  cells (NF),  $n = 289$  cells (FR)) with all replicated pooled. Differential expression is defined by  $\text{FDR} < 0.01$  and  $|\log_2\text{FC}| > 0.3$  (red points) via a two-sided Mann-Whitney test. Remote-memory-associated DEGs (which remain differentially expressed in three-quarters or more of replicates and when FR is compared to the NR and HC groups (see Extended Data Fig. 3)) are labelled in black. **b**, The number of DEGs per neuron subtype (left) and the number of shared DEGs between each neuron subtype (using pooled cells) (right). **c**,  $\log_2\text{FC}$  of remote-memory-associated DEGs (FR versus

NF) per neuron subtype. Each gene is further annotated with potential biological functions. DE, differentially expressed; EXH, excitatory; GO, Gene Ontology; INH, inhibitory; NS, not significant. *Skiv2l2* is also known as *Mtrex*, and *Fam134a* is also known as *Retreq2*. **d**, The percentile in which a TRAPed neuron (from C0) lies in the distribution of expression of a C0 DEG for all TRAPed C0 cells. The box plots (median  $\pm$  s.d.) show the  $\log_2\text{cpm}$  distribution for each C0 DEG. Hierarchical clustering reveals one common C0 transcriptional programme that is concertedly upregulated. **e**, The fraction of cells in each neuron subtype that is activated with the transcriptional programme (DEGs) from C0.

FR to NR mice, suggesting that a large portion of DEGs is specific to the recall experience itself and not merely a remnant of the fear experience.

By contrast, DEGs from microglial cells were enriched in innate immunity (*Il6r*, *Stat6*, *Csf3r*, *Il1a*, *Irf5*, *Cd86*, *Tnfrsf1b*, *Ywhaz*, *Litaf*, *Ptgs1*, *Gdi2* and *Rnf13*) and cytoskeletal reorganization/focal adhesion maintenance pathways (*Cdc42*, *Rhoa*, *Rhoc*, *Prkcd*, *Vasp*, *Arf6*,

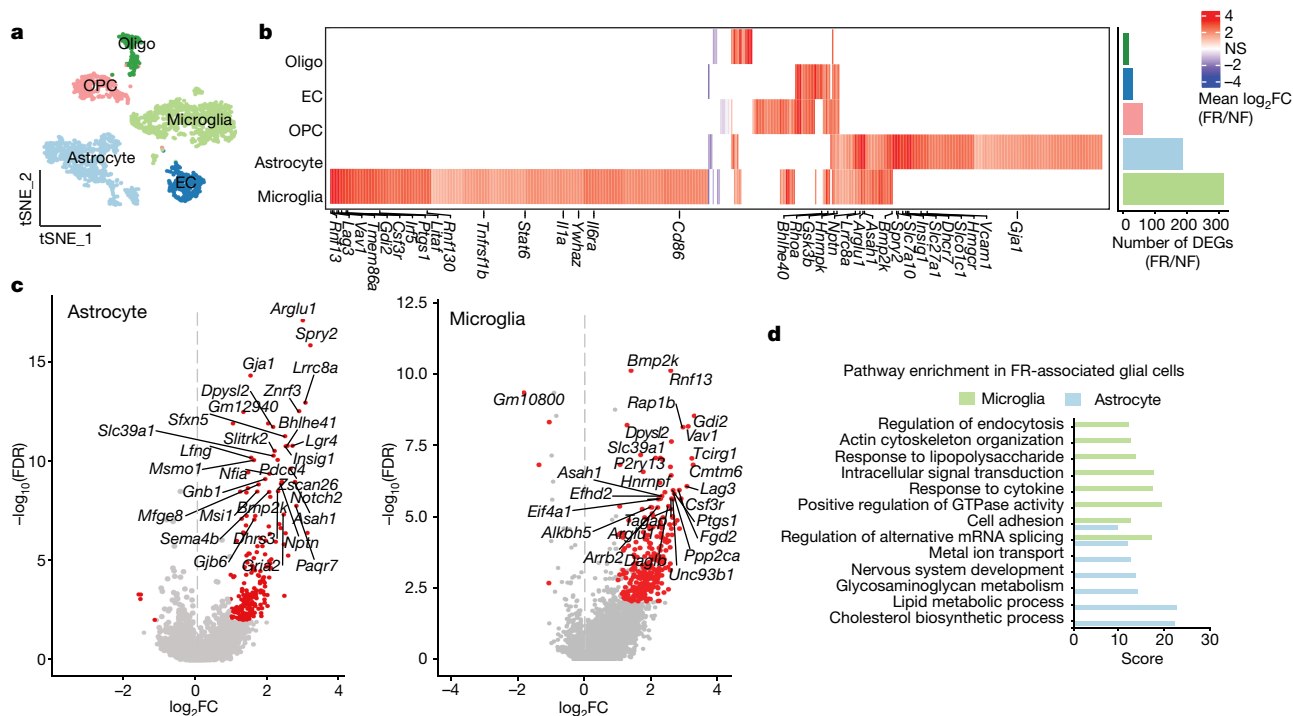
*Vav1* and *Actr2*), suggesting that upregulation of specific inflammatory molecules and enhancement of cell migration may be involved in the maintenance of memory. While less is known regarding the immunomodulatory roles of microglia in memory and learning, previous studies have shown that low levels of inflammatory cytokines (such as IL-1, IL-6 and tumour necrosis factor) can regulate neuronal circuit remodelling and long-term potentiation<sup>17</sup>.



**Fig. 4 | Remote memory consolidation is associated with specific markers for vesicle exocytosis.** **a**, A subset of remote-memory-associated DEGs that are known to regulate vesicle membrane fusion and exocytosis at the presynaptic terminal. The violin plots are overlaid with a drumstick plot indicating the average expression per mouse. The red points represent the median. The bubble plots depict the  $\log_2FC$  and the degree of significance (FDR) for each replicate and are coloured by such. In addition to the FR/NF  $\log_2FC$  (first column), DEGs were also confirmed to be differentially expressed when compared to NR (second column), and their activation is specific only to the active (inactive (In)) (last column). The red dots in coloured panels indicate which neuronal subtype these particular genes are upregulated in. **b**, Representative in situ images of *Serinc3* expression (purple) in *Dkk1<sup>+</sup>/tdT<sup>+</sup>* / DAPI<sup>+</sup> cells in the mPFC. Scale bars, 100  $\mu$ m. **c**, In situ validation of key genes involved in vesicle exocytosis in various neuron subtypes, including *Serinc3* (in the *Dkk1<sup>+</sup>* subtype,  $n = 268$  (FR),  $n = 62$  (NF) cells), *Stx1b* (in the *Rprm<sup>+</sup>* subtype,  $n = 342$  (FR),  $n = 144$  (NF) cells), *Syt13* (in the *Tnfai81p<sup>+</sup>* subtype,  $n = 244$  (FR),  $n = 44$  (NF) cells) and *Vamp2* (in the *Tesc<sup>+</sup>* subtype,  $n = 326$  (FR),  $n = 292$  (NF) cells). Each point represents the normalized integrated intensity of the probe per cell analysed.

In addition to neuron–neuron coupling, communication programmes between neurons and non-neuronal cells may support the memory trace over long periods. We looked for the expression

of receptors or ligands in non-neuronal cells whose known binding partner<sup>18</sup> is perturbed in TRAPed FR neurons (Extended Fig. 8c–e). We focused on genes that were differentially expressed in both the ligand-bound and the receptor-bound cell type (Extended Data Fig. 8c). In FR mice, we found upregulation of neuronal neuroligin-1 and neuroligin-3 (encoded by *Nlgn1* and *Nlgn3*, respectively) and its binding partner neuexin-1 (encoded by *Nrxn1*) on astrocytes, complexes that may enhance neuron–glia adhesions and modulate synaptic function<sup>19</sup>. Thus, the concerted upregulation of these binding pairs in FR mice strongly suggests a role for astrocyte–neuroligin



**Fig. 5 | Transcriptomic changes in non-neuronal cells associated with remote memory consolidation.** **a**, Sub-clustering of non-neuronal cells reveals five non-neuronal cell types (astrocyte, endothelial (EC), microglia, OPC and oligodendrocyte (oligo)) that were collected in an unbiased manner through tdT-negative flow cytometry gates. **b**, DEGs in non-neuronal cells (FR versus NF) (left). DEGs are defined as  $\log_2FC > 1$  and  $FDR < 0.01$ . The number of DEGs that satisfy these criteria in each non-neuronal cell type is also shown (right). The top DEGs (FR versus NF) for glial cells (astrocytes and microglia)

that are also differentially expressed in FR versus NR are labelled. **c**, DEGs (determined by two-sided Mann–Whitney test) that are upregulated and downregulated in astrocytes (left) and microglia (right) in FR versus NF mice. DEGs ( $FDR < 0.01$ ,  $\log_2FC > 1$ ) are indicated by red dots, and the top DEGs are labelled in black. **d**, Pathway analysis of the DEGs (FR versus NF) in microglia and astrocytes. The score is defined as the  $-\log_2(P\text{value})$  using the GeneAnalytics software.

interactions in the maintenance of synaptic strength during fear memory storage.

## Discussion

While high-resolution gene expression atlases of the brain have provided invaluable information about cellular taxonomy<sup>10,20</sup>, characterization of activity-dependent states within these cell types is necessary to understand how experience modulates gene expression, synaptic plasticity and neuronal circuitry. The ability to form and maintain unique synaptic connections that encode a particular memory out of a vast pool of other experiences requires a complex coding space. By using a combination of activity-dependent labelling of neurons and single-cell transcriptomics, we discovered that (1) all mPFC neuron types can be activated during consolidation of remote memory via heterogenous transcriptional programmes; (2) enhanced membrane fusion and vesicle exocytosis may be a critical mode of synaptic strengthening during memory consolidation; (3) a specific set of exocytosis-related genes out of a vast coding space may be involved in allowing highly unique, experience-specific connections to be made; (4) these particular transcriptional programmes are detectable at remote time points and thus are probably involved in maintaining the memory trace weeks after learning; and (5) consolidation of remote memory also induces a persistent transcriptional programme in astrocytes and microglia.

Deciphering the temporal evolution of engram populations and their associated gene programmes through the various stages of initial learning, recent memory and remote memory is crucial for understanding the basis of conversion of short-term memories to long-term memories. We found that the majority of gene programmes affected in activated neurons during early stages of learning<sup>21–23</sup> and recent memory<sup>24</sup> do not intersect with our remote-memory-associated DEGs (Extended Data Fig. 9a), nor with genes enriched in TRAPed FR populations over inactive ones (Extended Data Fig. 9b). This suggests that remote memory could be governed by temporally unique transcriptional programmes. However, future experiments using unified technologies to deconvolve the neuronal compositions of recent and remote engrams and identify the immediate transcriptional changes in recent memory will be of great importance. Relevant to this point, TRAPed neurons in NR mice also exhibited continuous transcriptional changes at moderate levels (when compared to NF mice) (Extended Data Fig. 10). However, these DEGs are largely non-intersecting with remote-memory-associated DEGs, which suggests that the experience of fear itself can induce long-lasting changes in gene expression programmes and that the process of recall induces new transcriptional programmes in a different set of neurons. The current data therefore provide the first step towards deciphering the transcriptional coding landscape that is specifically associated with remote memory consolidation.

## 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/s41586-020-2905-5>.

- Kandel, E. R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030–1038 (2001).
- Flexner, L. B. & Flexner, J. B. Effect of acetoxycycloheximide and of an acetoxycycloheximide–puromycin mixture on cerebral protein synthesis and memory in mice. *Proc. Natl Acad. Sci. USA* **55**, 369–374 (1966).
- Alberini, C. M. & Kandel, E. R. The regulation of transcription in memory consolidation. *Cold Spring Harb. Perspect. Biol.* **7**, a021741 (2014).
- Squire, L. R. Mechanisms of memory. *Science* **232**, 1612–1619 (1986).
- Kitamura, T. et al. Engrams and circuits crucial for systems consolidation of a memory. *Science* **356**, 73–78 (2017).
- DeNardo, L. & Luo, L. Genetic strategies to access activated neurons. *Curr. Opin. Neurobiol.* **45**, 121–129 (2017).
- DeNardo, L. et al. Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nat. Neurosci.* **22**, 460–469 (2019).
- McKenzie, S. & Eichenbaum, H. Consolidation and reconsolidation: two lives of memories? *Neuron* **71**, 224–233 (2011).
- Guenther, C. J., Miyamichi, K., Yang, H. H., Heller, H. C. & Luo, L. Permanent genetic access to transiently active neurons via TRAP: targeted recombination in active populations. *Neuron* **78**, 773–784 (2013).
- Zeisel, A. et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
- O'Sullivan, C. N., Sheridan, G. & Murphy, K. in *Transcription Factors CREB and NF-κB: Involvement in Synaptic Plasticity and Memory Formation* 43–65 (Bentham Science, 2012).
- Suzuki, A. et al. Upregulation of CREB-mediated transcription enhances both short- and long-term memory. *J. Neurosci.* **31**, 8786–8802 (2011).
- Kida, S. et al. CREB required for the stability of new and reactivated fear memories. *Nat. Neurosci.* **5**, 348–355 (2002).
- Inuzuka, M., Hayakawa, M. & Ingi, T. Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. *J. Biol. Chem.* **280**, 35776–35783 (2005).
- Zhang, X., Rizo, J. & Südhof, T. C. Mechanism of phospholipid binding by the C2A-domain of synaptotagmin I. *Biochemistry* **37**, 12395–12403 (1998).
- Bélanger, M., Allaman, I. & Magistretti, P. J. Brain energy metabolism: focus on astrocyte–neuron metabolic cooperation. *Cell Metab.* **14**, 724–738 (2011).
- Williamson, L. L., Sholar, P. W., Mistry, R. S., Smith, S. H. & Bilbo, S. D. Microglia and memory: modulation by early-life infection. *J. Neurosci.* **31**, 15511–15521 (2011).
- Ramilowski, J. A. et al. A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).
- Südhof, T. C. Synaptic neurexin complexes: a molecular code for the logic of neural circuits. *Cell* **171**, 745–769 (2017).
- Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e16 (2018).
- Lacar, B. et al. Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* **7**, 11022 (2016).
- Cho, J.-H., Huang, B. S. & Gray, J. M. RNA sequencing from neural ensembles activated during fear conditioning in the mouse temporal association cortex. *Sci. Rep.* **6**, 31753 (2016).
- Hrvatin, S. et al. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).
- Rao-Ruiz, P. et al. Engram-specific transcriptome profiling of contextual memory consolidation. *Nat. Commun.* **10**, 2232 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

### Mice

All animal experiments were conducted following protocols approved by the Administrative Panel on Laboratory Animal Care at Stanford University. The *TRAP2; Ai14* mouse line was kindly gifted by the Luo laboratory at Stanford. TRAP2 mice were heterozygous for the *Fos*<sup>2A-iCreER</sup> allele, and homozygous for *Ai14*, and were bred with *Ai14* homozygous mice in the C57BL/6 background. Mice were group-housed (maximum five mice per cage) on a 12 h light–dark cycle (07:00 to 19:00, light) with food and water freely available. Male mice 42–49 days of age were used for all the experiments. Mice were handled daily for 3 days before their first behavioural experiment.

### Genotyping

The following primers: GAG GGA CTA CCT CCT GTA CC (forward) and TGC CCA GAG TCA TCC TTG GC (reverse) were used for genotyping of the *Fos*<sup>2A-iCreER</sup> allele.

### Fear conditioning

The fear conditioning training was performed as previously described<sup>25</sup>. Briefly, mice were individually placed in the fear conditioning chamber (Coulbourn Instruments) located in the centre of a sound attenuating cubicle, which was cleaned with 10% ethanol to provide a background odour. A ventilation fan provided a background noise at approximately 55 dB. After a 2-min exploration period, three tone–foot shock pairings separated by 1-min intervals were provided. The 85 dB 2-kHz tone lasted for 30 s, and the foot shocks were at 0.75 mA and lasted for 2 s. The foot shocks were co-terminated with the tone. The mice remained in the training chamber for another 60 s before being returned to the home cages. For the context recall, mice were placed back into the original conditioning chamber for 5 min 16 days after the training. 4-OHT injections were performed immediately (within 30 min) before the recall experiments. For the HC and the NR groups, 4-OHT was injected at a similar time when the other two groups were subjected to recall. The behaviour of the mice was recorded and analysed with the FreezeFrame software (v.4; Coulbourn Instruments). Motionless bouts that lasted more than 1 s were considered as freeze. Data were analysed with the tracking software Viewer III (Biobserve).

### Food deprivation

Mice were deprived of food for 16 h, and then 4-OHT was injected to the animals and food was returned to one group (salience group) immediately afterwards (within 30 min), while no food was returned to the other group (no salience group) until 10 h later.

### tdT fluorescence examination

Mice were deep anaesthetized with tribromoethanol and perfused with PBS followed by fixative (4% paraformaldehyde diluted in PBS). The brains were then removed and postfixed in 4 °C overnight and immersed in 30% sucrose solution for 2 days before being sectioned at a thickness of 50 µm on a cryostat (CM3050 S, Leica Biosystems). Imaging was performed with a scanning microscope (BX61VS, Olympus).

### Single-cell dissociation and flow cytometry

mPFC regions were microdissected from live vibratome sections (300 µm thick) of the prefrontal cortex. Tissue pieces were enzymatically dissociated via a papain-based digestion system (LK003150, Worthington). Briefly, tissue chunks were incubated in 1 ml of papain (containing L-cysteine and EDTA), DNase and kynurenic acid for 1 h at 37 °C and 5% CO<sub>2</sub>. After 10 min of incubation, tissues were triturated briefly with a P1000 wide bore pipette tip and returned. Cells were triturated another four times (around 30 each) with a P200 pipette tip over the rest of the remaining incubation time. At room temperature, cell suspensions were centrifuged at 350g for 10 min, resuspended in 1 ml EBSS with 10% v/v

ovomucoid inhibitor, 4.5% v/v DNase and 0.1% v/v kynurenic acid, and centrifuged again. Supernatant was removed and 1 ml ACSF was added to cells. ACSF was composed of: 1 mM KCl, 7 mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub>, 1.3 mM NaH<sub>2</sub>PO<sub>4</sub>, 110 mM choline chloride, 24 mM NaHCO<sub>3</sub>, 1.3 mM Na ascorbate, 20 mM glucose and 0.6 mM sodium pyruvate. Cells were passed through a 70-µm cell strainer to remove debris. Hoechst stain was added (1:2,000; H3570, Life Technologies) and incubated in the dark at 4 °C for 10 min. Samples were centrifuged (350g for 8 min at 4 °C) and resuspended in 0.5 ml of ACSF and kept on ice for flow cytometry.

Cells were sorted via the Sony SH800 into 96-well or 384-well plates (Bio-Rad) directly into lysis buffer<sup>26</sup> with oligodT, and immediately snap frozen until processing. A positive 'TRAP' gate was set for cells that were both Hoechst<sup>+</sup> and tdT<sup>+</sup>. A negative 'TRAP' gate was set for all Hoechst<sup>+</sup> and tdT<sup>-</sup> cells in general. No gating on forward or backscatter was used to avoid size biases that might be present in a heterogenous neuronal population. Each plate was kept on the sorter for <25 min to prevent evaporation.

### Sequencing

Whole-cell lysis, first-strand synthesis and cDNA synthesis were performed using the Smart-seq-2 protocol as described previously<sup>26</sup> in both 96-well and 384-well formats, with some modifications. After cDNA amplification (23 cycles), cDNA concentrations were determined via capillary electrophoresis (96-well format) or the PicoGreen quantitation assay (384-well format), and wells were chosen to improve quality and reduce cost of sequencing. Only wells with >0.2 ng µl<sup>-1</sup> of cDNA were selected and cDNA concentrations were subsequently normalized to ~0.2 ng µl<sup>-1</sup> per sample, using the TPPLabtech Mosquito HTS and Mantis (Formulatrix) robotic platforms. Libraries were prepared, pooled and cleaned using the Illumina Nextera XT kits or in-house Tn5, following the manufacturer's instructions. Libraries were then sequenced on Nextseq or Novaseq (Illumina) using 2 × 75-bp paired-end reads and 2 × 8-bp index reads with a 200 cycle kit (20012861, Illumina). Samples were sequenced at an average of 1.5 million reads per cell.

### RNAscope

The RNAscope experiment was performed following the manufacturer's instructions using the RNAscope multiplex fluorescent reagent kit v2 (323100, ACD). All probes were purchased from existing stocks or custom designed from ACD.

### Bioinformatics and data analysis

**Mapping to the genome.** Sequences from Nextseq or Novaseq were demultiplexed using bcl2fastq, and reads were aligned to the mm10 genome augmented with ERCC (External RNA Controls Consortium) sequences, using STAR version 2.5.2b. Gene counts were made using HTSEQ version 0.6.1p1. All packages were called and run through a custom Snakemake pipeline. We applied standard algorithms for cell filtration, feature selection and dimensionality reduction. Briefly, genes that appeared in fewer than five cells, samples with fewer than 100 genes and samples with less than 50,000 reads were excluded from the analysis. Out of these cells, those with more than 30% of reads as ERCC, and more than 10% mitochondrial or 10% ribosomal were also excluded from analysis. Counts were log-normalized and then scaled where appropriate.

Next, the 'canonical correlation analysis' function from the Seurat package<sup>27</sup> was used to align raw data from multiple experiments. Only the first ten canonical components were used. After alignment, relevant features were selected by filtering expressed genes to a set of ~2,500 with the highest positive and negative pairwise correlations. Genes were then projected into principal component space using the robust principal component analysis. Single-cell principal component scores and gene loadings for the first 20 principal components were used as inputs into Seurat's (v2) FindClusters and RunTsn functions to calculate 2D tSNE coordinates and search for distinct cell populations.

# Article

Briefly, a shared-nearest-neighbour graph was constructed based on the Euclidean distance metric in the principal component space, and cells were clustered using the Louvain method. Cells and clusters were then visualized using 3D tSNE embedding on the same distance metric. A neuron was characterized as 'TRAPed' trapped if it satisfied two conditions: (1) from the tdT<sup>+</sup> sort gate (tdT protein positive) and (2) *tdT* mRNA raw count > 0. Neuron subtype marker genes were found by using the FindAllMarkers function in Seurat (min.pct = 0.3, thresh.use = 0.25, min.diff.pct = 0.2). DEG analysis was done by applying the Mann–Whitney *U*-test on various cell populations. Raw *P* values were adjusted to an FDR. Permutation tests were then performed on all genes of interest. All graphs and analyses were generated and performed in R. GeneAnalytics and GeneCards packages offered by the gene set enrichment analysis tool were used for GO/KEGG/REACTOME pathway analysis and classification of enriched genes in each subpopulation.

**Finding remote-memory-associated DEGs.** To reduce our list of DEGs (FR TRAP versus NF TRAP results in 1,291 DEGs, cells from 4 biological replicates pooled, log<sub>2</sub>FC > 0.3, FDR < 0.01) to only the most recall-specific, 4 steps were taken. Analysis was limited to C0–C4 neuron subtypes due to insufficient numbers of cells in C5 and C6 across all experimental conditions to make meaningful comparisons. First, DEGs are recalculated by assessing each experiment individually using the whole transcriptome, and only DEGs (via the same criteria as pooled) that intersect in three-quarters of replicates are kept. Three out of four criteria were chosen as a compromise due to the high strictness of four out of four, which yielded only a maximum of seven DEGs (for a neuron subtype). All resulting DEGs are found in the initial DEG list (all replicates pooled), indicating that no additional DEGs were found as a result of analysing replicates separately. Second, 'inactive' (tdT-negative) populations were also compared (FR inactive versus NF inactive) and any DEGs that intersected with the DEGs left after the first criteria, were removed. This ensures that DEGs are activity-dependent, and not merely an overall upregulation in all cells due to the experience. This routinely removed genes such as *Hsp90aa1* and *Pcna-ps2*. Third, the remaining DEGs had to be differentially expressed when FR TRAP was compared to either NR TRAP or HC TRAP. This ensures that the DEGs are specific to only neuronal ensembles that labelled by memory recall, and not due to forms of baseline activity (HC) or activity that remained from the initial fear learning (NR). Last, the remaining DEGs must pass a permutation test in which the training labels are shuffled and a distribution of log<sub>2</sub>FC is computed based on these labels. The true observed log<sub>2</sub>FC must be above the 95th percentile of the distribution of the shuffled distribution. After placing these constraints, 99 genes remain from the original list of 1,291.

**Assessment of activation score.** A TRAPed (or inactive) cell is considered to be 'activated' by the remote-memory DEG programme if 25%, 50% or 75% of the subtype-specific DEGs (remote-memory-associated DEGs only) is expressed above the 90th percentile of the distribution of that gene in NF TRAP controls from the same subtype. This calculation is then repeated with DEG programmes that are specific to each neuronal subtype. The fraction of cells activated with the subtype-specific signature is calculated as the number of activated cells divided by all cells in the subtype/activity group.

**Regulatory motif analysis.** Enrichment of known and de novo motifs was found using HOMER by inputting the list of 99 remote-memory-associated DEGs and using the function findMotifs.pl and

the criteria '-start -400 -end 100 -len 8,10 -p 2'. The locations of the motifs in specific DEGs were found using the -find <motif file> option of findMotifs.pl.

**RNAscope image analysis.** Images were taken using a Nikon Confocal Microscope (at ×10 or ×20, NA = 0.45) and images were processed in ImageJ to only obtain the mPFC regions. The resulting images were fed into a custom image analysis pipeline on CellProfiler (using a combination of the functions IdentifyPrimaryObjects, RelateObjects, FilterObjects, MeasureObjectIntensity, ClassifyObjects and CalculateMath). The custom pipeline can be found in Supplementary Methods). Briefly, images were corrected with control slides (unstained sample and negative control probes) and cells were segmented using the DAPI signal. Those harbouring a signal (above a set threshold level) for both the subtype marker and the tdT probe were retained. The integrated fluorescence intensity of the DEG probe was calculated for each DAPI<sup>+</sup>/subtype<sup>+</sup>/tdT<sup>+</sup> cell. Cells that were not double-positive were not considered. The integrated fluorescent intensity was then normalized to the integrated DAPI signal per cell and results were plotted with custom scripts in R.

## Reporting summary

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

## Data availability

The accession number for the single-cell RNA sequencing data reported in this paper is GSE152632.

## Code availability

Custom scripts can be found at <https://github.com/mbchen-424/memory-sc-rnaseq>.

- Zhou, M. et al. A central amygdala to zona incerta projection is required for acquisition and remote recall of conditioned fear memory. *Nat. Neurosci.* **21**, 1515–1519 (2018).
- Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

**Acknowledgements** We thank L. Chen, M. Zhou and J. Li for discussion of the experimental design; S. Kolluru and D. Henderson for assistance in library preparation; N. Neff and J. Okamoto for assistance with sequencing; J. Lui for advice on brain dissociation; L. Denardo, J. Lui and L. Luo for the gift and help with the TRAP2 line; and W. Wang, G. Stanley and F. Horns for helpful discussions and computational assistance. S.R.Q. is a Chan Zuckerberg Investigator. This work is supported by a grant from the NIH (MH115999 to T.C.S.)

**Author contributions** X.J. and T.C.S. designed the animal experiments. M.B.C. and S.R.Q. designed the single-cell RNA sequencing experiments. X.J. performed the animal experiments, brain dissection, and in situ hybridization and imaging. M.B.C. performed the brain dissociation, flow cytometry, single-cell library preparation and sequencing pipelines. M.B.C. performed all single-cell RNA sequencing data and image analysis, with input from X.J., S.R.Q. and T.C.S. M.B.C. wrote the manuscript with substantial contributions from X.J., S.R.Q. and T.C.S. T.C.S. and S.R.Q. oversaw the project.

**Competing interests** The authors declare no competing interests.

## Additional information

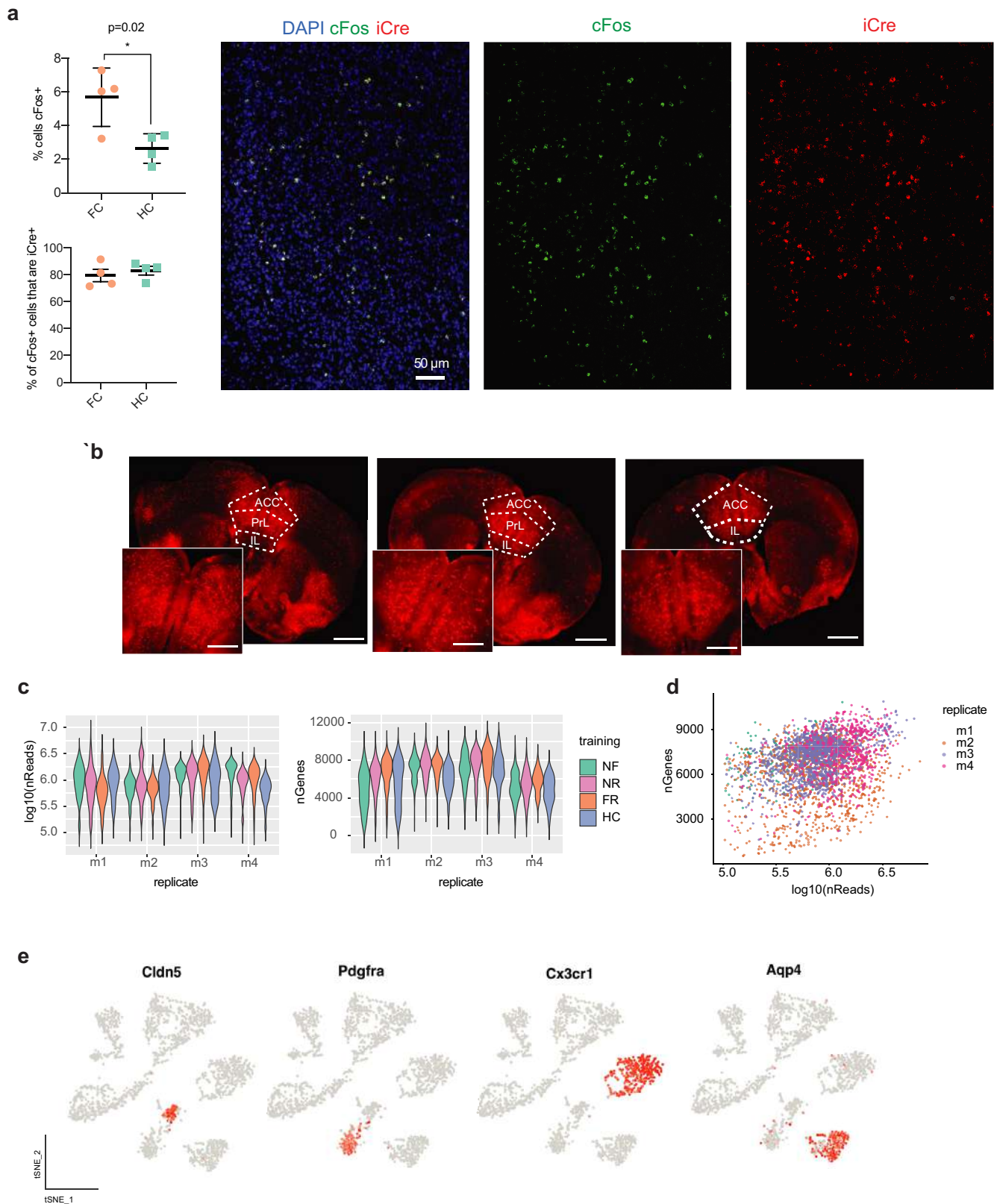
**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2905-5>.

**Correspondence and requests for materials** should be addressed to S.R.Q. or T.C.S.

**Peer review information** Nature thanks Steve Ramirez and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

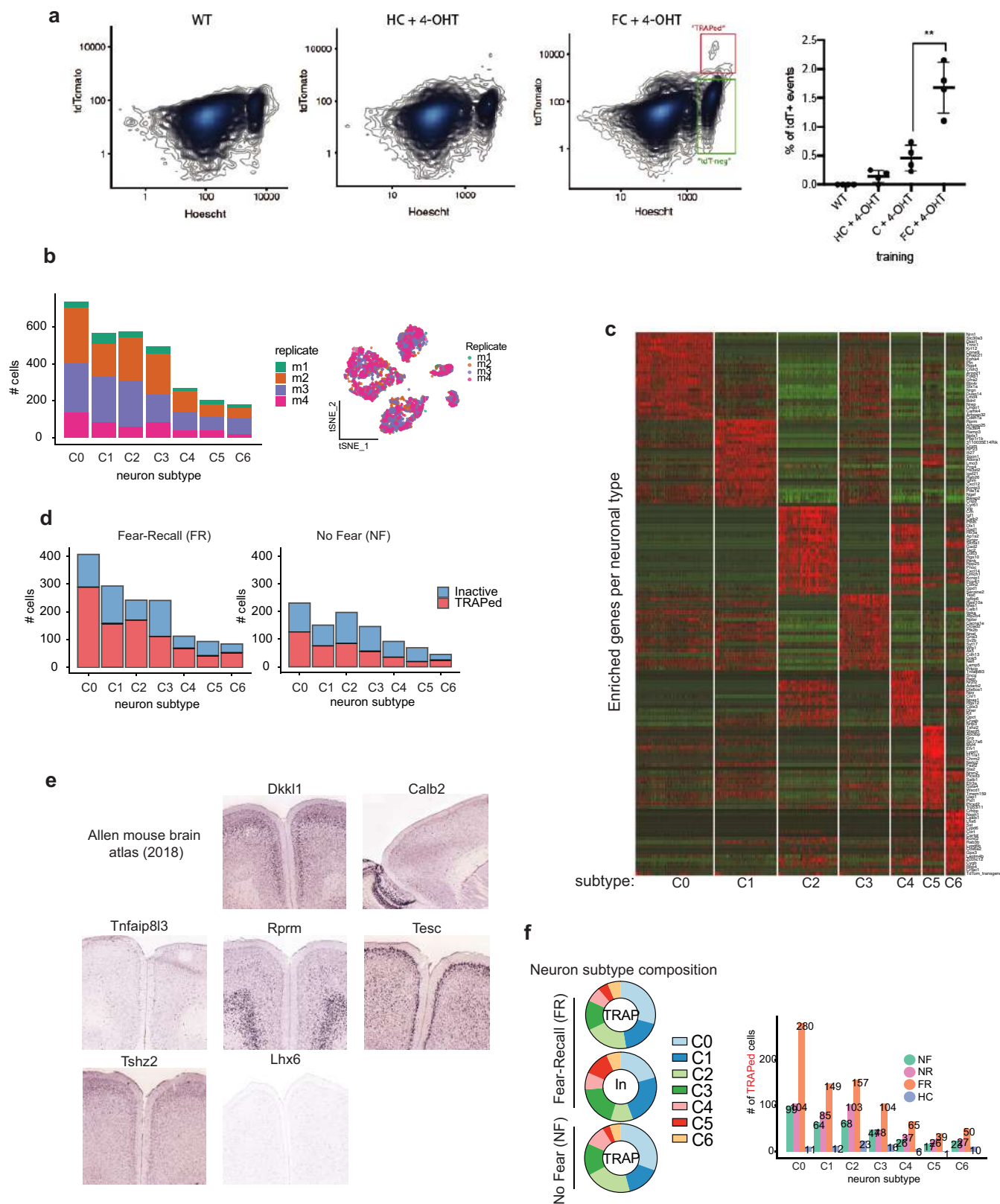
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.





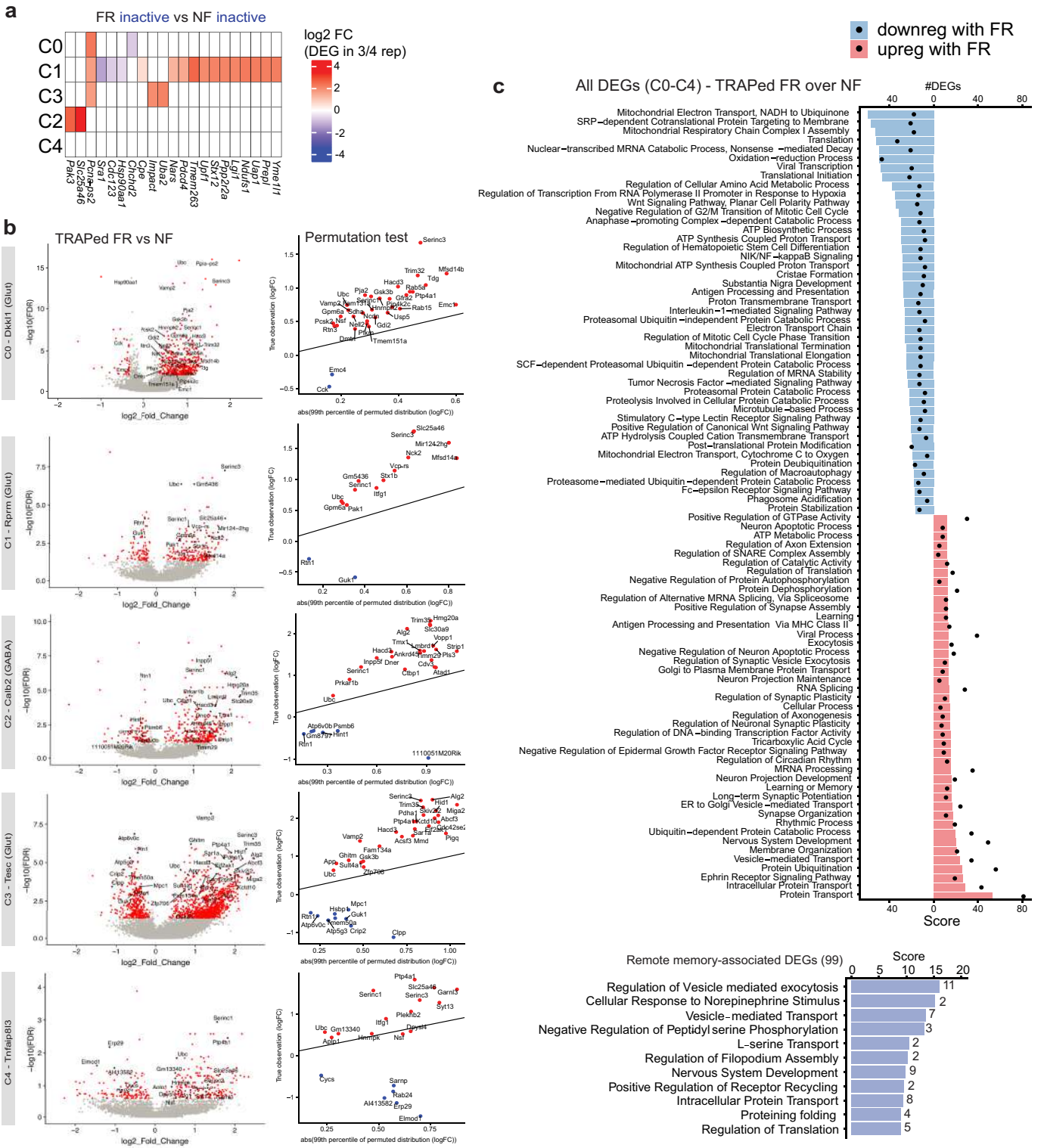
**Extended Data Fig. 1 | Fidelity of the *TRAP2;Ai14* line and sequencing quality metrics. a**, Fidelity with which the tdTomato reporter in *TRAP2;Ai14* mice captures endogenous *cFos* expression during fear memory encoding by in situ hybridization of *cFos* and *iCre* in the mPFC directly following fear conditioning. Left, percentage of *cFos*<sup>+</sup> cells in either fear conditioned (FC) or homecage (HC) mice as seen by in situ staining (mean  $\pm$  s.d.). Percentage of *cFos*<sup>+</sup> cells that are also *iCre*<sup>+</sup> (mean  $\pm$  s.e.m). Right, representative in situ hybridization images of FC mice (mPFC). **b**, Representative images of regions of

the mPFC analysed in this study (anterior cingulate cortex (ACC), the prelimbic (PrL) and the infralimbic cortex (IL)). Scale bars are 1mm (and 0.5mm in the insets). **c**, Violin plots of the number of reads and number of genes per biological replicates (m1–m4), per cell. **d**, Scatterplot depicting the lack of strong relationship between number of genes detected and number of reads obtained per cell. **e**, Scaled expression of canonical markers in non-neuronal cells (*Cldn5*-BECs, *Pdgfra*-OPCs, *Cx3cr1*-microglia, *Aqp4*-astrocytes).



**Extended Data Fig. 2 | Distribution of cell numbers and neuronal subtypes across various training conditions.** **a**, Representative flow cytometry plot of the amount of tdT<sup>+</sup> events per training condition. In scatter plot, each point represents one mouse (mean  $\pm$  s.d.). **b**, Number of cells from each biological replicate that were annotated as one of 7 defined neuron subtypes (C0–C6). **c**, Scaled expression of the top marker genes for each neuron subtype.

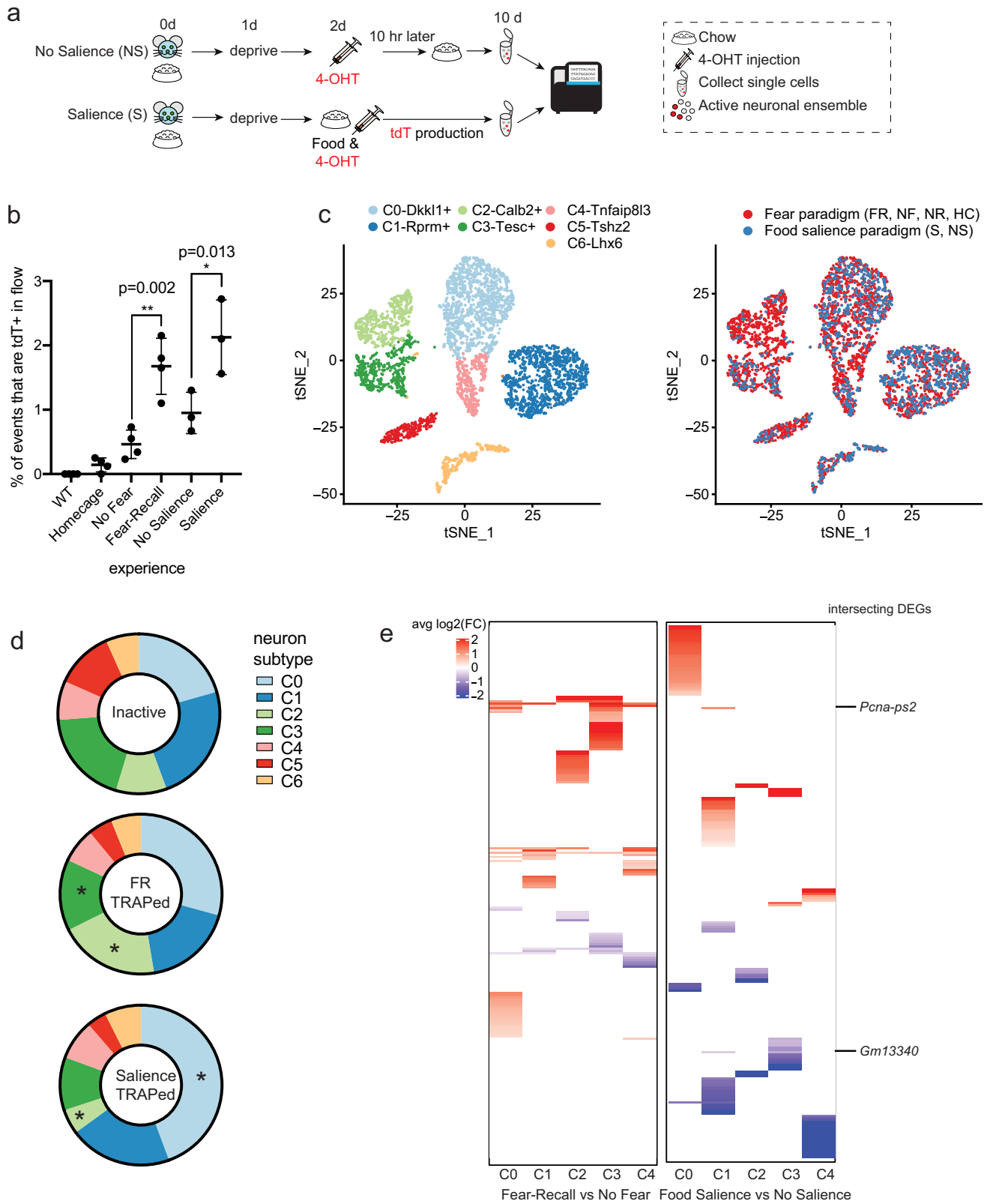
**d**, Number of TRAPed and Inactive cells (as defined by non-zero expression of *tdT* mRNA) collected per neuron subtype, in either fear-recall (FR) or no-fear (NF) mice. **e**, Representative images of subtype marker genes (C0–C6) in the Allen Brain ISH atlas. **f**, Neuron subtype composition of FR and NF populations. Number of TRAPed (tdT mRNA+) cells collected in each experimental condition that fall in one of 7 neuronal subtype categories.



**Extended Data Fig. 3 | Differential gene expression in distinct neuronal subtypes (FR over NF TRAPed populations).** **a**, DEGs in each neuron subtype when inactive (tdT) neurons are compared between FR and NF mice. **b**, Left, volcano plots of DEGs in FR vs NF mice for each neuron subtype (C0 to C4). DEGs found when all replicates are pooled analysed in a combined manner are shown in red. Recall-dependent DEGs (defined as being differentially expressed in at least 3 out of 4 replicates, when analysed replicates are analysed individually) are labelled in black. Right, permutations are performed for every recall-dependent DEG for each neuronal population. Upregulated DEGs lying above the  $y = x$  line (red) and downregulated DEGs lying below the  $y = x$  line

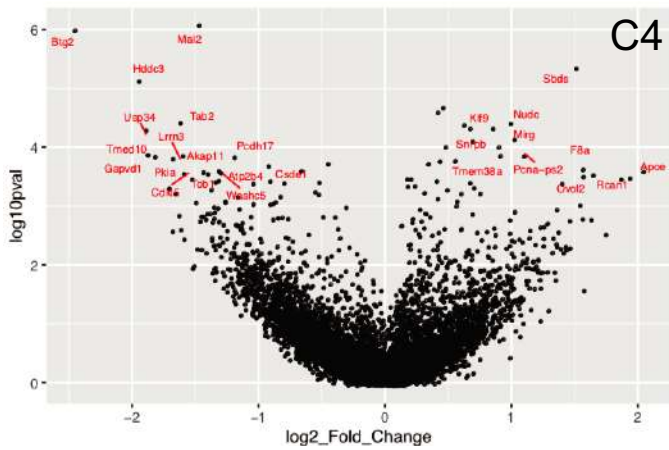
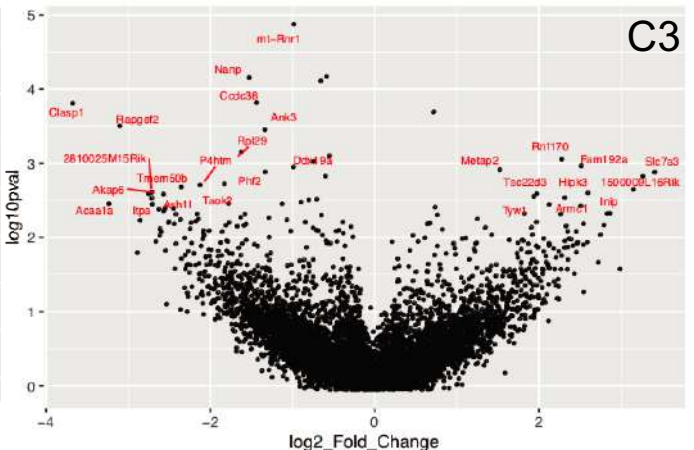
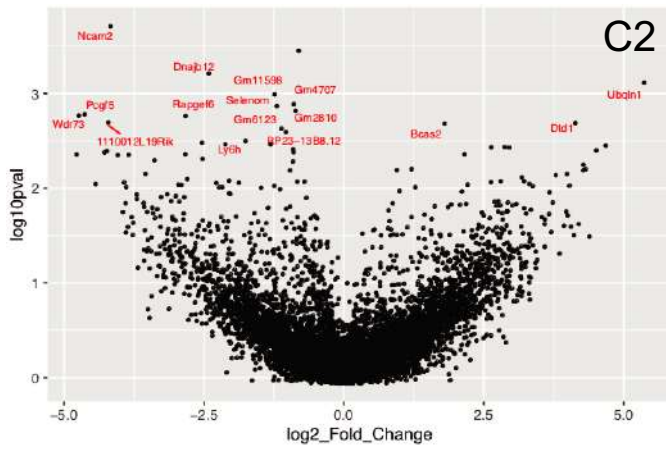
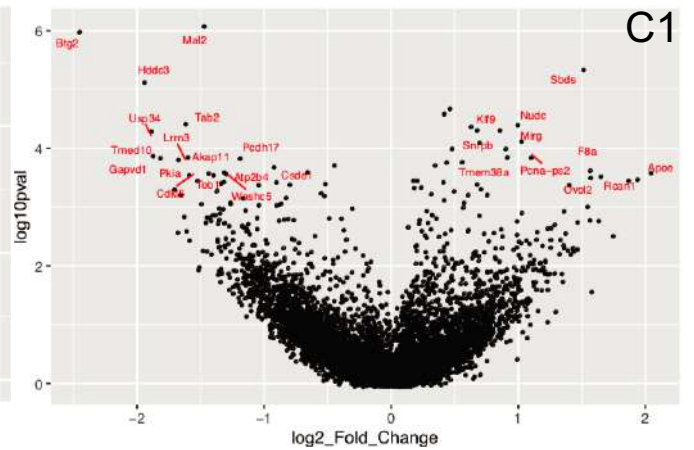
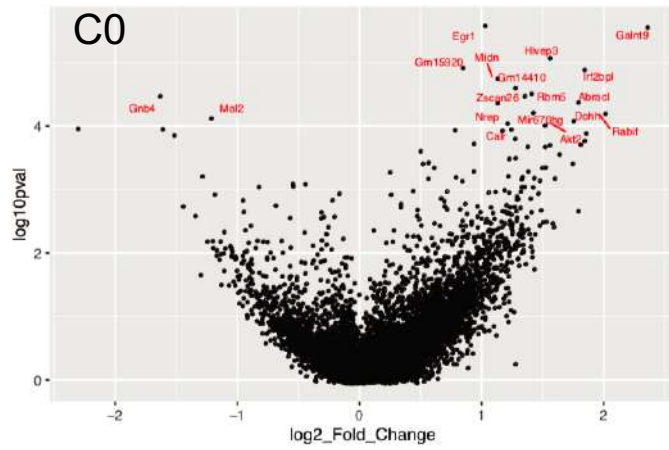
(blue) are considered to be above the 99th percentile of the permuted distribution. **c**, GO enrichment analysis of all up- and downregulated DEGs (941 DEGs up, 384 DEGs down, all neuron subtypes combined) when all replicates are pooled. Bars show the enrichment scores (GeneAnalytics) for the GO pathway and dot indicates the number of DEGs involved in that pathway. **d**, GO enrichment analysis of only the upregulated remote-memory-specific DEGs (from Fig. 3c) from all neuron subtypes combined. Bar indicates enrichment score (Gene Analytics) and number indicates number of recall-dependent DEGs involved in that pathway.

# Article



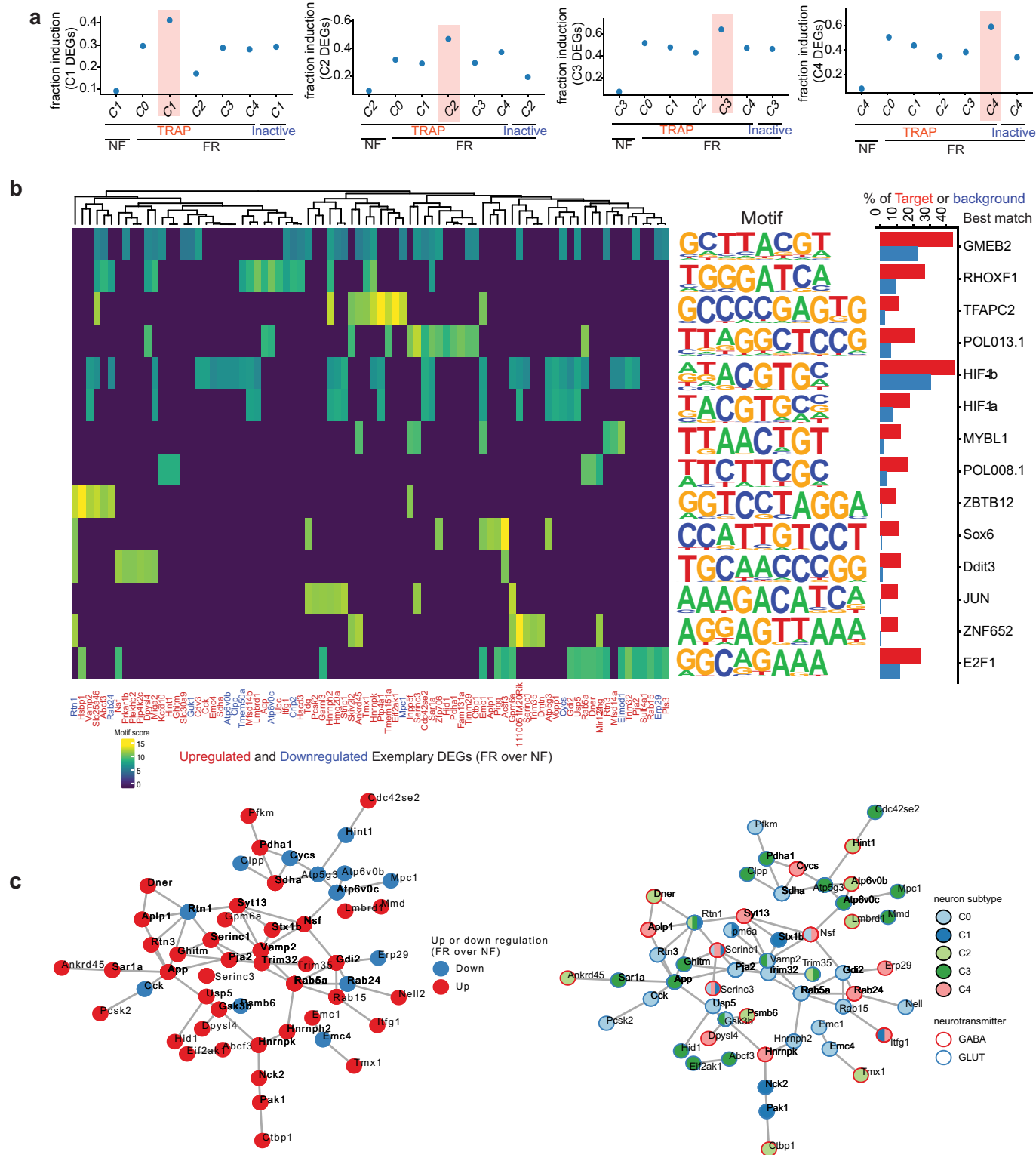
**Extended Data Fig. 4 | Analysis of TRAPed ensembles in food salience (S) versus no salience (NS) mice. a**, Schematic of experimental paradigm for generating TRAPed neuronal ensembles as a result of food salience (food deprivation followed by food return (salience) or no food return (no salience)). **b**, Percentage of events in flow cytometry that were tdT<sup>+</sup>, by experience (mean  $\pm$  s.d.). **c**, tSNE of the merge of data from fear-recall experiments and

food-salience experiments coloured by neuron subtype (left) and experimental paradigm (right). **d**, Subtype composition differences between TRAPed fear-recalled ensembles and TRAPed food salience ensembles, as compared to background Inactive ensembles. **e**, Heat map of the average log<sub>2</sub> fold change of DEGs in each neuron subtype when comparing fear-recall vs no-fear, and salience vs no salience. Only DEGs with FDR < 0.01 are shown.



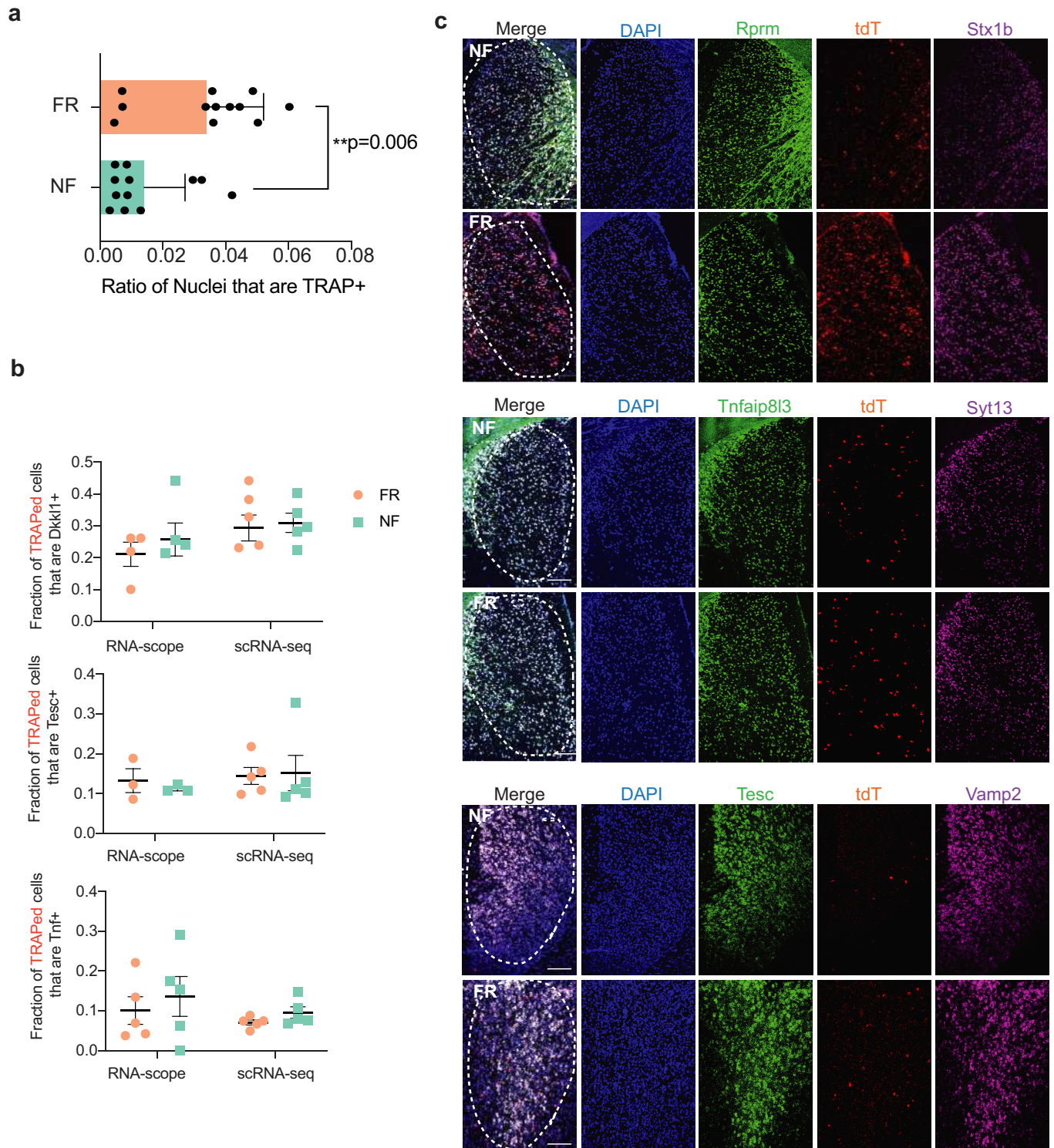
**Extended Data Fig. 5 | DEGs when comparing ensembles from food salience (S) to no salience (NS) mice.** Volcano plots showing the  $\log_2$  fold change and adjusted  $P$  values (in  $\log_{10}$  scale) of genes when comparing food

salience over no salience groups within each neuron subtype (C0 through C4). Top DEGs per neuron subtype are labelled in red. Positive  $\log_2$  fold change indicates upregulation in food salience (S) group.



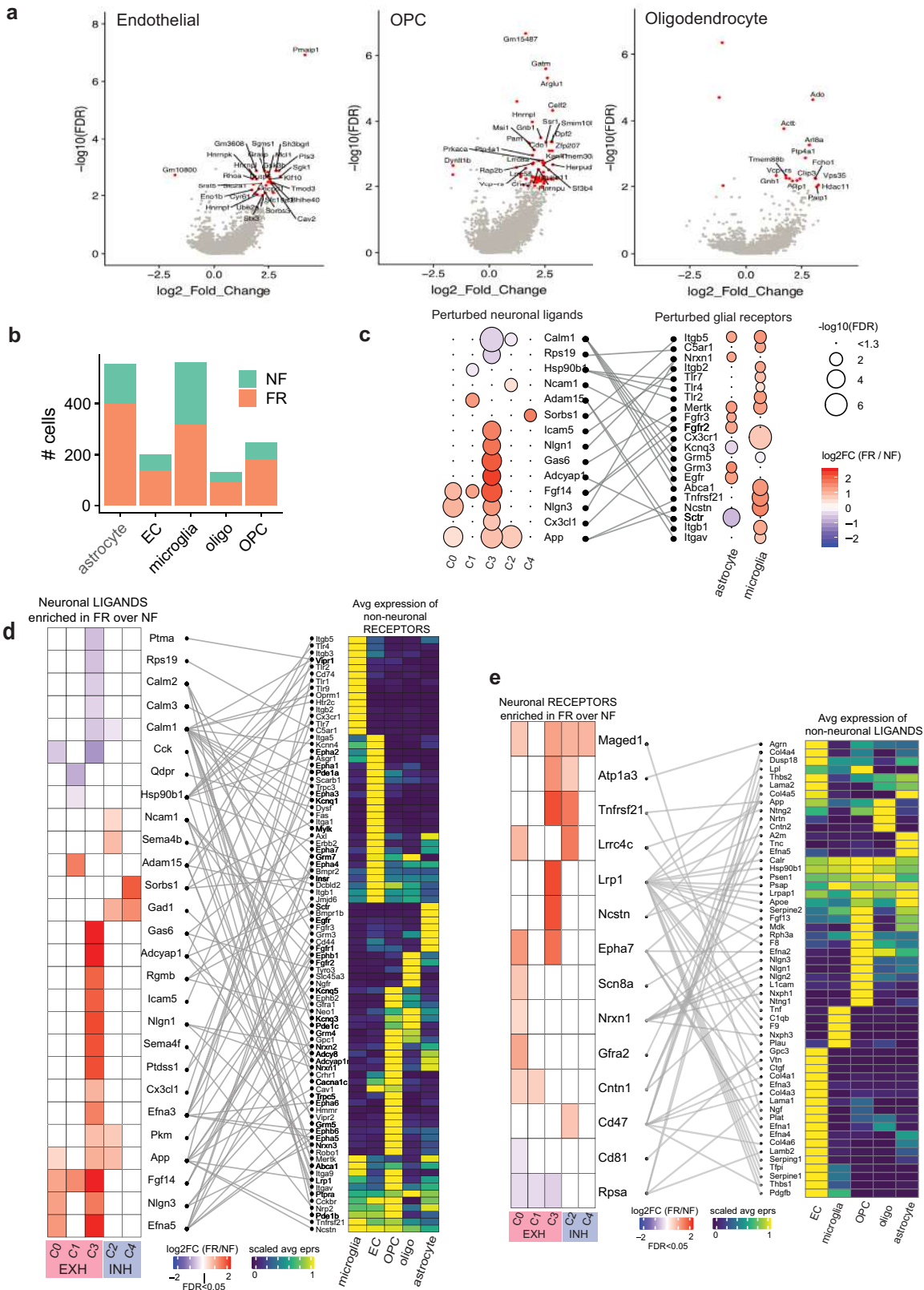
**Extended Data Fig. 6 | Neuron subtype-specific activation programs, hypothesized protein-protein interactions and upstream regulatory motifs. a**, Fraction of cells in each neuron subtype that are induced with the transcriptional program (that is, DEGs) from a neuron subtype. Overall, the activation program of each TRAPed neuron subtype is found to be more specific to it than the inactive population, or other subtypes. **b**, Left, de novo regulator motif discovery: analysis was performed using HOMER on the subset of 99 remote-memory-associated DEGs by looking at the sequences -400 to +100 bp from the TSS. 12 de novo and 2 known motifs were found (only motifs with an enrichment  $P$  value  $<10^{-2}$  were kept). Heat map depicts the 'motif score' of each DEG for each motif, and genes and motifs were clustered via the ward.D

method. Right, bar graph depicting the percentage of the DEGs (target sequences) that possess a match for the motif within -400 to +100 bp from the TSS, vs the percentage of background sequences. For de novo motifs, the best match gene is listed on the right. HIF1b and HIF1a are matches to known motifs. **c**, Left, hypothesized protein-protein interactions of a subset of recall-dependent DEGs (TRAPed FR/NF) using the STRING database (<https://string-db.org/>). Only genes that are connected at a confidence level of 0.4 (medium) are shown. Connections indicate a possible existence of an interaction between two proteins. Genes are coloured by up or downregulation in FR/NF. Right, same network plot, with nodes coloured by the neuron subtype which differentially regulates the DEG.



**Extended Data Fig. 7 | In situ validation of *tdT* levels, neuronal subtype compositions and remote-memory-specific DEGs in the mPFC.** **a**, Ratio of Nuclei that are *tdT*<sup>+</sup> (mRNA level) per training condition. Each data point represents one region of interest. (mean  $\pm$  s.d.) **b**, Ratio of TRAPed cells that are positive for a neuronal subtype marker obtained either via the RNA-scope method, or by scRNA-seq (mean  $\pm$  s.e.m.) (see Fig. 2). TRAPed cells are defined as DAPI<sup>+</sup>/*tdT*<sup>+</sup> in RNAscope quantification, and as *tdT* mRNA count >1 in

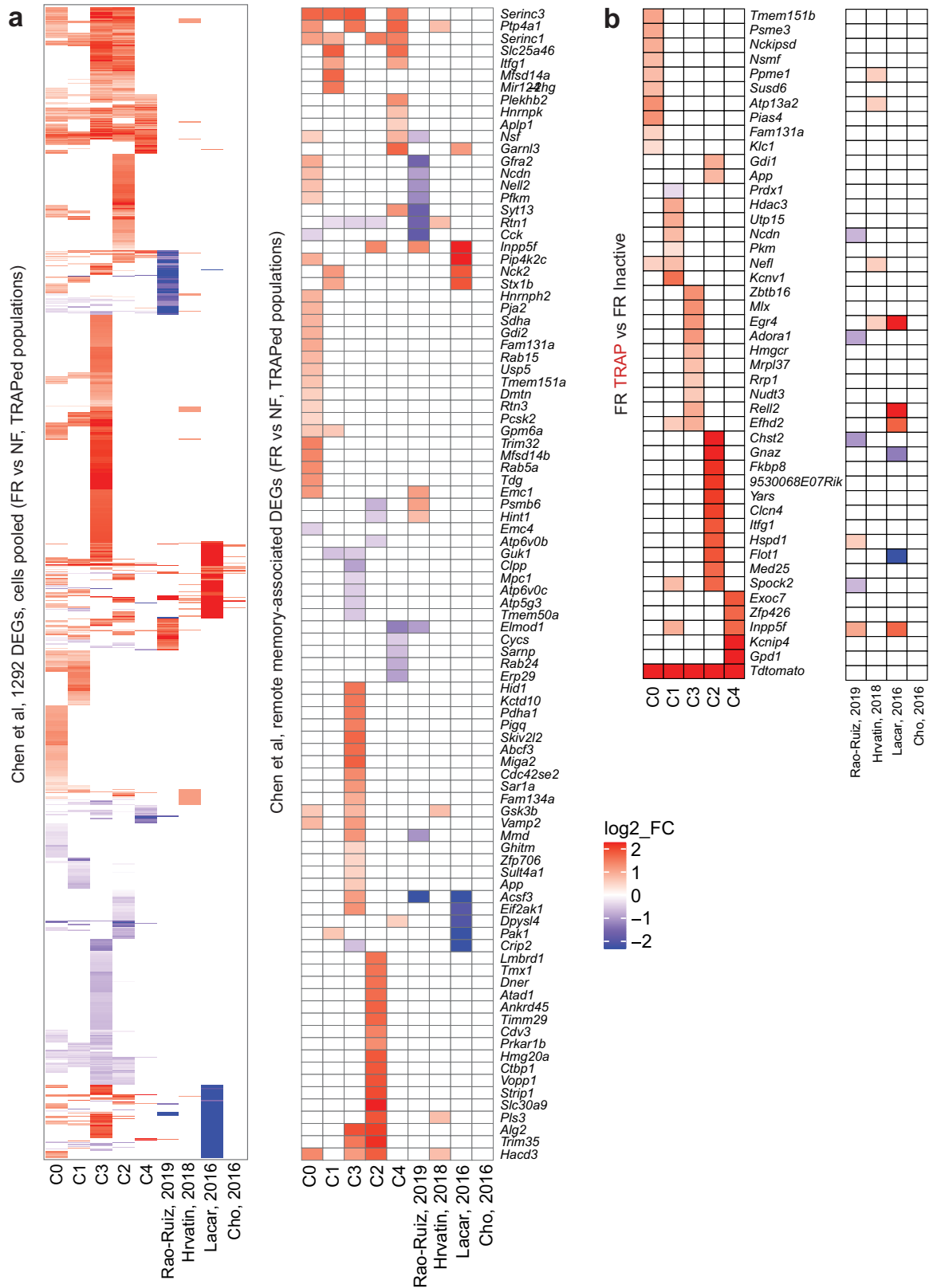
scRNA-seq (post-QC). No significant differences are found between FR and NF within either RNAscope or scRNA-methods, indicating no major changes in neuronal subtype composition of active populations in different training conditions. **c**, in situ hybridization of key remote-memory specific DEGs including *Stx1b* in *Rprm*<sup>+</sup>/*tdT*<sup>+</sup> cells, *Syt13* in *Tnfaip8l3*<sup>+</sup>/*tdT*<sup>+</sup> cells, *Vamp2* in *Tesc*<sup>+</sup>/*tdT*<sup>+</sup> cells. Scale bars, 100  $\mu$ m.



**Extended Data Fig. 8 | DEGs and potential cell-cell interactions in non-neuronal cells during memory consolidation.** **a**, Volcano plots of non-neuronal cell types when comparing cells in FR over NF nice. DEGs ( $\text{FDR} > 0.01, \log_2\text{FC} > 1$ ) are labelled in red, and exemplary DEGs (high  $\log_2\text{FC}$  and  $\log_{10}\text{FDR}$ ) are labelled in black. **b**, Number of non-neuronal cells collected in this study, for each cell type and experimental condition. **c**, Heat map of a subset of neuronal ligands and glial receptors that are found to be differentially perturbed upon memory consolidation. Only receptors and ligands which were

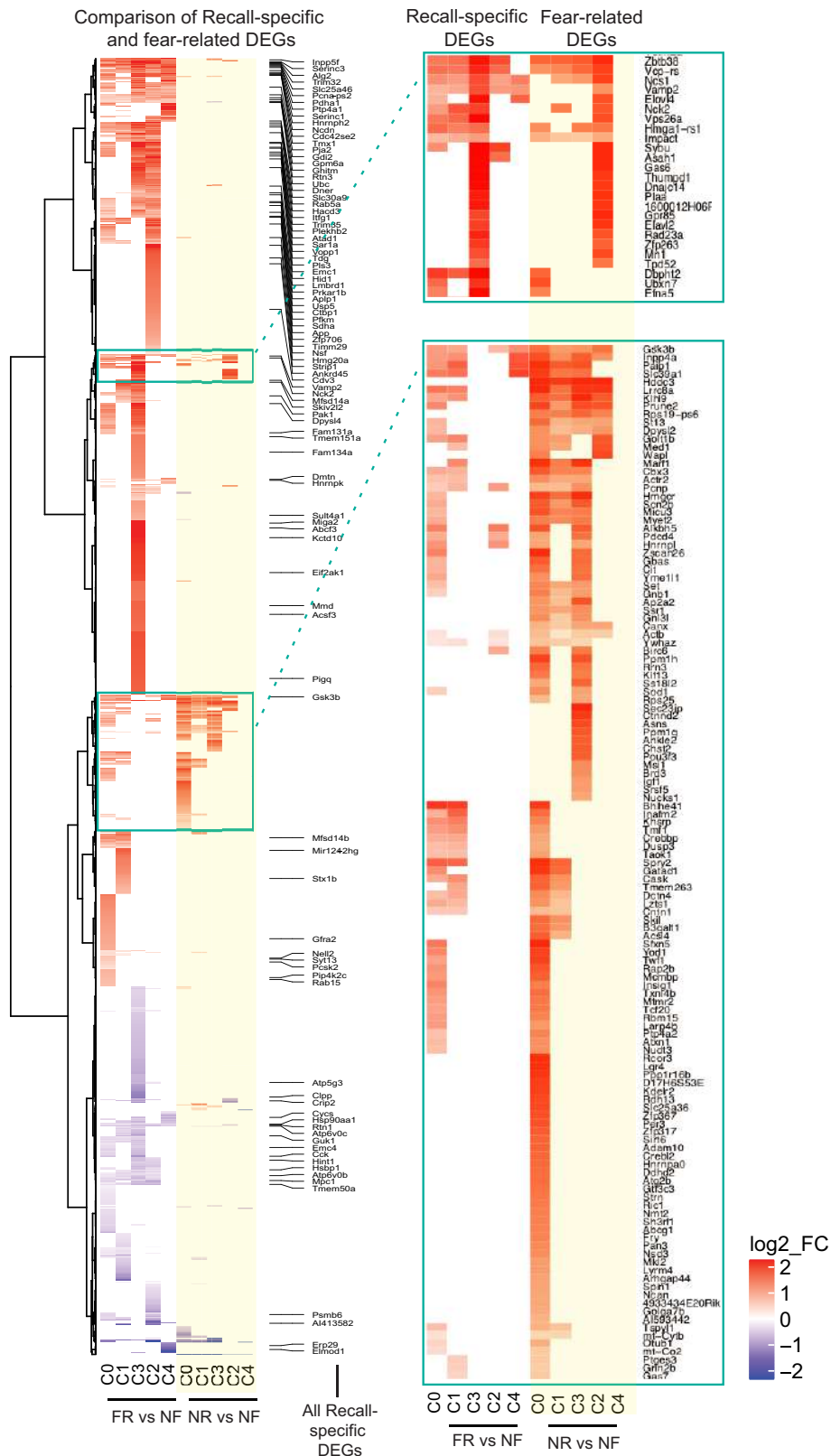
found to be (differentially) expressed are shown. **d**, Left, heat map of the  $\log_2\text{FC}$  of DEGs (FR over NF) in neurons that are classified as ligands. Middle and right, Sankey plot of known ligand-receptor pairs and heat map of the average scaled expression level of the corresponding receptors in each cell type. **e**, Left, heat map of the  $\log_2\text{FC}$  of DEGs (FR over NF) in neurons that are classified as receptors. Middle and right, Sankey plot of known ligand-receptor pairs and heat map of the average scaled expression level of the corresponding ligands in each cell type.





**Extended Data Fig. 9 | Comparison of remote-memory DEGs with previously published datasets of experience-dependent transcriptional activity.** **a**, Left, heat map of the log<sub>2</sub>FC of all 1,292 DEGs (FDR < 0.01, FR over NF, all cells pooled) in this manuscript, and their log<sub>2</sub>FC values in previously published datasets of experience-dependent DEGs in activated neurons during: recent fear memory retrieval<sup>24</sup>, associative fear-learning<sup>22</sup>, post-visual stimulus<sup>23</sup>, or novel environment exposure<sup>21</sup>. A value of zero log<sub>2</sub>FC indicates the gene was not differentially expressed in a dataset. Right, same as left, but

now DEGs are filtered down to the 'Recall-dependent DEG' set derived from this manuscript. Only genes differentially expressed in three out of four replicates are remaining. **b**, Left, log<sub>2</sub> fold change heat map of the recall-dependent DEGs between tdT<sup>+</sup> vs tdT<sup>-</sup> neurons in FR mice (genes are differentially expressed in >3/4 replicates) undergoing remote fear memory consolidation. Right, the log<sub>2</sub>FC values of these genes if they are found in previously published datasets of experience-dependent DEGs (see **a**). A value of zero log<sub>2</sub>FC indicates the gene was not differentially expressed in that dataset.



**Extended Data Fig. 10 | Comparison of remote-memory-specific DEGs and fear-experience-related DEGs.** Left, log<sub>2</sub> fold change heatmap of the union of DEGs from comparing FR vs NF (remote-memory specific) and NR vs NF (fear-related). Sustained transcriptional changes from the fear-experience

itself is shown in the yellow highlighted columns. Right, zoomed in view of the portion of the heat map within the green boxes where most fear experience-related DEGs are located.

## 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

- Sequences from the Nextseq or Novaseq were demultiplexed using bcl2fastq, and reads were aligned to the mm10 genome augmented with ERCC sequences, using STAR version 2.5.2b. Gene counts were made using HTSEQ version 0.6.1p1. All packages were called and run through a custom Snakemake pipeline  
- FreezeFrame (version 4) by Coulbourn Instruments

Data analysis

- R (version 4.0.2 Taking Off Again)  
- RStudio Version 1.1.456  
- Seurat v3 by the Satija Lab (<https://satijalab.org/seurat/>)  
- HOMER (version 4.8) (<http://homer.ucsd.edu/homer/motif/>)  
- GeneAnalytics (2020) (<https://geneanalytics.genecards.org/>)

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

- The accession number for the single cell RNA-sequencing data reported in this paper is GSE152632.

- STRING database (<https://string-db.org/>)
- Malacards: The human disease database (<https://www.malacards.org/>)
- RNA-seq data from Rao-Ruiz, 2019 (NCBI GEO Accession GSE129024)
- RNA-seq data from Hrvatin, 2017 (GSE102827)
- RNA-seq data from Lacar, 2016 (NCBI GEO Accession Number: GSE77067)
- RNA-seq data from Cho, 2016 (GSE85128)
- Receptor-ligand data from Ramilowski, 2015 ([http://fantom.gsc.riken.jp/5/suppl/Ramilowski\\_et\\_al\\_2015/](http://fantom.gsc.riken.jp/5/suppl/Ramilowski_et_al_2015/))

## 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](https://www.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      Sample sizes (both number of mouse replicates and number of cells sequences) were determined via a combination of factors including: the total expected size of the population of active neurons in the prefrontal cortex (~1.5%), (2) cost of sequencing and mice (3) expected margin of error and confidence levels
- Data exclusions      No data exclusions
- Replication      Replication was achieved by performing the behavioral training on 4 sets of mice, on 4 different independent days. Results were consistent in all replicates (4).
- Randomization      Mice were randomly grouped. Mice were fed and house in the same situation just prior to exposure to various behavioral trainings
- Blinding      Blinding was not performed during this study as the number of mice per training per biological replicate was too few to effectively allow this

## 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   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology                          |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                          |

### Methods

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Involved in the study                              |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq                  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging    |

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

- Laboratory animals      TRAP2 was generated in a 129Sv/SvJ background. For behavior experiments, they were backcrossed to C57Bl6/J for 3 generations. Mice were housed at room temperature and 40-60% humidity. We have described the usage of C57bl6J mice in the method part of the manuscript. Male mice at age between 6-7 weeks were used in this study.
- Wild animals      None
- Field-collected samples      None
- Ethics oversight      All animal procedures followed animal care guidelines approved by Stanford University's Administrative Panel on Laboratory Animal Care (APLAC)

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

## Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

Sample preparation

mPFC regions were micro-dissected from live vibratome sections (300  $\mu$ m thick) of the prefrontal cortex. Tissue pieces were enzymatically dissociated via a papain-based digestion system (Worthington, Cat # LK003150). Briefly, tissue chunks were incubated in 1mL of papain (containing L-cysteine and EDTA), DNase, and kynurenic acid for 1 hour at 37C and 5% CO<sub>2</sub>. After 10 min of incubation, tissues were triturated briefly with a P1000 wide bore pipette tip and returned. Cells were triturated another 4 times (~30 each) with a P200 pipette tip over the rest of the remaining incubation time. At room temperature, cell suspensions were centrifuged at 350g for 10 min, resuspended in 1mL of EBSS with 10% v/v ovomucoid inhibitor, 4.5% v/v DNase and 0.1% v/v kynurenic acid, and centrifuged again. Supernatant was removed and cells 1mL ACSF was added. ACSF was composed of: 1mM KCl, 7mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub>, 1.3 mM NaH<sub>2</sub>PO<sub>4</sub>, 110 mM choline chloride, 24mM NaHCO<sub>3</sub>, 1.3 mM Na Ascorbate, 20mM glucose and 0.6mM sodium pyruvate. Cells were passed through a 70  $\mu$ m cell strainer to remove debris. Hoescht stain was added (1:2000, Life Technologies, Cat #H3570) and incubated in the dark at 4C for 10 min. Samples were centrifuged (350g for 8 min at 4C) and resuspended in 0.5mL of ACSF and kept on ice for flow cytometry.

Instrument

Sony SH800

Software

Sony SH800

Cell population abundance

tdtomato+/ Hoescht+ cells were ~0.5-2% of all total events

Gating strategy

There was gating on BSC or FSC as size distribution of neurons/non-neuronal cells of interest is unknown. Only gating for a population that was double positive for Hoescht (above 1000 arbitrary fluorescent units) and tdTomato (above 10e4 arbitrary fluorescent units) was used (see Extended Data 2 for examples).

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.