

Representational drift in primary olfactory cortex

<https://doi.org/10.1038/s41586-021-03628-7>

Carl E. Schoonover^{1,3}✉, Sarah N. Ohashi^{1,2}, Richard Axel^{1,3}✉ & Andrew J. P. Fink^{1,3}✉

Received: 2 September 2020

Accepted: 11 May 2021

Published online: 9 June 2021

 Check for updates

Perceptual constancy requires the brain to maintain a stable representation of sensory input. In the olfactory system, activity in primary olfactory cortex (piriform cortex) is thought to determine odour identity^{1–5}. Here we present the results of electrophysiological recordings of single units maintained over weeks to examine the stability of odour-evoked responses in mouse piriform cortex. Although activity in piriform cortex could be used to discriminate between odorants at any moment in time, odour-evoked responses drifted over periods of days to weeks. The performance of a linear classifier trained on the first recording day approached chance levels after 32 days. Fear conditioning did not stabilize odour-evoked responses. Daily exposure to the same odorant slowed the rate of drift, but when exposure was halted the rate increased again. This demonstration of continuous drift poses the question of the role of piriform cortex in odour perception. This instability might reflect the unstructured connectivity of piriform cortex^{6–12}, and may be a property of other unstructured cortices.

In primary sensory neocortices, tuning to basic features such as retinotopy, somatotopy and tonotopy is stable^{13–19}. Responses may vary from day to day, but in the absence of perturbation or training, this variability is bounded and differences do not accumulate over time^{14,16}. In the olfactory system, sensory neurons that express the same receptor project with precision to spatially invariant glomeruli in the olfactory bulb²⁰. Each odorant evokes a distinct pattern of activity in the bulb that is stable over several months^{21,22}. Axonal projections from individual glomeruli discard this spatial patterning and diffusely innervate the piriform cortex without apparent structure^{6–9}. Thus, a second transformation occurs in piriform cortex, where individual odorants activate unique, distributed and readily distinguishable ensembles of neurons^{3,4,23}. Stimulus-evoked activity in piriform cortex is therefore thought to determine the identity of an odorant^{1–5}. If piriform cortex is the arbiter of odorant identity over the long term, then perceptual constancy requires that its activity, or a function of its activity, is stable.

interval tested was 3.5 μm (first quartile (Q1) = 2.1 μm , third quartile (Q3) = 5.6 μm), indicating that movement of the probe with respect to the neurons was negligible (Extended Data Fig. 1g–i). Moreover, single units were most similar to themselves across days than they were to any other single unit simultaneously recorded within a day (Extended Data Figs. 2g, 3). Thus, signals were sufficiently stable to permit isolation of single units across multiple weeks, with a total of 379 single units held over an interval of 32 days from 6 mice (Extended Data Fig. 1b). The number of single units isolated within a session (105 ± 29 (mean \pm s.d.), $n = 97$ recording sessions in 16 mice) exhibited only a weak dependence on implant age that was not significant over at least five months (Pearson's correlation $\rho = -0.1$, $P = 0.23$, Extended Data Fig. 1c). Although the electrode sites of our silicon probes sampled superficial and deep pyramidal neurons in anterior piriform cortex layers 2 and 3, it is likely that some of the single units isolated corresponded to semilunar cells or inhibitory interneurons.

Stable recordings in piriform cortex over weeks

We examined the stability of odour-evoked activity in mouse anterior piriform cortex in response to a panel of odorants over several weeks (Fig. 1a). We developed a methodology to follow the spiking of individual neurons over multiple days, using a silicon probe cemented in a fixed position with respect to the tissue (Extended Data Fig. 1a, Supplementary Discussion). We monitored both spike waveforms and non-waveform-based features on a daily basis to assess the stability of our recordings (Extended Data Figs. 1–3, Supplementary Discussion). The shapes of single unit waveforms, as well as their autocorrelograms, were preserved across days (Extended Data Figs. 1e, f, 2c–e). The median displacement of the neurons' estimated positions across the longest

Odour responses drift over time

The stability of single units over weeks permitted us to examine whether the odour responses of individual neurons are maintained over time. We recorded neural signals across a 32-day interval and measured the responses of single units to a panel of either four or eight neutral odorants, presented seven times per day every eight days (Fig. 1a). The panel of odorant molecules was selected to maximize diversity in functional groups and organoleptic properties. Mice were awake and head fixed but were not engaged in any task other than sampling the odour stimuli that were presented.

Within each day, we observed selective and reliable responses to the odorant panel (Fig. 1b, c, Extended Data Figs. 4, 5a–c). However,

¹Howard Hughes Medical Institute, Mortimer B. Zuckerman Mind Brain Behavior Institute, Department of Neuroscience, Columbia University, New York, NY, USA. ²Present address: Immunobiology Graduate Program, Yale School of Medicine, New Haven, CT, USA. ³These authors contributed equally: Carl E. Schoonover, Andrew J. P. Fink. ✉e-mail: ces2001@columbia.edu; ra27@columbia.edu; af2243@columbia.edu

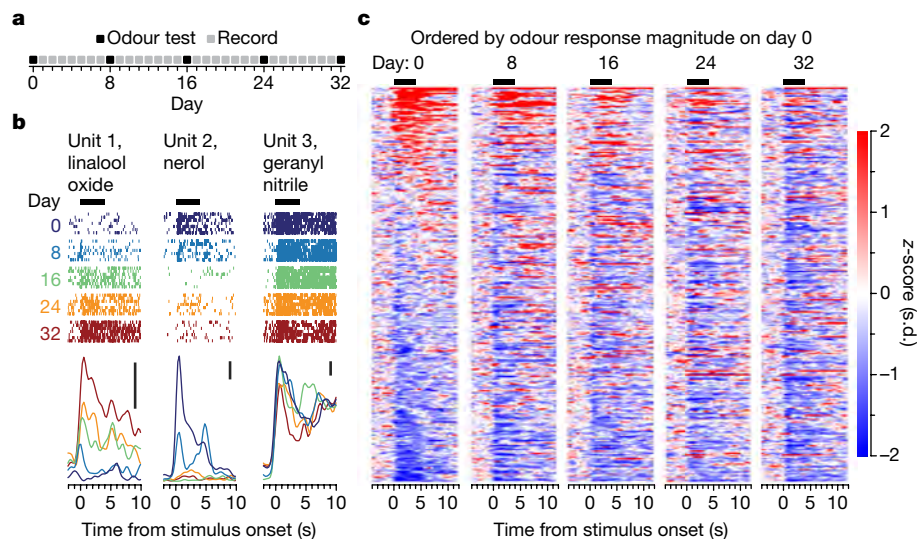


Fig. 1 | Odour-evoked activity in piriform cortex over 32 days. **a**, Experiment timeline. Black boxes, sessions in which odourants were administered (days 0, 8, 16, 24, and 32); grey boxes, sessions in which spontaneous activity was recorded without odourant administration. **b**, Odour-evoked activity on each odour test day for three example odour–unit pairs. Top, spike rasters (rows: 7 trials per day); bottom, peristimulus time histograms, superimposed across

days and coloured by day. Vertical black bars, 5 spikes per s firing rate. **c**, z-scored odour-evoked activity of 300 randomly selected odour–unit pairs, ordered by odour response magnitude on day 0. **b, c**, Horizontal black bars indicate 4-s odourant stimulus epoch. See Extended Data Fig. 4 for additional single-unit examples and population response maps.

there were gradual and pronounced changes in odour responses across days. Single units either gained or lost responsivity to a given odourant (Fig. 1b, c, Extended Data Figs. 4, 5d) and only rarely exhibited stable responses to the odourant panel over 32 days ($2.5 \pm 0.6\%$ of single units; Extended Data Fig. 5e, right). The probability of a single unit maintaining a response to an odour across a 32-day interval was $6.6 \pm 0.9\%$ (Extended Data Fig. 5d, e, left). The progressive changes in the responses of single units were due to continual alterations in odour-evoked responses (Fig. 2a–e), not to a global loss of responsiveness in piriform cortex (Fig. 2f, Extended Data Figs. 4b, 5a–c).

We quantified this drift by comparing the single-unit response magnitude for each odourant across days, and found that responses became increasingly dissimilar over time (Fig. 2a, within-day $R^2 = 0.94$, 8-day interval $R^2 = 0.52$, 16-day interval $R^2 = 0.31$, 24-day interval $R^2 = 0.22$, 32-day interval $R^2 = 0.08$, $n = 6$ mice). By contrast, changes in spontaneous firing rates across the same intervals were relatively small (Extended Data Fig. 2h) and neither baseline spontaneous firing rate nor changes in spontaneous firing rate predicted across-day changes in odour responses (Extended Data Fig. 2i).

Notably, variance in single-unit waveform stability metrics did not explain variance in odour response similarity across days (Extended Data Fig. 2j, k). Unstable tracking of single units across days therefore cannot account for changes in the odour responses of single units over time. Additionally, given that these changes in odour responses did not vary around a constant mean value but rather accumulated across days, we can rule out inconsistency in stimulus delivery across test days (Extended Data Fig. 6), as well as changes in the animals' internal state (for example, arousal) as the cause of the representational drift. Finally, it has been suggested that activity in piriform cortex during the early phase of the odour response epoch is sufficient to accurately establish its identity^{4,5,24}. However, when we isolated the first 200 ms of the odour response, using the local field potential as an indirect estimate of the sniff phase, we found that odour responses during this brief epoch exhibited drift (Extended Data Fig. 7, Supplementary Discussion).

We also measured changes in odour-evoked population responses. We constructed population vectors by binning the evoked response of each single unit into four 2-s windows beginning at odour onset—a binning that yielded maximal within-day linear classification (Extended

Data Fig. 5f–h). The performance of a linear classifier trained on earlier days and tested on later days deteriorated as a function of time between training and testing, approaching chance levels after 32 days (Fig. 2b). Across-day drift was symmetric in time (Extended Data Fig. 5i) and classification accuracy was not improved by concatenating multiple days in the training set (Extended Data Fig. 5j). Thus, it is not possible to establish which single units would be most informative about odour identity on day 32 using their responses across days 0–24. These data indicate that changes in odour-evoked responses in piriform cortex accumulate over time.

We next estimated the rate at which odour-evoked responses in piriform cortex drift. First, we computed the correlation between trial-averaged population odour responses across all pairs of recording days (Fig. 2c, left). Consistent with across-day changes in single-unit response magnitude (Fig. 2a), population vector correlations across days decreased as a function of interval (Fig. 2c, right). We then computed the angle between trial-averaged population vectors (Extended Data Fig. 5k–m). An exponential fit to these measurements gave a time constant of 28.3 days with an asymptote of 89.4° (an angle of 90° indicates complete decorrelation; Extended Data Fig. 5l, right). We estimated the rate of drift in two ways. First, we computed the rate of change of the exponential fit over the 32-day interval, finding a mean rate of 1.0° per day. Second, we estimated the drift rate between every pair of days. The angle between days reflects a combination of both across-day drift in odour-evoked responses and within-day trial-to-trial variability²⁵. We therefore corrected for within-day variability by computing a corrected angle, subtracting within-day variability from across-day variability. The corrected angle increased significantly as a function of interval (Fig. 2d). Finally, we divided the corrected angle by the time interval over which it was computed (Extended Data Fig. 5k). This estimate produced a drift rate of $1.3 \pm 1.2^\circ$ per day across all intervals (mean \pm s.d., $n = 6$ mice; Fig. 2e), with comparable rates for all odourants used (Extended Data Fig. 6e, f), and in accordance with the estimate based on the exponential fit (Extended Data Fig. 5l, right). Thus, piriform cortex exhibits representational drift: changes in odour-evoked responses accrete continuously over time and trend towards complete decorrelation within weeks, well under the lifespan of a mouse²⁶.

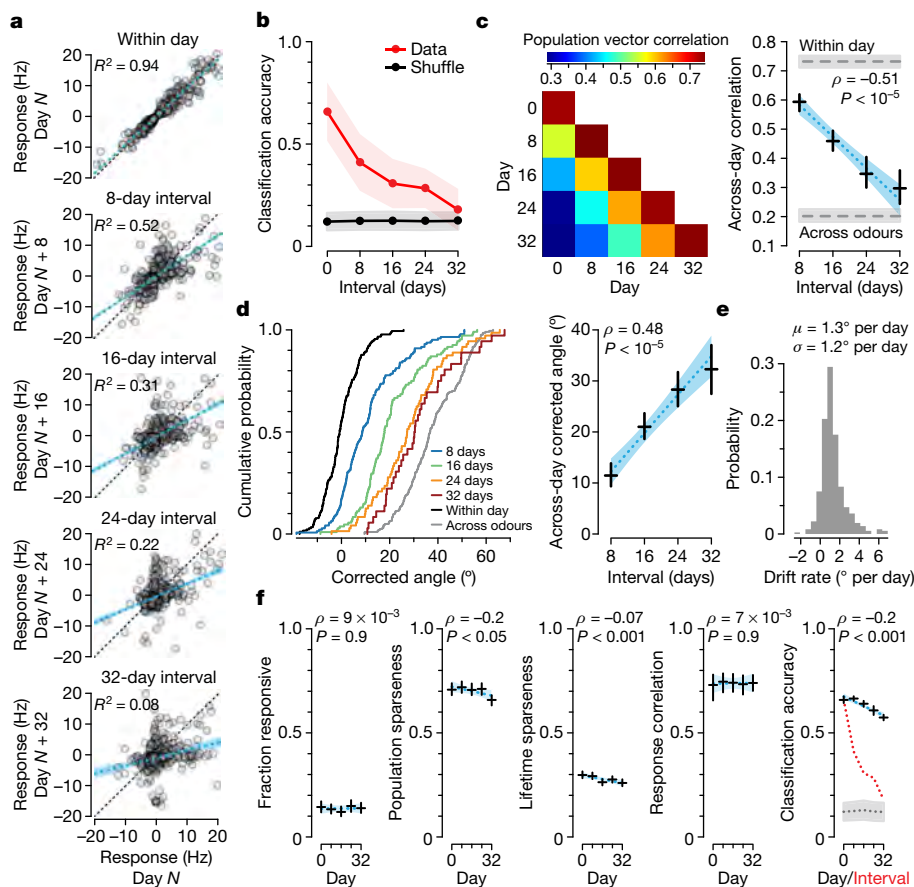


Fig. 2 | Drift of odour responses across days despite conserved statistics within each day. **a**, Regression of baseline-subtracted odour response magnitude within day (even versus odd trials) and across 8- to 32-day intervals (later versus earlier days) for, respectively, $n = 2,353$, $n = 2,843$, $n = 1,879$, $n = 1,170$, and $n = 577$ odour-unit pairs that were responsive on at least one of the two comparisons (577 random odour-unit pairs plotted). Black dashed line, unity. **b**, Classification accuracy (eight-way, support vector machine (SVM), linear kernel, L2 regularization, single-trial population vectors, random subsets of 41 single units) within-day (leave-one-out cross-validation) and across intervals (model trained on earlier and tested on later days using all responses). Mean \pm s.d. **c**, **d**, Pearson's correlations (**c**) and corrected angles (**d**) within-day and across 8- to 32-day intervals ($n = 360$, $n = 144$, $n = 108$, $n = 72$ and $n = 36$ trial-averaged population vector pairs, respectively). **c**, Right, grey

dashed lines, mean \pm 95% CI within-day (top) and across odours across 32 days (bottom). **d**, Left, cumulative distributions of corrected angles across days, within-day, and across-odour (across 32 days). **e**, Distribution of drift rates for all 360 across-day population vector pairs. **f**, Within-day population statistics. Fraction of responsive odour-unit pairs, population and lifetime sparseness, days 0–32: $n = 584$, $n = 635$, $n = 629$, $n = 593$ and $n = 545$ single units, respectively. Within-day correlations: $n = 72$ trial-averaged population vector pairs per day (even versus odd trials). Classification accuracy as in **b** (within-day, leave-one-out cross-validation); red dotted line, mean across-day performance, replotted; grey, shuffled (mean \pm s.d.). All panels: black crosses, mean \pm 95% CI; blue dotted line, linear regression; blue shading, 95% CI. Classification performed on the three mice that were presented with an eight-odorant panel; otherwise, $n = 6$ mice.

The changes in odour-evoked responses we have observed could be attributed to gross changes in the population response. We found, however, that the basic response properties of the piriform cortex population changed only marginally over time (Fig. 2f, Extended Data Figs. 4b (right), 5a–c). The fraction of single units that responded to an odorant stimulus was largely unaltered across the 32-day interval (Fig. 2f; $R^2 = 8.0 \times 10^{-5}$). Moreover, population sparseness ($R^2 = 3.1 \times 10^{-2}$), lifetime sparseness ($R^2 = 4.4 \times 10^{-3}$), within-day variability ($R^2 = 5.3 \times 10^{-5}$), and within-day classifier performance ($R^2 = 5.2 \times 10^{-2}$) varied little over this period (Fig. 2f). Within-day classifier performance exhibited a modest deterioration (Fig. 2f, right, black crosses), which may contribute a small fraction of the decrease in across-day classification accuracy we measure over time (Fig. 2f, right, red dotted line); however, for all intervals, across-day performance was significantly lower than within-day performance ($P < 4.8 \times 10^{-78}$ for all comparisons). These observations show that drift in odour-evoked responses is not the result of a global diminution in responsivity and occurs against a background of grossly stable population response properties.

Odour response geometry drifts over time

Regions downstream of piriform cortex could in principle compensate for drift in the response of neurons over time if the geometry of odour-evoked population activity were conserved. It has recently been shown that the structure of piriform cortex responses to highly similar odorant molecules is conserved across individuals, but that no such structure is apparent for relatively dissimilar stimuli, such as those we used²⁷. To determine whether the geometry of odour responses in piriform cortex is conserved despite drift, we defined the shape of a set of odour responses, where each vertex is the mean population response to each odorant (Fig. 3a). We described the geometry of this object by measuring the angles formed between pairs of connected edges (Fig. 3a, b). We then computed the extent to which those angles varied across days. This metric is insensitive to shape-preserving linear transformations such as translation, rotation, and uniform expansion. We found that changes in edge angles accumulated significantly across days (Fig. 3c). Thus, the geometry of population activity evoked by relatively dissimilar

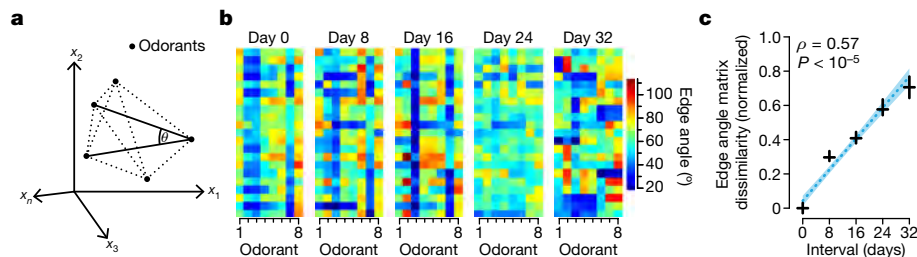


Fig. 3 | Odour response geometry drifts over time. **a**, Diagram illustrating edge angles. Each black dot symbolizes the population response to an individual odorant: the trial-averaged population response vector in n -dimensional neural activity space. Dotted lines, edges between population responses to different odorants. An example edge angle (θ) is illustrated between two connected edges (solid lines). **b**, The 21 edge angles (rows) for each of the 8 test odorants (columns) measured on each of the 5 odour test days from

a representative animal. **c**, Edge angle matrix dissimilarity, normalized by scaling results from each subject between 0 (mean within-day matrix dissimilarity) and 1 (mean shuffled matrix dissimilarity). Within-day, 525 edge angle matrix pairs and across-day, 1,050 edge angle matrix pairs from the three mice that were presented with an eight-odorant panel. Within-day versus across-day dissimilarity for all intervals $P < 6.7 \times 10^{-42}$ (Wilcoxon rank-sum). Black crosses, mean \pm 95% CI; blue dotted line, linear regression; blue shading, 95% CI.

odorant molecules is not conserved, but rather changes gradually over time.

We used numerous additional measures of odour response geometry, and invariably observed drift (Extended Data Figs. 8, 9, Supplementary Discussion). These include a wide variety of temporal binnings of the odour response (Extended Data Fig. 8d, e), measures that rely on correlation matrices instead of edge angle matrices (Extended Data Fig. 8a–c), and estimates computed in low-dimensional projections of the data (Extended Data Fig. 8c, second and forth panels). This drift was also observed when analysis was restricted to stimulus-coding dimensions of the evoked response²⁸ (Extended Data Fig. 9). These analyses do not depend on longitudinal observation of individual neurons across days and therefore provide independent evidence that odour-evoked responses in piriform cortex change over time.

Drift persists after odour conditioning

We next investigated whether evoked responses to a behaviourally salient odorant might exhibit greater stability. We placed implanted mice in a conditioning chamber and presented conditioned stimuli: one paired with foot shock (CS+) and one presented without shock (CS-) eight times each over the course of a single session (Fig. 4a, Extended Data Fig. 10a). The following day we used a virtual burrow assay²⁹ to confirm that animals responded selectively to the CS+ but not to the CS- or to other odorants that were never presented in the conditioning chamber. Animals showed selective ingress (aversion) to the CS+ across the duration of the protocol (Extended Data Fig. 10b).

Simultaneously acquired recordings showed that evoked population responses to conditioned stimuli drifted at a rate comparable to that of neutral odours (Fig. 4b; CS+, 1.2 (0.7–1.8)° per day (mean (95% confidence interval (CI))); CS-, 0.9 (0.7–1.7)° per day; neutral, 1.0 (0.8–1.0)° per day; for all comparisons $P \geq 0.61$, Wilcoxon rank-sum, $n = 5$ mice). Moreover, the across-day correlations of the single-unit response magnitude for each odorant were not greater for CS+ than for CS- or neutral odorants and there was no significant difference in corrected angles (Extended Data Fig. 10c). Thus, fear conditioning does not appear to stabilize odour-evoked responses in piriform cortex: we observed stable behaviour in spite of drifting neural activity.

Frequent experience reduces drift

These data pose the question of the role of piriform cortex in odour perception. Representational drift in piriform cortex may reflect a learning system that continuously updates and overwrites itself³⁰. We made two observations that support this view. We presented a panel of odorants daily across a 32-day interval. Beginning on day 16, we also

presented a set of unfamiliar odorants at 8-day intervals and compared the stability of the evoked responses to the familiar and unfamiliar odorants (Fig. 4c, top, cohort A). We found that the rate of drift of the evoked responses to familiar odorants was less than half that of unfamiliar odorants (Fig. 4d, solid lines; familiar odours, 0.4 (0.2–0.6)° per day (mean (95% CI)) versus unfamiliar odours, 0.9 (0.7–1.2)° per day; $P = 4.3 \times 10^{-4}$, Wilcoxon rank-sum, $n = 5$ mice). Moreover, daily exposure resulted in higher correlations in single-unit response magnitude over 16 days (Extended Data Fig. 10d, e), consistently higher classification performance across time (Extended Data Fig. 10f), and consistently lower corrected angles (Extended Data Fig. 10g, h). Thus, the drift rate

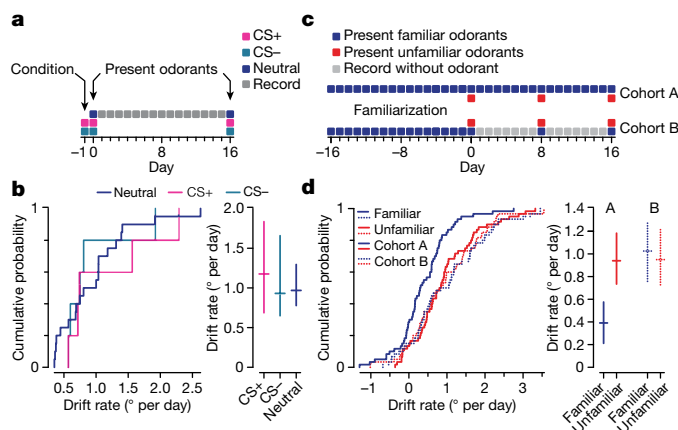


Fig. 4 | Frequent experience, but not fear conditioning, reduces drift rate.

a, Conditioning timeline. Day -1: present one odorant paired with shock (CS+) and one without shock (CS-) in a conditioning chamber. Days 0 and 16: administer conditioned odorants (CS+ and CS-) and four neutral odorants to head-fixed mouse while recording neural signals and measuring behavioural responses in a virtual burrow assay²⁹. Days 1–15: record neural signals without test odorant administration. **b**, Cumulative distributions (left) and mean \pm 95% CI (right) of drift rates for all three classes of stimulus. CS+ and CS-, $n = 5$ population vector pairs; neutral odorants, $n = 20$ population vector pairs (5 mice). For all comparisons, $P > 0.05$ (Wilcoxon rank-sum). **c**, Odour frequency timeline. All mice were first presented with a panel of four neutral odorants daily (familiar odorants) over 16 days (familiarization; days -16 to -1). Then a panel of unfamiliar odorants was presented at 8-day intervals (days 0, 8, 16). Between days 0 and 16, cohort A (5 mice) was presented familiar odorants daily and cohort B (5 mice) was presented familiar odorants at 8-day intervals. Neural signals were recorded daily. **d**, Cumulative distributions (left) and mean \pm 95% CI (right) of drift rates for cohorts A and B. Cohort A familiar odorants drift rate versus all others, $P < 9.4 \times 10^{-4}$; all other comparisons, $P > 0.05$; Wilcoxon rank-sum, $n = 60$ trial-averaged population vector pairs.

exhibits history dependence: continual experience with an odourant enhances the stability of piriform cortex's evoked response to it.

Notably, this observation provides independent evidence for our ability to follow a fixed population of single units over time. We observed rapidly drifting evoked responses to unfamiliar odours and slowly drifting evoked responses to familiar odours on separate trials within the same recording session of the same population of neurons. Thus, it is unlikely that the changes in odour-evoked responses we have observed are caused by a failure to follow the same population of neurons.

Theoretical models predict that, in a highly plastic learning system, the stable representations that encode memories will be overwritten unless the circuit has a mechanism to store them for the long term³⁰. We therefore investigated whether piriform cortex has a mechanism to retain stable odour-evoked responses after daily experience ceases. We presented a panel of odourants every day for sixteen days. This panel, along with a set of unfamiliar odourants, was then presented only at 8-day intervals (Fig. 4c, bottom, cohort B). Once daily odour exposure ceased, evoked responses to the familiar and unfamiliar stimuli drifted at similar rates (Fig. 4d, dotted lines; familiar odours, $1.0 (0.8-1.3)^\circ$ per day versus unfamiliar odours, $1.0 (0.7-1.2)^\circ$ per day; $P=0.81$, Wilcoxon rank-sum, $n=5$ mice). Thus, daily exposure slows representational drift, but without ongoing exposure drift rate increases once again.

Discussion

We have shown that odour-evoked activity in the mouse anterior piriform cortex exhibits rapid and cumulative reorganization over time. If piriform cortex facilitates stimulus identification over the long term, there must be a function of piriform cortex activity whose output is stable. Regions downstream of piriform cortex could conceivably compensate for representational drift³¹. For example, replay of previously experienced odourants could regularly update the readout³². This would require continual replay without continual odour exposure over the life of the organism. Alternatively, a downstream reader could leverage some invariant geometry in odour-evoked population activity, even as the responses of individual neurons drift over time³³. Although we failed to uncover evidence for such a geometry, we note that our analysis does not rule out the existence of an invariant structure that lies on a nonlinear manifold. A third possibility is that the small subset (2.5%) of single units that show preserved odour-evoked responses over 32 days are responsible for encoding odour identity. This would leave unresolved the question of what role is played by the vast majority of drifting piriform cortex output. Finally, our observations do not rule out the possibility that activity in piriform cortex encodes odour identity along with other unknown variables in a format not readily captured by simply examining responses to odour stimuli over time³⁴.

However, piriform cortex may not be the ultimate arbiter of odour identity. Piriform cortex is not the sole recipient of olfactory sensory information. Stable odour-evoked responses in the olfactory bulb^{21,22} are broadcast to diverse regions that may link sensation with valence and action: the anterior olfactory nucleus, olfactory tubercle, cortical amygdala and lateral entorhinal cortex⁶.

If piriform cortex does not determine odour identity over the long term, what function may be implemented by this structure? The three principal observations in this study—representational drift, history-dependent stabilization, and subsequent drift of previously stabilized odour-evoked responses—are consistent with a model in which piriform cortex functions as a fast learning system that continually learns and continually overwrites itself. Piriform cortex may operate as a 'scratch pad', rapidly encoding memory traces upon limited odourant exposure. Lacking a mechanism to stabilize these memory traces, however, piriform cortex cannot store those memories over the long term³⁰. This may explain why the evoked odour response changes when an animal is not experiencing the odourant. In a fast learning system,

experience with one stimulus will drive plastic changes that alter the evoked responses to other stimuli. In our experiments, ongoing experience with the odourants of the home cage, for example, may result in cumulative changes in the evoked responses to our test odourants. In this model, representational drift in piriform cortex is a consequence of continual learning and concomitant overwriting.

Learned representations in piriform cortex that reflect recent olfactory experience may be only transiently useful to the animal, because they are continuously updated as the environment changes. Alternatively, learned representations in piriform cortex may be transferred to a more stable region downstream for long-term storage. Models that posit the consolidation of memory traces from a highly plastic, unstable network to a less plastic, stable network computationally outperform single-stage learning systems by combining both flexibility and long-term memory³⁵. This framework has been proposed to account for the initial formation of episodic memories in hippocampus and later consolidation in neocortex³⁶. Piriform cortex may therefore operate in conjunction with a slow learning system downstream to support long-term access to stored representations. This hypothetical slow learning system must receive input from the olfactory bulb, where odour-evoked responses are stable^{21,22}, as well as from piriform cortex, where they are drifting. This third region might then accommodate the identification of an olfactory stimulus while incorporating learned information from piriform cortex. Thus, in this model, although day-to-day activity in piriform cortex does not provide the basis for identifying odourants, learning in piriform cortex upon initial exposure to stimuli, followed by consolidation to a slow learning system downstream, contributes to the encoding of stimulus identity.

The stability of neural activity varies across brain regions. Representational drift has been previously reported in hippocampus CA1³⁷⁻⁴⁰, motor cortex⁴¹, and posterior parietal cortex⁴². In primary sensory neocortices, however, in the absence of a training paradigm or a gross perturbation, responses to basic stimulus features (such as retinotopy, tonotopy or somatotopy) vary across days but tuning remains centred around a fixed mean^{14,16,17,19,43-45}. Moreover, tuning changes induced in sensory neocortex by training or perturbation have been found to reverse after the intervention is discontinued^{14,18}. By contrast, our data show that in primary olfactory cortex changes in odour-evoked responses accrete continuously: the evoked activity at every successive time point is increasingly dissimilar to the first.

What conditions promote drifting versus stable neural activity? The anatomical organization of most sensory cortices may set a bound on changes in stimulus tuning to basic features. In primary visual cortex, for example, inputs from thalamus obey ordered retinotopic organization⁴⁶. The weakening or loss of one synapse onto a visual cortical neuron is therefore likely to be matched by the strengthening or gain of a synapse tuned to a similar retinotopic region, limiting representational drift with respect to the simple artificial stimuli commonly used in most studies of visual cortex. However, the organization of piriform cortex differs from that of sensory neocortex: piriform cortex receives distributed⁶⁻⁹ and unstructured^{3,10} inputs from olfactory bulb, and recurrent cortical connections are broadly distributed and appear to lack topographic organization^{11,12}. Thus, representational drift in piriform cortex may be explained by ongoing plasticity in an unstructured network. This same reasoning may account for why evoked responses to complex naturalistic stimuli appear to drift in visual cortex⁴⁷, even though those to simple artificial stimuli, which are tethered by anatomical constraints, do not. The structured connectivity observed in most early sensory regions may not be present in higher centres, suggesting that representational drift may be a more pervasive property of cortex.

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/s41586-021-03628-7>.

1. Haberly, L. B. Single unit responses to odor in the prepiriform cortex of the rat. *Brain Res.* **12**, 481–484 (1969).
2. Kadohisa, M. & Wilson, D. A. Separate encoding of identity and similarity of complex familiar odors in piriform cortex. *Proc. Natl Acad. Sci. USA* **103**, 15206–15211 (2006).
3. Stettler, D. D. & Axel, R. Representations of odor in the piriform cortex. *Neuron* **63**, 854–864 (2009).
4. Miura, K., Mainen, Z. F. & Uchida, N. Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. *Neuron* **74**, 1087–1098 (2012).
5. Bolding, K. A. & Franks, K. M. Recurrent cortical circuits implement concentration-invariant odor coding. *Science* **361**, eaat6904 (2018).
6. Haberly, L. B. & Price, J. L. The axonal projection patterns of the mitral and tufted cells of the olfactory bulb in the rat. *Brain Res.* **129**, 152–157 (1977).
7. Sosulski, D. L., Bloom, M. L., Cutforth, T., Axel, R. & Datta, S. R. Distinct representations of olfactory information in different cortical centres. *Nature* **472**, 213–216 (2011).
8. Miyamichi, K. et al. Cortical representations of olfactory input by trans-synaptic tracing. *Nature* **472**, 191–196 (2011).
9. Ghosh, S. et al. Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature* **472**, 217–220 (2011).
10. Davison, I. G. & Ehlers, M. D. Neural circuit mechanisms for pattern detection and feature combination in olfactory cortex. *Neuron* **70**, 82–94 (2011).
11. Johnson, D. M., Illig, K. R., Behan, M. & Haberly, L. B. New features of connectivity in piriform cortex visualized by intracellular injection of pyramidal cells suggest that “primary” olfactory cortex functions like “association” cortex in other sensory systems. *J. Neurosci.* **20**, 6974–6982 (2000).
12. Franks, K. M. et al. Recurrent circuitry dynamically shapes the activation of piriform cortex. *Neuron* **72**, 49–56 (2011).
13. Gilbert, C. D. & Wiesel, T. N. Receptive field dynamics in adult primary visual cortex. *Nature* **356**, 150–152 (1992).
14. Rose, T., Jaepel, J., Hübener, M. & Bonhoeffer, T. Cell-specific restoration of stimulus preference after monocular deprivation in the visual cortex. *Science* **352**, 1319–1322 (2016).
15. Clark, S. A., Allard, T., Jenkins, W. M. & Merzenich, M. M. Receptive fields in the body-surface map in adult cortex defined by temporally correlated inputs. *Nature* **332**, 444–445 (1988).
16. Margolis, D. J. et al. Reorganization of cortical population activity imaged throughout long-term sensory deprivation. *Nat. Neurosci.* **15**, 1539–1546 (2012).
17. Mayrhofer, J. M., Haiss, F., Helmchen, F. & Weber, B. Sparse, reliable, and long-term stable representation of periodic whisker deflections in the mouse barrel cortex. *Neuroimage* **115**, 52–63 (2015).
18. Weinberger, N. M., Javid, R. & Lapan, B. Long-term retention of learning-induced receptive-field plasticity in the auditory cortex. *Proc. Natl Acad. Sci. USA* **90**, 2394–2398 (1993).
19. Kato, H. K., Gillet, S. N. & Isaacson, J. S. Flexible sensory representations in auditory cortex driven by behavioral relevance. *Neuron* **88**, 1027–1039 (2015).
20. Mombaerts, P. et al. Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
21. Bhalla, U. S. & Bower, J. M. Multiday recordings from olfactory bulb neurons in awake freely moving rats: spatially and temporally organized variability in odorant response properties. *J. Comput. Neurosci.* **4**, 221–256 (1997).
22. Kato, H. K., Chu, M. W., Isaacson, J. S. & Komiyama, T. Dynamic sensory representations in the olfactory bulb: modulation by wakefulness and experience. *Neuron* **76**, 962–975 (2012).
23. Poo, C. & Isaacson, J. S. Odor representations in olfactory cortex: “sparse” coding, global inhibition, and oscillations. *Neuron* **62**, 850–861 (2009).
24. Wilson, C. D., Serrano, G. O., Koulakov, A. A. & Rinberg, D. A primacy code for odor identity. *Nat. Commun.* **8**, 1477 (2017).
25. Chestek, C. A. et al. Single-neuron stability during repeated reaching in macaque premotor cortex. *J. Neurosci.* **27**, 10742–10750 (2007).
26. Flurkey, K., Curren, J. M. & Harrison, D. E. in *The Mouse in Biomedical Research* 637–672 (Elsevier, 2007).
27. Pashkovski, S. L. et al. Structure and flexibility in cortical representations of odour space. *Nature* **583**, 253–258 (2020).
28. Kobak, D. et al. Demixed principal component analysis of neural population data. *eLife* **5**, e10989 (2016).
29. Fink, A. J., Axel, R. & Schoonover, C. E. A virtual burrow assay for head-fixed mice measures habituation, discrimination, exploration and avoidance without training. *eLife* **8**, e45658 (2019).
30. Fusi, S., Drew, P. J. & Abbott, L. F. Cascade models of synaptically stored memories. *Neuron* **45**, 599–611 (2005).
31. Rule, M. E., O’Leary, T. & Harvey, C. D. Causes and consequences of representational drift. *Curr. Opin. Neurobiol.* **58**, 141–147 (2019).
32. Káli, S. & Dayan, P. Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nat. Neurosci.* **7**, 286–294 (2004).
33. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).
34. Brette, R. Is coding a relevant metaphor for the brain? *Behav. Brain Sci.* **42**, e215 (2018).
35. Roxin, A. & Fusi, S. Efficient partitioning of memory systems and its importance for memory consolidation. *PLOS Comput. Biol.* **9**, e1003146 (2013).
36. McClelland, J. L., McNaughton, B. L. & O’Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
37. Kentros, C. G., Agnihotri, N. T., Streater, S., Hawkins, R. D. & Kandel, E. R. Increased attention to spatial context increases both place field stability and spatial memory. *Neuron* **42**, 283–295 (2004).
38. Mankin, E. A. et al. Neuronal code for extended time in the hippocampus. *Proc. Natl Acad. Sci. USA* **109**, 19462–19467 (2012).
39. Rubin, A., Geva, N., Sheintuch, L. & Ziv, Y. Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife* **4**, e12247 (2015).
40. Lee, J. S., Briguglio, J. J., Cohen, J. D., Romani, S. & Lee, A. K. The statistical structure of the hippocampal code for space as a function of time, context, and value. *Cell* **183**, 620–635. e22 (2020).
41. Rokni, U., Richardson, A. G., Bizzi, E. & Seung, H. S. Motor learning with unstable neural representations. *Neuron* **54**, 653–666 (2007).
42. Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N. & Harvey, C. D. Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell* **170**, 986–999. e16 (2017).
43. Tolias, A. S. et al. Recording chronically from the same neurons in awake, behaving primates. *J. Neurophysiol.* **98**, 3780–3790 (2007).
44. Mank, M. et al. A genetically encoded calcium indicator for chronic in vivo two-photon imaging. *Nat. Methods* **5**, 805–811 (2008).
45. Jeon, B. B., Swain, A. D., Good, J. T., Chase, S. M. & Kuhlman, S. J. Feature selectivity is stable in primary visual cortex across a range of spatial frequencies. *Sci. Rep.* **8**, 15288 (2018).
46. White, E. L. Thalamocortical synaptic relations: a review with emphasis on the projections of specific thalamic nuclei to the primary sensory areas of the neocortex. *Brain Res.* **180**, 275–311 (1979).
47. Marks, T. D. & Goard, M. J. Stimulus-dependent representational drift in primary visual cortex. Preprint at <https://doi.org/10.1101/2020.12.10.420620> (2020).

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 2021

Methods

Ethical compliance

All procedures were approved by the Columbia University Institutional Animal Care and Use Committee (protocol AC-AAAT5466) and were performed in compliance with the ethical regulations of Columbia University as well as the Guide for Animal Care and Use of Laboratory Animals.

Animals

We used 10–17-week-old (12 ± 2 weeks, mean \pm s.d.; $n = 45$ mice) male C57BL/6J mice (Jackson laboratories). Sample size was not predetermined; we established that our sample sizes were sufficient from the size and statistical significance of the effects. Randomization is not relevant to this study as comparisons were not made across groups. Blinding is not relevant to this study as manual curation of single-unit templates was done without knowledge of stimulus responsiveness, and behaviour was analysed automatically without manual scoring. Mice were group-housed before silicon probe implantation and housed singly thereafter.

Stereotactic targeting and head-plate attachment surgery

Animals were anaesthetized with isoflurane (3% induction, 1.5–2% maintenance), placed on a feedback-controlled heating pad (Fine Science Tools) and secured within a stereotactic frame (Model 1900-B Head Holder Assembly with Sagittal Plate from a Model 1900 Stereotaxic Alignment Instrument, David Kopf Instruments) affixed to an air table (TMC). All stereotactic targeting was performed using a separate motorized manipulator (Scientifica) mounted on the air table under a Leica M80 dissection stereo microscope (Leica Microsystems), with images captured using the Leica Application Suite. Carprofen (5 mg/kg) was administered via subcutaneous injection as a preoperative analgesic and bupivacaine (2 mg/kg) was delivered underneath the scalp to numb the area of the incision. The skull was exposed and cleaned with sterile cotton swabs.

To reliably target the anterior piriform cortex (APCx), we used a series of stereotactic procedures to reduce targeting variability across animals. The skull was aligned by first adjusting the angle of the head, dorsal tilt (pitch), such that lambda and the rostral confluence of the sinus (RCS) both lay in the horizontal plane of the motorized manipulator. The lateral edges of the frontal and parietal bones were mapped at 500- μ m increments from 1,000 μ m posterior to RCS (pRCS) to 4,500 μ m pRCS. These measurements were compared to mean values and the angle of the head, coronal tilt (roll) and sagittal tilt (yaw) were adjusted using the head-holder assembly to minimize differences from a mean reference mouse ($n = 164$ mice). The midline was defined as the mean midpoint between the lateral edges of the frontal bones at 1,000 μ m pRCS, 1,500 μ m pRCS, and 2,000 μ m pRCS. If the difference between any of these landmarks exceeded two standard deviations from mean values, mice were not used for probe implantation.

Following head alignment, the area surrounding the target location for APCx was marked with hatch marks scored with a scalpel and then labelled with oil-based ink. The skull was covered in a thin layer of cyanoacrylate adhesive (Krazy Glue, Elmer's Products). A glass capillary was then positioned at the target for APCx (2,250 μ m lateral to the midline, 1,150 μ m pRCS) and a camera attached to the dissection microscope was used to take a photograph of the capillary position for positioning the silicon probe on the day of implantation.

A portion of skin adjacent to the incision was removed and the area of skin and muscle surrounding the skull was temporarily covered with silicone elastomer (Kwik-Cast, World Precision Instruments) to protect it during application of cement. A coating of adhesive luting cement (C&B-Metabond, Parkell) was applied atop the layer of cyanoacrylate adhesive, covering the skull except for the probe target area. A titanium head plate (27.4 mm \times 9.0 mm \times 0.8 mm, G. Johnson, Columbia

University) was lowered onto the skull using the micromanipulator. The head plate was mounted on a custom adaptor (G. Johnson, Columbia University) that positioned it within the horizontal plane of the micromanipulator. The headplate was secured with additional applications of the luting cement, again ensuring that cement covered the probe target area. A well was then formed using dental acrylic (Tru-Pak Orthodontic Acrylic, Stoelting) such that the entire edge of the opening of the headplate was in contact with cement and the probe insertion site was within this well, separated from any exposed tissue.

Two bone screws, each connected to 15-mm-long 36 American Wire Gauge wire (Phoenix Wire) soldered to an Amphenol pin (A-M Systems) trimmed to 1 mm, were then inserted bilaterally above the cerebellum opposite the midline to serve as ground and reference electrodes. The bone screws were then sealed in place with the luting cement and the wires were directed to the anterior portion of the headplate, opposite the probe target area, and sealed in position with dental acrylic. The silicone elastomer was removed and the incision posterior to the headplate was closed with sutures. Mice were allowed to recover for 6–35 days (19 ± 10 days, mean \pm s.d., $n = 45$ mice) before probe implantation was attempted.

Probe preparation

All recordings were performed using A1x32-Poly3-5mm-25 s-177 silicon probes (177 μ m² site surface area, 3-column honeycomb site geometry with 18 μ m lateral and 25 μ m vertical site spacing, 36 μ m centre-to-centre horizontal span, 275 μ m centre-to-centre vertical span, 114 μ m maximum shank width near the sites, 15 μ m shank thickness) with an H32 connector (NeuroNexus Technologies). Before implantation, a NanoZ (White Matter) was used to electroplate each of the 32 electrodes (+5 nAmp, 30 s) with poly(3,4-ethylenedioxythiophene) (PEDOT) to reduce site impedance^{48,49}.

The probe tip base was affixed with cyanoacrylate adhesive to a 3D-printed titanium anchor, which itself was coupled to a probe tip connector, 3D-printed using a water-soluble polymer (polyvinyl alcohol, PVA). The probe tip connector was attached to custom-designed fittings (T. Tabachnik, Columbia University), which also held the probe's printed circuit board (PCB) and Omnetics connector and were secured to a two-axis gimbal (G. Johnson, Columbia University).

The two-axis gimbal was mounted on the Scientifica manipulator and was used to align the probe shank to the vertical axis of travel of the manipulator. Alignment was optimized until the probe shank, as it passed through the approximately 3 mm that would enter the brain during implantation, deviated by no more than 10 μ m from a fixed point, visualized using a reticle (alignment within 12 arcminutes). The probe was then left, mounted, on the two-axis gimbal with the alignment preserved. During the implantation procedure, the two-axis gimbal was again mounted on the Scientifica manipulator and the probe's alignment was confirmed before penetration was initiated.

Probe implantation

Mice were anaesthetized with ketamine and xylazine (100 mg/kg ketamine and 10 mg/kg xylazine initial dose). An intraperitoneal cannula was inserted to administer maintenance doses of ketamine (30 mg/kg every 20–30 min). The mouse was placed on the feedback-controlled heating pad and carprofen (5 mg/kg) was administered via subcutaneous injection. The headplate was then inserted into the stereotactic frame using custom holders and levelled to ensure its alignment to the horizontal plane of the Scientifica manipulator. The alignment of the headplate was confirmed using a dial test indicator affixed to the Scientifica manipulator. To minimize vibration during implantation, all procedures were performed on an air table (TMC).

Using the photographs previously taken during head-plate attachment, a glass guide pipette attached to the manipulator was positioned at the previously targeted location for probe implantation. A craniotomy was then performed using a dental drill (Osada Success 40, Osada

Article

Electric Co.) and a 0.5-mm drill bit (Fine Science Tools). Considerable care was taken when performing the craniotomy to avoid damage to the underlying tissue; in the event of pial bleeding, oedema, or other signs of damage to the brain the probe implantation was aborted. After craniotomy had been successfully performed, the glass guide pipette was lowered to the surface of the brain and its position relative to the blood vessels on the pial surface was photographed to guide later placement of the silicon probe. The craniotomy was then covered with cotton moistened with phosphate-buffered saline. A custom-designed well moulded out of silicone elastomer (Kwik-Cast) was affixed atop the head plate to permit later dissolution of the PVA probe tip connector once the probe was cemented in place. The glass guide pipette was then removed from the manipulator and the silicon probe was positioned above the insertion site using the photographs of the guide pipette.

The probe was lowered to 1,000 μm below the pial surface under manual control of the motorized manipulator. Throughout the penetration, neural signals were monitored both visually and aurally. Beyond depths of 1,000 μm below the pial surface, software was used to control the speed of descent. The probe was lowered from 1,000 μm to 2,200 μm at 2 $\mu\text{m}/\text{s}$ and beyond 2,200 μm was advanced at 1 $\mu\text{m}/\text{s}$ until the desired target area was reached. Proper probe alignment was confirmed by visually tracking the movement of distinctive spike waveforms as the array of electrode sites passed by them during penetration. Good alignment was further confirmed by temporarily retracting the probe (between approximately 500 μm and 1,500 μm) and confirming that spike waveforms, which the probe sites had already passed, reappeared. If signals were attenuated during these two checks (indicating that the probe was causing damage to the tissue, possibly due to slight misalignment), implantation was aborted.

The cell-dense layer 2 of APCx was identified by a rapid increase in threshold-crossing events and tight coupling of the local field potential to the animal's breathing cycle. In 15 preliminary APCx targeting experiments, we coated a glass capillary with Dil (Thermo Fisher Scientific) and determined post hoc in histology appropriate stereotactic coordinates. Then, in five preliminary acute mapping experiments, during which we applied Dil to the back of the probe, we related clear physiological signatures along the penetration track with physical locations determined post hoc in histology (Extended Data Fig. 1a, right). These physiological signatures permit us to determine unambiguously whether our electrode sites are located in the cell-dense layer of piriform cortex, and thus to decide whether to commit a probe to chronic implantation. Once the probe reached its target position (depth from pial surface $3,118 \pm 131 \mu\text{m}$, $n = 32$ mice, mean \pm s.d.) it was retracted to accommodate settling of the tissue (retraction from target depth $117 \pm 41 \mu\text{m}$, mean \pm s.d., $n = 32$ mice). If the probe failed to target the cell-dense layer 2 on the first attempt, implantation was aborted; to minimize tissue damage, we never made more than a single penetration during implantation.

Once the probe was positioned within layer 2 of APCx, a thin protective coating of uncured (liquid) silicone (DOWSIL 3-4680 Silicone Gel, Dow) was carefully applied to the pial surface through a 23-gauge blunt syringe needle affixed to a 1-ml syringe. After the silicone had cured, very liquid (uncured) dental acrylic (Tru-Pak Orthodontic Acrylic, Stoelting) was carefully applied through a 23-gauge blunt syringe needle affixed to a 1-ml syringe, surrounding the shank of the probe up to its base, similar to previous work in which silicon probes were implanted without using a microdrive^{50,51}. The dental acrylic was then allowed to cure for at least 20 min. The dental acrylic and still-exposed probe tip holder were then carefully covered in Grip Cement (Dentsply), again using a 23-gauge blunt syringe needle affixed to a 1-ml syringe, in two phases. The Grip Cement was allowed to harden for at least 20 min after each phase. The Grip Cement surrounded the probe tip up to the PVA probe tip connector and flexible polyimide cable that connected the probe tip to the probe's PCB and Omnetics connector. Between the time the probe entered the brain and the time it was gently decoupled

from the micromanipulator (see below), considerable care was taken to minimize movement of the probe relative to the brain tissue in any direction other than vertical. Only fluids (uncured silicone, dental acrylic, Grip Cement) were permitted to contact the probe/micromanipulator assembly; contact with a solid object (for example, the 23-gauge needle) typically resulted in degradation of the continuously monitored spike waveforms and abortion of the probe implantation.

The silicone well was then filled with distilled water in order to dissolve the PVA probe tip connector, thus ensuring gentle decoupling of the probe from the micromanipulator. Once the PVA had dissolved (~30 min), the only remaining physical connection between the probe tip and the PCB and Omnetics connector was the thin, flexible polyimide cable. Thus, at this point, the PCB and Omnetics connector could be disconnected from the manipulator without disturbing the implant. The PCB and Omnetics connector were then gently affixed with Grip Cement within a copper mesh casing. Note that the PCB and Omnetics connector were not affixed to the headplate or Grip Cement surrounding the probe itself, to isolate the implant from physical insult when plugging or unplugging the headstage during subsequent recording. The copper mesh casing was then secured to the headplate using Grip Cement and a lead connected to the probe's ground wire was soldered to the copper mesh. The probe was unplugged from the headstage and, to protect the pin contacts from debris and dander, the exposed sites of the Omnetics connector were covered with tape (Scotch Magic Tape, 810-NA, 3M). The animal was then returned to its homecage.

Stabilization following implantation

We have observed that in some cases the brain tissue appears to relax around the probe over as long as 2–3 weeks, as evidenced by large, distinctive single unit waveforms slowly creeping up the vertical axis of the electrode sites over several days. In order to ensure complete stabilization of the tissue around the probe and thus permit longitudinal recordings of single units across days, mice were given 9 ± 3 weeks (mean \pm s.d., minimum 5 weeks, $n = 16$ mice) to recover (Extended Data Fig. 2a, b). Thereafter, we began daily recordings. During an initial monitoring phase, animals were head fixed and the head stage was plugged into the probe's Omnetics connector for daily 30-min recording sessions to assess single-unit stability across days. After at least 10 days of daily monitoring, experiments commenced (see 'Experiment schedule' and Extended Data Fig. 2a, b).

Histology

Mice were transcardially perfused with 4% paraformaldehyde 0.1 M phosphate buffer fixative. The brain was postfixed overnight, washed in phosphate-buffered saline and sectioned along the coronal plane (100- μm sections) using a vibratome (Leica VT1000s, Leica Microsystems). Cell bodies were stained using NeuroTrace (Life Technologies) for 30 min at a 1:50 dilution. Sections were imaged under a Zeiss LSM 510 microscope (Carl Zeiss) under the control of ZEN imaging software (Carl Zeiss).

Experimental apparatus

During all recording sessions, animals were head fixed and positioned atop a trimmed absorbent underpad (Fisher Scientific) within a 3D printed polylactic acid tube (45.5 mm inner diameter, 49 mm outer diameter, 7 cm long, with a sealed back). For most experiments the tube was secured in a fixed position with respect to the headplate holders, but to monitor behavioural responses to aversively conditioned stimuli the tube floated on a pair of frictionless air bearings using a virtual burrow assay²⁹. A background of bandpass filtered acoustic white noise (1,000–45,000 Hz; approximately 7 dB) was played throughout all experiments and the mouse and apparatus were within a custom-made sound-attenuating chamber resting on an air table (TMC). Experiments were conducted in dark conditions, with animals illuminated with infrared light to permit simultaneous video monitoring.

Stimulus delivery

Odorant stimuli (Extended Data Fig. 6) were administered using a 4-s odorant pulse with an average 60-s inter-trial interval (ITI), drawn at random from a uniform distribution between 40 and 80 s. Each odour test session consisted of seven stimulus presentation blocks in which each odorant stimulus was delivered pseudorandomly within each block. Animals were given 3 min to acclimate to head fixation before we started stimulus presentation.

Odorant stimuli were administered using a custom-built olfactometer. At all times a constant flow of air (0.7 l/min) was directed through a nose port constructed of polyether ether ketone (PEEK), placed approximately 1 mm from the nose of the mouse. Air was routed to the nose port via one of two independent air lines: an air stream, normally routed to the nose port, and an odour stream, normally routed to exhaust but routed to the nose port upon administration of odorant stimuli. Both the odour and air streams were supplied with dedicated tanks of medical grade air and the flow rate of each was regulated by a dedicated mass flow controller (GFCS-010201, Aalborg). Both lines were directed through 50-ml glass bottles containing 15 ml dipropylene glycol (DPG, MilliporeSigma).

Monomolecular odorants were dissolved in 15 ml DPG in separate 50 ml bottles and replenished before each odour test session. The inlets and outlets of all odorant bottles were connected to the odour stream via two PEEK manifolds (Western Analytical). The inlet of each bottle was affixed with a normally closed two-way valve. The outlet was affixed with a three-way valve normally open to an exhaust line (two-way and three-way valves manufactured by the Lee Company). In preparation for odorant stimulus delivery, the two-way inlet valve was opened and the three-way outlet valve was switched from the exhaust to the odour stream. The odour stream was then allowed to stabilize for at least 30 s before the odorant stimulus was administered. To deliver odorant stimuli, a four-way valve (NResearch) routed the air stream to exhaust, replacing it with the now odorized odour stream; the odorant stimulus was switched off when the four-way valve routed the odour stream back to exhaust and the air stream back to the nose port. The four-way valve was housed outside the experiment chamber in a sound-attenuating box. The timing of olfactometer valve state changes and the trial structure of the experiment were controlled by a National Instruments board (National Instruments) using custom software written in Python (PyDAQmx³² was used to interface with the NI-DAQmx driver).

After passing through the nose port, all gas was routed to a photoionization detector (miniPID, Aurora Scientific). To minimize contamination, all material in contact with the odorized air stream, including the wetted material of all valves, was constructed from Teflon, Tefzel, or PEEK. The flows of the air and odour streams were equalized before each experiment (using mass flow meter GFMS-010786 from Aalborg) and the tubing carrying the two streams from the four-way valve was set to equal length and impedance to minimize variation in flow rate upon switching between the air and odour streams. PID signals for each experiment were recorded to monitor stimulus delivery (Extended Data Fig. 6). In subsequent analysis, we scaled all signals using a set of reference molecules to correct for small, uniform changes in PID sensor amplitude across days, resulting from fluctuations in bulb intensity, for example.

Odorant stimuli

The following odorants were used in these experiments, with their concentration (volume by volume in DPG) titrated during preliminary concentration-dependence experiments to evoke responses in comparable fractions of piriform neurons: 2% *cis*-3-hexen-1-ol, 2% octanal, 2% anisole, 4% ethyl *trans*-3-hexenoate, 2% or 4% isopentyl acetate, 4% nerol, 4% decanal, 4% or 6% linalool oxide, 10% methyl salicylate, 10% cuminaldehyde, 10% geranyl nitrile (1:1 mixture of (E)- and (Z)-isomers),

10% R(-)-carvone (Extended Data Fig. 6). All odorants were purchased from MilliporeSigma, Burlington, MA with the exceptions of nerol, ethyl *trans*-3-hexenoate, and cuminaldehyde, which were purchased from Thermo Fisher Scientific, and geranyl nitrile, which was purchased from W. W. Grainger and manufactured by Tokyo Chemical Industry.

For measurements of representational drift over a 32-day interval, we tested two cohorts of mice. One cohort ($n = 3$ mice) was presented with a panel of eight odorants every eight days. The second cohort ($n = 3$ mice) was presented with a panel of four odorants every eight days, as well as a panel of four familiar odorants presented daily from days 0 to 16 and at 8-day intervals thereafter. For this second cohort, only responses to the four odorants presented at 8-day intervals throughout were considered in our analysis of representational drift over a 32-day interval (Figs. 1, 2). In one animal from the first cohort, the panel of eight odorants was presented once additionally, nine days before odour testing commenced. Otherwise, all odorants included in our measure of drift over a 32-day interval were not presented to the animals before the first test day (day 0).

For experiments measuring the effect of aversive conditioning on drift rate (Fig. 4a, b), the CS+ was nerol; the CS- was linalool oxide; the neutral odorants were isopentyl acetate, methyl salicylate, cuminaldehyde, and R(-)-carvone. For experiments measuring the effect of daily experience on drift rate (Fig. 4c, d), the four familiar odorants were taken from *cis*-3-hexen-1-ol, octanal, anisole, ethyl *trans*-3-hexenoate, decanal, and geranyl nitrile (1:1 mixture of (E)- and (Z)-isomers). The four unfamiliar odorants were taken from isopentyl acetate, nerol, decanal, linalool oxide, methyl salicylate, cuminaldehyde, geranyl nitrile (1:1 mixture of (E)- and (Z)-isomers), and R(-)-carvone. In all cases, the mouse had not previously experienced the unfamiliar odorant stimuli when the experiment commenced.

We avoided using odorant stimuli that had been previously shown to elicit innate attraction or aversion (we used 'neutral' odorants) and selected a structurally diverse array of odorant molecules with distinct organoleptic properties.

Experiment schedule

For experiments monitoring representational drift across a 32-day interval, test odorants were administered every eight days, presented on five 'odour-test days' (days 0, 8, 16, 24, and 32). On intervening days, animals were placed on the recording rig under identical conditions as on odour-test days but no odorant stimuli were administered. Except when otherwise noted, we grouped analysis into days 0–8, 8–16, 16–24, and 24–32 (8-day intervals), days 0–16, 8–24, and 16–32 (16-day intervals), and days 0–24 and 8–32 (24-day intervals).

For experiments monitoring the effect of daily exposure on representational drift, animals were presented with a panel of four 'familiar' odorants daily across a 16-day interval (days -16 to 0). On day 0, in addition to the four 'familiar' odorants, a panel of four 'unfamiliar' odorants that the animals had not previously experienced was presented. For cohort A ($n = 5$ mice), daily presentation of the 'familiar' odorants continued, with presentation of the 'unfamiliar' odorants occurring only every eight days (on days 0, 8, and 16). For cohort B ($n = 5$ mice) both 'familiar' and 'unfamiliar' odorants were presented every 8 days (days 0, 8, and 16). All analyses were performed on odour responses obtained on days 0, 8, and 16, looking at two 8-day intervals (days 0–8, 8–16) and one 16-day interval (days 0–16).

We note that daily experience (seven trials per odour per day) has a weak but statistically significant effect on within-day angle ($\rho = 0.18$, $P = 5.8 \times 10^{-5}$). It is therefore necessary to correct for these changes in within-day variability in order to make comparisons across different conditions (for example, familiar versus unfamiliar), which we do by subtracting within-day variability from across-day variability (corrected across-day angle). Daily exposure also resulted in slightly decreased within-day classification (familiar odorants 0.84 versus unfamiliar odorants 0.95), which we corrected for by normalizing classification

Article

performance between chance and the within-day performance, thus permitting comparison between classification accuracies for familiar and unfamiliar stimuli.

Odour–shock conditioning and testing

Conditioning and testing were performed using methods previously described²⁹. In brief, implanted mice were placed in a fear conditioning chamber under conditions of darkness with an acoustic background of white noise and allowed to acclimate for 5 min. Eight blocks of CS+ and CS– odourants were presented in pairs of pseudorandomly interleaved trials. The odourant stimuli were 10 s in duration with a 300 ± 100 s ITI. During the final 2 s of presentation of the CS+ only, the floor of the fear conditioning chamber was electrified (intensity 0.8 mAmp). Upon completion of all eight blocks, the mouse was permitted to recover for 5 min in the fear conditioning chamber and then returned to its home cage.

One day after the fear conditioning protocol (day 0), as well as 17 days after (day 16), animals were placed within a virtual burrow assay (VBA)²⁹ to monitor their behavioural responses to odourant stimuli (CS+, CS–, and four neutral odourants) and to record neural signals. Animals were acclimated to the VBA for 5 min after head fixation (3 min in the ‘open loop’ and 2 min in the ‘closed loop’ configuration²⁹). The odourants were presented as in previously described experiments (4-s odourant pulse, 60 ± 20 s ITI). The position of the virtual burrow was measured using a laser displacement sensor. Each trial began with the animal’s head well outside the tube. Movement of the virtual burrow around the body during the odourant stimulus constituted ‘ingress’, akin to a flight response²⁹. The position of the virtual burrow was quantified during the final second of odourant administration to measure the response to the CS+, CS–, and neutral odourants. On all intervening days (days 1–15) neural signals were recorded for at least 45 min with the tube in a fixed position and without presentation of odourant stimuli.

Data acquisition and spike sorting

Neural signals were acquired using a Cerebus Neural Signal Processor (Blackrock Microsystems) using differential recordings with a Cereplex M headstage (bandwidth: 0.3 Hz–7.5 kHz) sampled at 30 kHz under the control of the Cerebus Central Suite software. Continuous signals were then digitally bandpass filtered at 500–6,000 Hz and preprocessed to subtract electrical artefacts associated with muscle activity in the jaw and/or mystacial pad: at each time bin, the voltage registered on the seven electrode sites closest to 0 μ V was averaged and subtracted from all 32 electrode sites. Owing to the magnitude of fields generated by piriform pyramidal neuron action potentials combined with the dense spacing between electrode sites, action potentials typically registered on a large fraction, but not all, of the probe’s 32 electrodes. The choice of seven electrodes reflects a balance between accurate estimation of electrical artefacts common across all sites and avoiding subtracting signals associated with veridical action potential events.

We then concatenated all recordings across all days included in a given analysis (for example, for recordings of responses over a 32-day interval data from all 33 days of recording were concatenated in the order in which the recordings were acquired). The concatenated data were then spike sorted using the template-matching algorithm Kilo-sort⁵³ (<https://github.com/cortex-lab/KiloSort>); thus, the spike-sorting algorithm was blind to recording day and treated data concatenated across days as a single continuous signal. We used 256 templates with six passes and a spike threshold of 4σ . We did not use any drift correction, so any drifting waveforms were assigned separate templates. The output of Kilosort was manually curated using Phy (<https://github.com/cortex-lab/phy>). During this manual curation phase, we merged pairs of templates whose cross-correlogram indicated the spiking of a single common neuron; we did not perform any template splitting in the low-dimensional visualization, instead relying on oversplitting by requesting a large number of templates. We excluded from analysis any templates that corresponded to electrical noise, as well as any templates

for which refractory period violations exceeded 2% (Extended Data Fig. 1d, right; refractory period defined as an inter-spike interval <1.5 ms; median 0.08% refractory period violations, Q1 = 0.01%, Q3 = 0.24%; 17% of templates had zero refractory period violations). In summary, we did not attempt to stitch templates across multiple days, but rather designed the spike-sorting pipeline to treat concatenated continuous signals as a single recording.

We then excluded from further analysis single units that were not stable across the recording. First, we eliminated single units whose firing rate was near zero for any of the comparison days, as this indicates that the single unit was either gained or lost. Next, we computed the correlation between each single unit’s waveforms measured on individual comparison days. For this, we computed each single unit’s mean waveforms (82 samples, 2.7 ms total duration), averaged on each day for each electrode, and then concatenated the mean waveforms from the probe’s 32 electrodes. Single-unit waveform similarity is defined as the Pearson’s correlation between a pair of concatenated average waveform vectors. Within a given recording day, the 99th percentile correlation between the waveforms of different single units was 0.93 (Extended Data Fig. 1d, left). We therefore excluded single units whose correlation across a pair of comparison days fell below this value. Variations in waveform shape are amplified in large-amplitude single units. Thus, we used a final manual check to identify and include those rare cases in which very large amplitude single units had across-day waveform correlations that fell slightly below 0.93. Each of these exceptions is shown in Extended Data Fig. 2j (left), k (left).

Assessment of single-unit stability across days

The shape and magnitude of spike waveforms may vary over time, confounding efforts to follow single units^{43,51,54,55}. We used three metrics to quantify the stability of single units across days: waveform similarity (computed as described above), displacement of the estimated single-unit position relative to the probe sites, and similarity in the shape of the spike–time autocorrelogram (ACG distance).

To estimate single-unit position, we computed a spatial average across electrode positions weighted by the square of the mean waveform amplitude at each electrode. Many single units registered low-amplitude waveforms on a majority of sites on the probe, which can confound the estimate of the centroid and cause underestimation of centroid displacement across days. For this reason we used the square of the waveform peak-to-peak amplitude on each site to minimize the influence of low-amplitude waveforms, estimating single unit centroid position (x, y) as

$$(x, y) = \left(\frac{\sum_{i=1}^N x_i a_i^2}{\sum_{i=1}^N a_i^2}, \frac{\sum_{i=1}^N y_i a_i^2}{\sum_{i=1}^N a_i^2} \right),$$

where N is the number of electrodes, x_i is the lateral position of the i th electrode, y_i is the vertical position of the i th electrode, and a_i is the peak-to-peak amplitude of the spike waveform recorded at the i th electrode. Single-unit displacement was then computed using the Euclidean distance between single-unit centroids on separate days. We note that owing to the diverse and complex morphologies and biophysics of the neurons we recorded from, it is not possible to establish with precision the location of the source relative to the probe sites.

We computed the spike-time ACG over a 2-s window using a bin size of 1 ms. To measure the similarity in the shape of the ACG (ACG distance), we computed the Euclidean distance between normalized, filtered and then logarithmically binned spike–time ACGs.

For a given pair of comparison days, we computed each of these three metrics for all single units that were stable across that interval, as well as for all pairs of single units that were simultaneously recorded within a given recording session, including single units that were not held across any other session. We then compared within-unit/across-day

distributions to the across-unit/within-day distributions to assess whether single units were more similar to themselves across days than to other single units simultaneously recorded within a day (Extended Data Figs. 1f, i, 2, 3). We also identified for each single unit the ten and one most similar simultaneously recorded single unit(s) within a given day by ranking the waveform similarity for each single unit relative to every other single unit recorded on that day (Extended Data Fig. 3).

We also tested whether our three single-unit stability metrics predicted the stability of a single unit's evoked response to a panel of odourants (Extended Data Fig. 2j, k). To estimate the similarity of a single unit's odour responses across days we computed, for each single unit, a vector whose elements consisted of the spontaneous baseline-subtracted, trial-averaged firing rate during the stimulus epoch for each odourant. We then measured the Pearson's correlation between these vectors for each single unit across pairs of days.

To measure the dependence of waveform similarity and centroid displacement on the similarity in ACG shape (Extended Data Fig. 2f), we computed the Pearson's correlation of ACG distance with either waveform similarity or centroid displacement for all pairs of single units that were stable across a given interval on all days within that interval.

Peristimulus time histograms

To visualize the firing of individual single units we computed a peristimulus time histogram (PSTH) averaged across all trials for a given odourant stimulus on a given day, smoothed using a Gaussian kernel ($\sigma = 0.4$ s; Fig. 1b, Extended Data Fig. 4a). To generate response maps of odour-evoked population responses, we z-scored the smoothed PSTH by subtracting, for each single unit, its mean spontaneous baseline firing rate across all trials on a given day during the baseline epoch (4 s before stimulus onset, 'baseline firing rate') and dividing by the standard deviation of firing rates during the baseline epoch. Thus, the response maps (Fig. 1c, Extended Data Fig. 4b) indicate changes in firing rate in units of standard deviation of spontaneous activity. These two methods were used for visualization purposes only; unless stated otherwise, all statistical analyses were performed on unsmoothed, spontaneous baseline-subtracted firing rate estimates.

Classifier analysis

We assessed the linear separability of population odour responses by measuring classification performance using an SVM (linear kernel, L2-regularization). For within-day classification, we used leave-one-out cross-validation, training on all but one of the $8 \times 7 = 56$ trials on a given day (8 odourant stimuli, 7 trials each) and testing classifier prediction on the trial that was left out. This procedure was repeated until all trials on a given day had been tested in this way. For across-day classification, we trained a model on all 56 trials on one day and tested that model's predictions on all 56 trials on another day. Performance on shuffled data was tested by randomly permuting the odourant stimulus labels on the test dataset.

In light of the broad diversity in temporal response profiles across single units (for example, Fig. 1b, Extended Data Fig. 4a), ranging from transient ON or OFF responses to sustained responses that lasted several seconds, we performed preliminary analyses to optimize the quantification of single trials with respect to the temporal profile of the odour response. In these analyses we used stitched data across all experimental subjects and normalized the population vectors. A sliding window analysis (1-s window, Extended Data Fig. 5f) confirmed that there was odour-coding activity throughout the duration of the odourant stimulus epoch as well as for several seconds following stimulus offset, with local peaks immediately following stimulus onset and offset. We then tested 15 temporal binnings of the odour response, using a variety of time windows ranging from 250 ms to 8 s (starting at odour onset), using between 1 and 32 bins, and spanning either the stimulus epoch alone or both the stimulus epoch and a post-stimulus epoch (Extended Data Fig. 5g). The optimal binning scheme (as assessed

by maximum performance and lowest number of single units required to exceed 50% accuracy) was found to be one that spanned both the odour epoch and the post-odour epoch, averaging in time across four 2-s windows (0–2 s, 2–4 s, 4–6 s, 6–8 s). Accordingly, this is what we used for subsequent SVM analyses: population vectors were composed of $N \times 4$ elements where N is the number of single units that were stable across the pair of comparison days. (Note that comparable performance can be achieved using diverse binning schemes (Extended Data Fig. 5g)). Other than for this preliminary characterization, all analyses were performed on simultaneously recorded single units in individual mice using non-normalized baseline-subtracted evoked firing rates to preclude the possibility that separate renormalization within each day would cause a spurious decline in classification accuracy across days. To permit pooling of classifier performance results across mice, we were limited by the lowest number of stable single units for any across-day comparison in any animal (41 single units). In all comparisons for which there were more than 41 stable single units, we tested 100 randomly selected 41-single-unit subsets. Thus, within-day performance is expected to be substantially lower when not using data stitched across multiple experimental subjects, owing to the lower number of available units.

When we measured classification performance during the first 200 ms of the odour response (Extended Data Fig. 7), the narrow temporal window provided relatively noisy spike rate estimates with which to train the classifier. We therefore optimized the analysis parameters to promote classification performance as follows. (1) We used the dataset for which we had the highest minimum single-unit yield across all intervals (from among those animals that were presented with eight odourant stimuli). (2) We performed classification on only subsets of four out of the eight odours. Thus, performance was promoted by requiring only four-way rather than eight-way classification. Moreover, this scheme permitted 70 permutations over the eight odour classes. (3) Finally, we used z-scored estimated firing rates, rather than baseline-subtracted firing rates, to denoise the training data, and used z-scored estimated firing rates for all analyses reported in Extended Data Fig. 7b–e.

Population response similarity and drift rate calculation

To measure the similarity of population responses over time, we computed the Pearson's correlation between pairs of trial-averaged population vectors for each odour across all pairs of comparison days. Population vectors were computed as described above (see 'Classifier analysis'), then averaged across all trials for across-day comparisons, or averaged across even and odd trials separately for within-day comparisons. We then computed the Pearson's correlation between pairs of trial-averaged population vectors (Fig. 2c).

To compute the rate of drift across days, we first computed the angle between a pair of population vectors as follows:

$$\theta_{p,q} = \cos^{-1} \left(\frac{\mathbf{u}_p \cdot \mathbf{u}_q}{\|\mathbf{u}_p\| \|\mathbf{u}_q\|} \right),$$

where $\theta_{p,q}$ is the angle between days p and q and \mathbf{u}_p is the trial-averaged population vector on day p . Variability within days was estimated as follows:

$$\bar{\theta} = \frac{1}{M} \sum_{k=1}^M \theta_k,$$

where θ_k is the mean angle across even and odd trials on each day for all days k out of M total days. We then estimated the rate of drift using two approaches. (1) We fit the trial-averaged population vector angles across all the intervals using a single exponential function: $\theta = C - (C - R)e^{-t/\tau}$, where θ is the variability (angle), C the asymptote, R the intercept at $t = 0$ (within-day variability, $\bar{\theta}$), and τ the time constant

Article

of the exponential in days (Extended Data Fig. 5i, right). We then computed the drift rate over the 32-day interval by measuring the mean rate of change of the exponential fit. (2) We measured the drift rate across individual pairs of comparison days for individual odours as follows. For each odour, for each pair of days p and q , we computed the across-day angle $\theta_{p,q}$ between trial-averaged population vectors. To correct for within-day variability, without which we would overestimate across-day drift²⁵, we computed a corrected angle by subtracting the within-day angle $\bar{\theta}$. Drift rate ($r_{p,q}$) across the interval $\Delta t_{p,q}$ between the p th and q th days was $r_{p,q} = \frac{\theta_{p,q} - \bar{\theta}}{\Delta t_{p,q}}$ (Extended Data Fig. 5k). We also measured drift rate for each odorant stimulus tested and found that all exhibited mean drift rates between 0.8° per day and 1.6° per day (Extended Data Fig. 6f).

Measurement of population response geometry

To assess whether the geometry of the population response in piriform cortex drifted or remained the same across days, we computed odour similarity matrices within individual days and measured by how much these matrices changed as a function of the interval between a pair of days.

To construct odour similarity matrices, we used two standard measures of population vector similarity—Pearson’s correlation and cosine distance, both of which yield $N \times N$ symmetric matrices for N stimuli. We also developed a third measure, ‘edge angle’, in which we computed all angles formed between pairs of connected edges in the odour population response (Fig. 3a). This measure is insensitive to shape-preserving translation, rotation, and uniform expansion in the original neural space.

The change between a pair of odour similarity matrices A was computed by taking the Frobenius norm (‘matrix dissimilarity’ in main text) on the difference between the two,

$$\|A^{p,q}\|_F = \|A^p - A^q\|_F = \sqrt{\sum_{i=1}^M \sum_{j=1}^N |a_{i,j}^p - a_{i,j}^q|^2},$$

where A^p and A^q are odour similarity matrices measured on days p and q , respectively, of dimension $M \times N$. (For odour–odour correlation or cosine distance matrices $M = N$, where N is the number of odorant stimuli; for edge angle matrices, N is the number of odorants and $M = N \binom{N-1}{2}$.) A Frobenius norm of 0 indicates identical matrices, whereas Frobenius norms of increasing value indicate an increasing difference between the odour similarity matrices, and consequently an increasing difference between the geometrical relationships encoded by the two matrices.

To permit estimation of within-day variability and comparison with across-day variability, we computed odour-similarity matrices using trial-averaged population vectors taken across all combinations of 3-trial/4-trial splits of the data for a given odorant on a given session. This approach permits normalization of each across-day measure by within-day variability, as both are computed using an equivalent number of trials (‘normalized matrix dissimilarity’ in main text). We therefore could pool across animals, by rescaling results from individual subjects as follows,

$$\widehat{\|A^{p,q}\|_F} = \frac{\|A^{p,q}\|_F - \overline{\|A^w\|_F}}{\|A^s\|_F - \overline{\|A^w\|_F}},$$

where $\widehat{\|A^{p,q}\|_F}$ is the scaled Frobenius norm between similarity matrices on days p and q , $\overline{\|A^w\|_F}$ is the mean within-day Frobenius norm computed on all sessions and $\|A^s\|_F$ is the mean Frobenius norm computed using similarity matrices computed on the first and last day after shuffling stimuli. These analyses were performed in individual animals (that is, without stitching single units across animals) and only in those that were presented with the panel of eight stimuli.

To account for the temporal profile of the odour response, population vectors were quantified using the same 15 temporal binnings of the odour response as for the SVM preliminary analysis, as well as using a 1-s sliding window. The choice of binning scheme had a negligible effect on the relationship between matrix dissimilarity and interval, with all binnings showing an increase in Frobenius norm as a function of interval (Extended Data Fig. 8d, e, right); example plots use a single 4-s bin that spans the odorant stimulus epoch.

We also recomputed the odour similarity matrices after projecting the data onto lower-dimensional spaces. We projected the data onto its first six principal components, computed individually within each day for each of the 15 binnings of the odour response, and again found a negligible effect of binning choice on the relationship between matrix dissimilarity and interval (data not shown). For the 4-s, single-bin scheme shown, the first six principal components captured on average 98% of the variance. We also projected the data onto the first ten odour-coding dimensions, identified using demixed principal component analysis (PCA)²⁸ using the full odour response profile averaged in time windows of 250 ms over the odorant stimulus epoch and 4 s following it, computed individually within each day, which accounted for on average 57% of the odour coding variance. We note that the variance captured by the first six principal components computed using PCA exceeds the variance captured by the first ten odour coding dimensions identified using dPCA. This discrepancy is due to the fact that dPCA considers the temporal profile of the odour response, and therefore requires more dimensions to capture variability in the data: not only variability across the eight odour-stimulus conditions but also variability across time.

Statistics

All statistical analyses were performed using custom software written in MATLAB (MathWorks). Support vector machine analyses were implemented using LIBLINEAR⁵⁶ (<https://www.csie.ntu.edu.tw/~cjlin/liblinear/>). Auto- and cross-correlograms were computed using spikes (<https://github.com/cortex-lab/spikes>). Demixed PCA²⁸ was performed using dPCA (<https://github.com/machenslab/dPCA>). Except when otherwise noted, we used the nonparametric two-sided Wilcoxon rank-sum test to test the null hypothesis that the distribution of measurements in a pair of populations have equal medians.

To identify responsive odour–unit pairs, we used a significance level of $\alpha = 0.001$ (Wilcoxon rank-sum between the spike count during the 4-s epoch before stimulus onset for all trials on a given day and the spike count on the seven trials during the odorant stimulus). Analysis results were qualitatively similar across different choices of α .

To quantify the fraction of single units that were stable across a 32-day interval, we first excluded non-responsive and broadly tuned single units, using a threshold for lifetime sparseness that we varied between 0.2 and 0.65 (Extended Data Fig. 5e, right), which corresponds to the 40th to 95th percentiles, respectively, of all single units. For each of these single units we then computed a binarized, signed tuning vector (0 if the single unit did not have a significant response to that odorant, 1 if the odorant evoked a significant increase in firing rate, -1 if the odorant evoked a significant decrease in firing rate, $\alpha = 0.001$, Wilcoxon rank-sum). We defined a single unit as stable if it had equal binarized, signed tuning vectors across the 32-day interval. This quantity is reported for all values of lifetime sparseness thresholds tested (Extended Data Fig. 5e, right). We also measured, for each single unit, the percentage of significant odour responses measured on one day that were preserved on subsequent days (Extended Data Fig. 5d, e (left)). (Reported percentages of stable units and responses are with respect only to the set of units that were included for analysis at a given lifetime sparseness threshold.)

We measured the sparseness of the piriform cortex population response across days using standard methods^{4,57–59}. The population sparseness (S_p) for each odorant was defined as

$$S_p = \frac{1 - \frac{\left[\sum_{j=1}^N r_j \right]^2}{\sum_{j=1}^N r_j^2}}{1 - \frac{1}{N}},$$

where N is the number of single units and r_j is the response (spontaneous baseline-subtracted spike count during the 4-s odour epoch) of the j th single unit. Lifetime sparseness for each single unit was computed using the same equation, but with N corresponding to the number of odours tested and r_j the response of a given single unit to the j th odourant.

Reporting summary

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

Data availability

Data will be made available upon reasonable request to the corresponding authors.

Code availability

Code will be made available upon reasonable request to the corresponding authors.

48. Cui, X., Wiler, J., Dzaman, M., Altschuler, R. A. & Martin, D. C. In vivo studies of polypyrrole/peptide coated neural probes. *Biomaterials* **24**, 777–787 (2003).
49. Ludwig, K. A. et al. Poly(3,4-ethylenedioxythiophene) (PEDOT) polymer coatings facilitate smaller neural recording electrodes. *J. Neural Eng.* **8**, 014001 (2011).
50. Okun, M., Lak, A., Carandini, M. & Harris, K. D. Long term recordings with immobile silicon probes in the mouse cortex. *PLoS ONE* **11**, e0151180 (2016).
51. Steinmetz, N. A. et al. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science* **372**, eabf4588 (2021).
52. Cladé, P. *PyDAQmx: a Python Interface to the National Instruments DAQmx Driver* <http://pythonhosted.org/PyDAQmx/> (2010).

53. Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M. & Harris, K. D. in *Adv. Neural Information Processing Systems* 4448–4456 (NeurIPS, 2016).
54. Dickey, A. S., Suminski, A., Amit, Y. & Hatsopoulos, N. G. Single-unit stability using chronically implanted multielectrode arrays. *J. Neurophysiol.* **102**, 1331–1339 (2009).
55. Dhawale, A. K. et al. Automated long-term recording and analysis of neural activity in behaving animals. *eLife* **6**, e27702 (2017).
56. Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. & Lin, C.-J. LIBLINEAR: A library for large linear classification. *J. Mach. Learn. Res.* **9**, 1871–1874 (2008).
57. Perez-Orive, J. et al. Oscillations and sparsening of odor representations in the mushroom body. *Science* **297**, 359–365 (2002).
58. Rolls, E. T. & Tovee, M. J. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* **73**, 713–726 (1995).
59. Willmore, B. & Tolhurst, D. J. Characterizing the sparseness of neural codes. *Network* **12**, 255–270 (2001).
60. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

Acknowledgements We thank L. F. Abbott, S. R. Datta, S. Fusi, J. W. Krakauer, A. Kumar, M. A. Long, A. M. Michalek, and S. L. Pashkovski for comments on the manuscript; G. W. Johnson and T. Tabachnik for assistance with instrumentation; G. Buzsáki and members of the Buzsáki laboratory, D. A. Gutnisky, and C. C. Rodgers for experimental advice; the instructors of the 2015 Advanced Course in Computational Neuroscience at the Champalimaud Foundation for instruction; D. F. Albeanu and members of the Albeanu laboratory, D. Aronov, K. A. Bolding, R. Costa, J. P. Cunningham, J. T. Dudman, W. M. Fischler, K. M. Franks, M. E. Hasselmo, V. Jayaraman, A. Y. Karpova, E. Marder, J. A. Miri, C. Poo, D. Rinberg and members of the Rinberg laboratory, E. S. Schaffer, and D. D. Stettler for comments; M. Gutierrez, C. H. Eccard, P. J. Kisloff, and A. Nemes for general laboratory support; and the Howard Hughes Medical Institute and the Helen Hay Whitney Foundation for financial support.

Author contributions This work is the result of a close collaboration between C.E.S. and A.J.P.F., who conceived the study and designed the experiments. Experiments were performed by C.E.S., S.N.O. and A.J.P.F. The data were analysed and the manuscript was written by C.E.S., R.A. and A.J.P.F.

Competing interests The authors declare no competing interests.

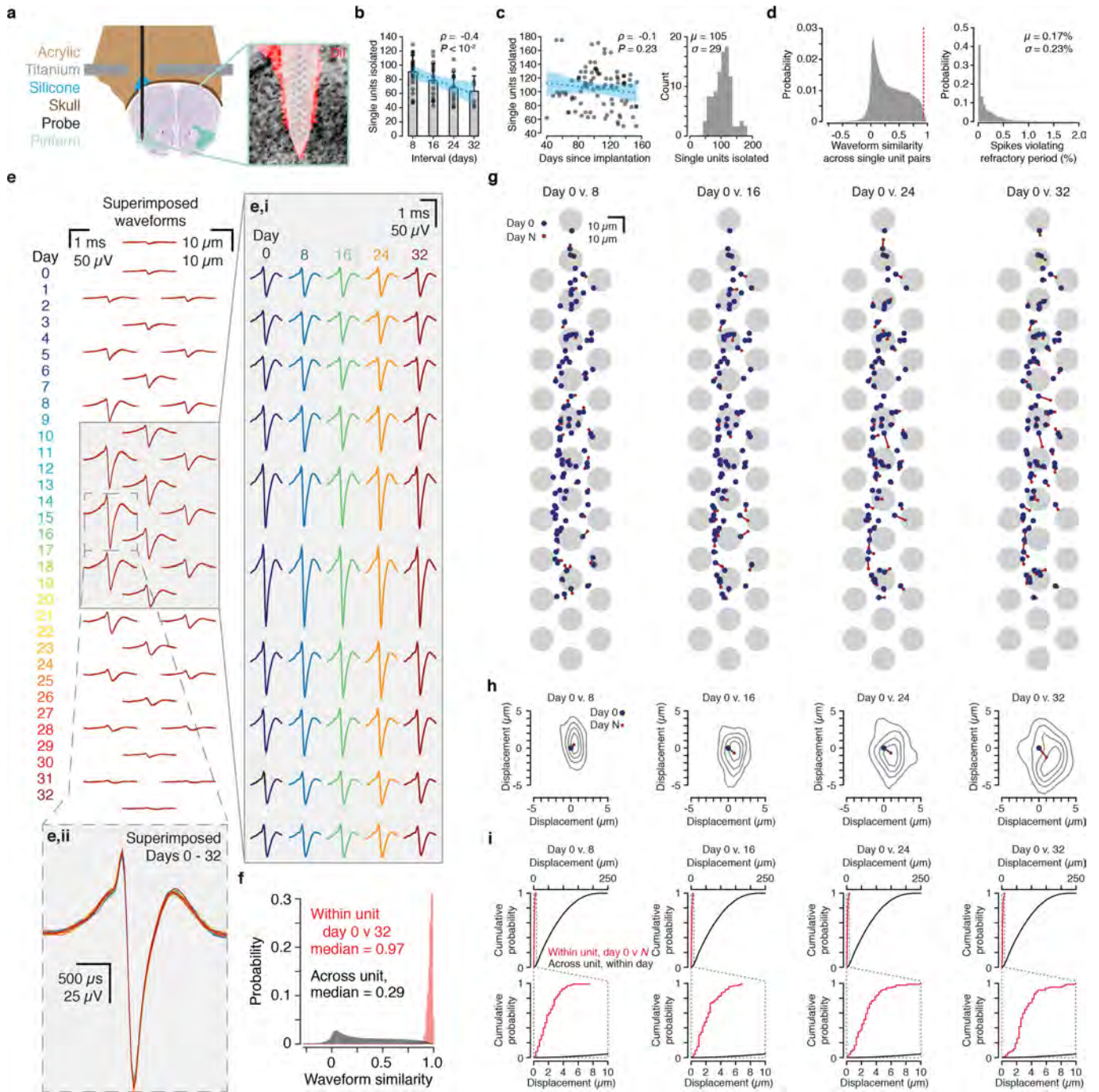
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03628-7>.

Correspondence and requests for materials should be addressed to C.E.S., R.A. or A.J.P.F.

Peer review information Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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

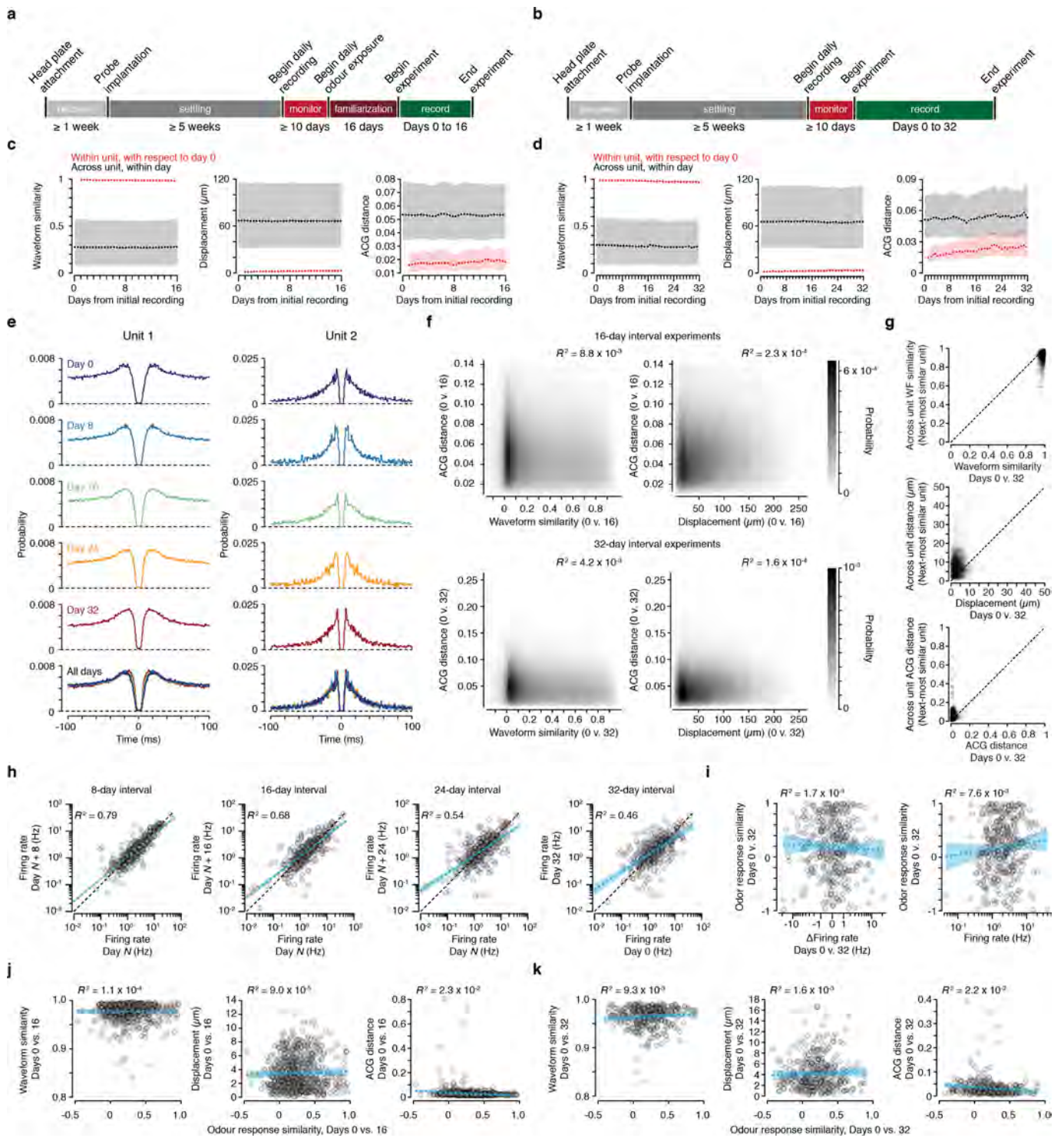


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Longitudinal tracking of single units in piriform cortex.

a. Chronic silicon probe implantation in anterior piriform cortex (green). Anatomical image and structure designations from the Allen Mouse Brain Atlas⁶⁰ (<http://www.brain-map.org>). Inset, probe diagram with relative positions of the 32 recording electrodes; red, Dil marking probe position; black, cell bodies (NeuroTrace). Diagrams not to scale. **b.** Number of single units retained as a function of recording interval duration. Mean \pm s.d. with individual data points; blue dotted line, linear regression; blue shading, 95% CI. $\rho = -0.41$, $P = 1.2 \times 10^{-3}$, $n = 24$ recordings across 8 days, $n = 18$ recordings across 16 days, $n = 12$ recordings across 24 days, $n = 6$ recordings across 32 days, all from 6 mice. **c.** Single-unit yield for single-day recording sessions. Left, per recording session as a function of time since probe implantation. Blue dotted line, linear regression; blue shading, 95% CI. $\rho = -0.12$, $P = 0.23$. Right, single-unit yields across all single-day recording sessions ($n = 97$ recording sessions in 16 mice). **d.** Left, probability density of waveform similarities for all pairs of single units simultaneously recorded within each day. Red dashed line indicates inclusion criterion (0.93, the distribution's 99th percentile) for rejection of candidate single units recorded across multiple days. Right, probability density of refractory period violations (refractory period is defined as an inter-spike interval < 1.5 ms). **e.** Average waveforms for a representative single unit recorded at each of the recording sites of the silicon probe. Waveforms from all days (0–32) are superimposed with each day plotted as a separate colour (colour scheme maintained throughout). Inset **e.i**, mean waveforms for days 0, 8, 16, 24, and 32 for a subset of recording sites, indicated by the grey box. Inset **e.ii**, mean waveforms for days 0 to 32, superimposed for a single recording site (dashed grey box). **f.** Waveform correlations for each

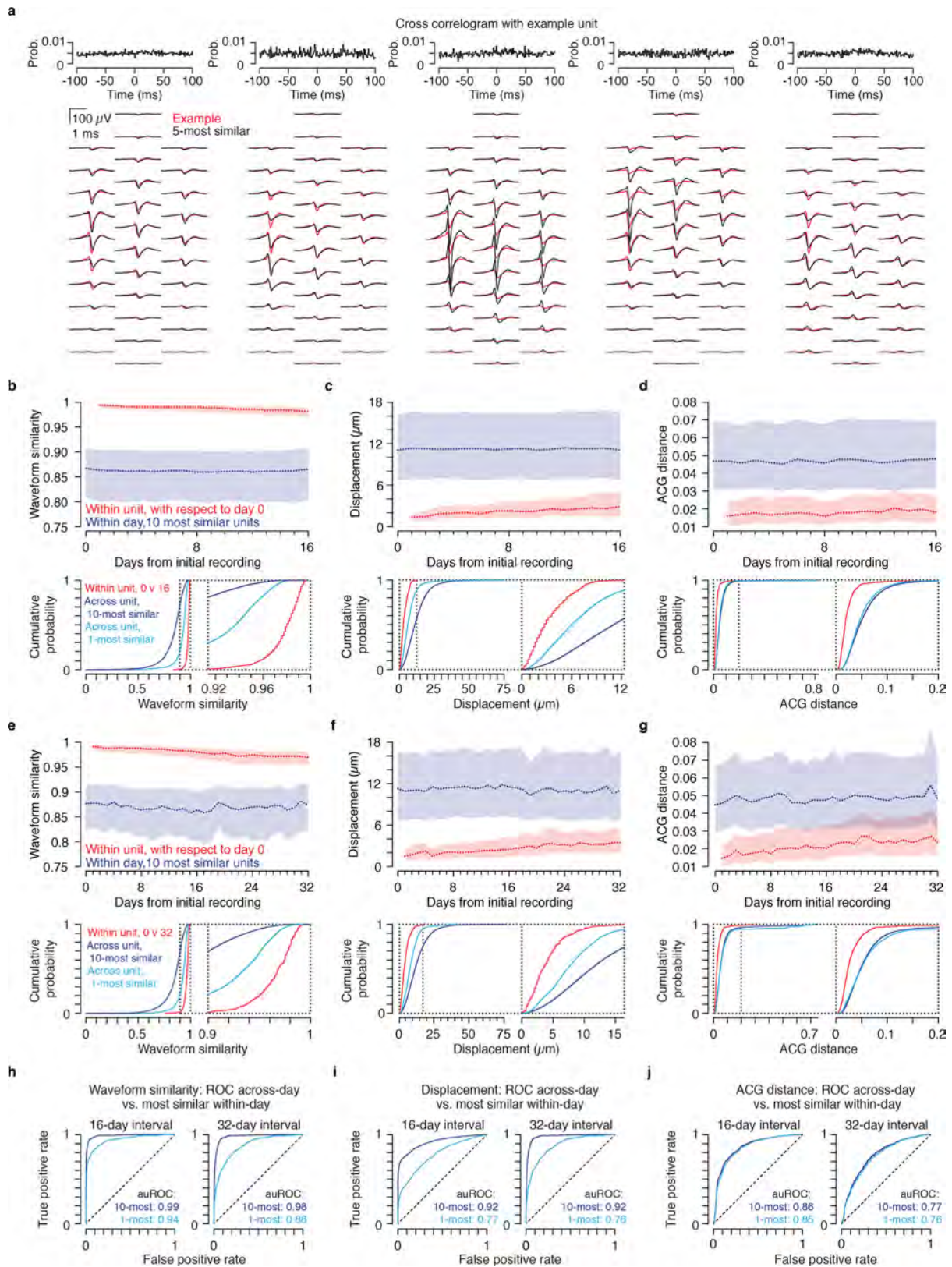
single unit across days 0 and 32 (red) and across all single units within-day (grey); within-unit, across-days versus within-day, across-units, $P = 4.8 \times 10^{-246}$, Wilcoxon rank-sum, $n = 379$ single units from 6 mice. The grey distribution is replotted from **d** (left). **g.** Single-unit waveform centroids across a 32-day interval from a representative mouse (centroid computed using spatial average across electrode positions weighted by the squared mean waveform amplitude at each electrode). Centroid for each single unit isolated on day 0 (blue circles) and days 8, 16, 24, and 32 (red circles, columns 1–4, respectively; days 0 versus 8, $n = 100$ single units; days 0 versus 16, $n = 94$ single units; days 0 versus 24, $n = 84$ single units; days 0 versus 32, $n = 77$ single units). Grey circles indicate the positions and sizes of the probe's 32 electrode sites. **h.** Mean displacement of single-unit centroids from this mouse between day 0 (blue circle, defined at origin) and days 8, 16, 24, and 32 (red circles, columns 1–4, respectively). Grey contours indicate quintile boundaries of the distribution of centroid position displacement for the population. **i.** Top, cumulative distribution of within-unit centroid displacement (red) between day 0 and days 8, 16, 24, and 32 (columns 1–4, respectively) and across-unit centroid displacement within day (black) for this mouse. Median on day 0 versus 8 within-unit = $1.8 \mu\text{m}$ (Q1 = $1.1 \mu\text{m}$, Q3 = $2.9 \mu\text{m}$), across-unit = $63.5 \mu\text{m}$ (Q1 = $31.5 \mu\text{m}$, Q3 = $107 \mu\text{m}$); day 0 versus 16 within-unit = $2.1 \mu\text{m}$ (Q1 = $1.3 \mu\text{m}$, Q3 = $3.2 \mu\text{m}$), across-unit = $64.2 \mu\text{m}$ (Q1 = $32.0 \mu\text{m}$, Q3 = $108 \mu\text{m}$); day 0 versus 24 within-unit = $2.6 \mu\text{m}$ (Q1 = $1.5 \mu\text{m}$, Q3 = $3.5 \mu\text{m}$), across-unit = $64.4 \mu\text{m}$ (Q1 = $32.2 \mu\text{m}$, Q3 = $108 \mu\text{m}$); day 0 versus 32 within-unit = $2.8 \mu\text{m}$ (Q1 = $2.2 \mu\text{m}$, Q3 = $4.2 \mu\text{m}$), across-unit = $64.5 \mu\text{m}$ (Q1 = $32.1 \mu\text{m}$, Q3 = $108 \mu\text{m}$); for all comparisons $P < 9.5 \times 10^{-51}$, Wilcoxon rank-sum, n as in **g**. Inset at bottom, x -axis 0 to $10 \mu\text{m}$.



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Assessing the stability of single units recorded across multiple days. **a, b**, Experiment time courses for 16-day (**a**) and 32-day (**b**) interval protocols. 'Recovery', period following headplate attachment and stereotactic targeting before silicon probe implantation to allow full recovery; 'settling', minimum five-week period after probe implantation to permit tissue settling and signal stabilization; 'monitor', minimum ten-day period during which neural signals were recorded daily to assess signal stability. Experiments began only once single units could be reliably tracked across days. 'Record', experiment protocol (Figs. 1a, 4a, c). 'Familiarization' (16-day interval experiments in **a** only), daily odour presentation for experiments described in Fig. 4c, d and Extended Data Fig. 10d–h. **c**, For single units held during 16-day interval experiments, waveform similarity (left; Pearson's correlation), centroid displacement (middle), and spike time ACG distance (right; Euclidean norm between normalized ACGs) measured between day 0 and subsequent days (red, 'within-unit, across-day') and across all single units within each day (black, 'across-unit, within-day'). Dotted lines, median. Shading, boundaries of top and bottom quartiles ($n = 690$ single units from 7 mice). **d**, As in **c** but for single units held during 32-day interval experiments ($n = 379$ single units from 6 mice). **e**, Example spike–time autocorrelograms from two single units recorded in the same mouse on five separate days. **f**, Density heatmap showing ACG distance of pairs of simultaneously recorded single units plotted against waveform similarity (left) and distance between their centroids (right) for those pairs. Top, 16-day interval experiments ($n = 1,248,216$ pairs of single units from 7 mice on 17 days); bottom, 32-day interval experiments ($n = 841,138$ pairs of single units from 6 mice on 33 days). This shows that waveform-based features (waveform similarity and centroid distance) vary independently of the

similarity of the spike–time ACGs. Thus, ACG distance is a measure of single-unit stability to which the spike-sorting pipeline is insensitive. **g**, Waveform similarity (top), centroid distance (middle), and ACG distance (bottom) for a given single unit between days 0 and 32, plotted against the same metric applied to the same single unit versus the most similar other simultaneously recorded single unit. Dashed line, unity. **h**, Mean single unit spontaneous firing rate on an individual day (baseline firing rate) compared across intervals of 8 days ($\rho = 0.89$, $n = 2,177$ single units), 16 days ($\rho = 0.82$, $n = 1,412$ single units), 24 days ($\rho = 0.74$, $n = 816$ single units) and 32 days ($\rho = 0.68$, $n = 379$ single units) from 6 mice. For all correlations, $P < 4.0 \times 10^{-52}$. Each plot shows a random subset of 379 single units, to match the number of single units recorded across the 32-day interval (right). Black dashed line, unity; blue dotted line, linear regression; blue shading, 95% CI. **i**, Odour response similarity plotted against change in mean spontaneous firing rate on a symlog scale (left; $\rho = -0.041$, $P = 0.43$) and absolute spontaneous firing rate on a log scale (right; $\rho = 0.087$, $P = 0.09$) across a 32-day interval ($n = 379$ single units from 6 mice). Blue dotted line, linear regression; blue shading, 95% CI. **j**, For single units held during 16-day interval experiments, waveform similarity (left), centroid displacement (middle) and ACG distance (right) measured between days 0 and 16, plotted against odour response similarity of that same single unit (Pearson's correlation of pairs of vectors computed on the two days, consisting of the response magnitudes for each odourant of a panel); black circles, individual units; blue dotted line, linear regression; blue shading, 95% CI. **k**, As in **j** but for single units held during 32-day interval experiments, plotting these features measured between days 0 and 32.

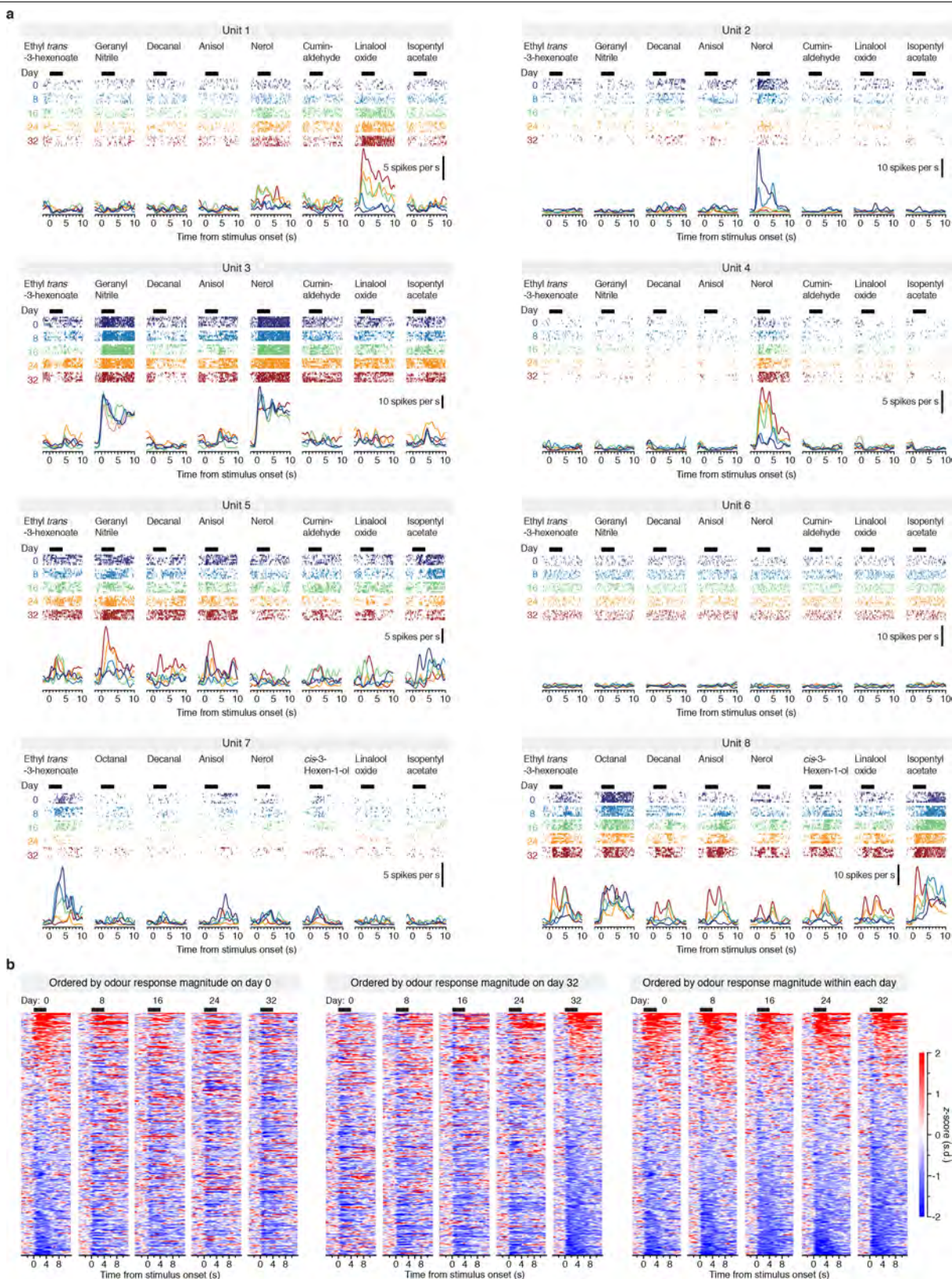


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | A single unit is more similar to itself across days than to any other single unit on the probe.

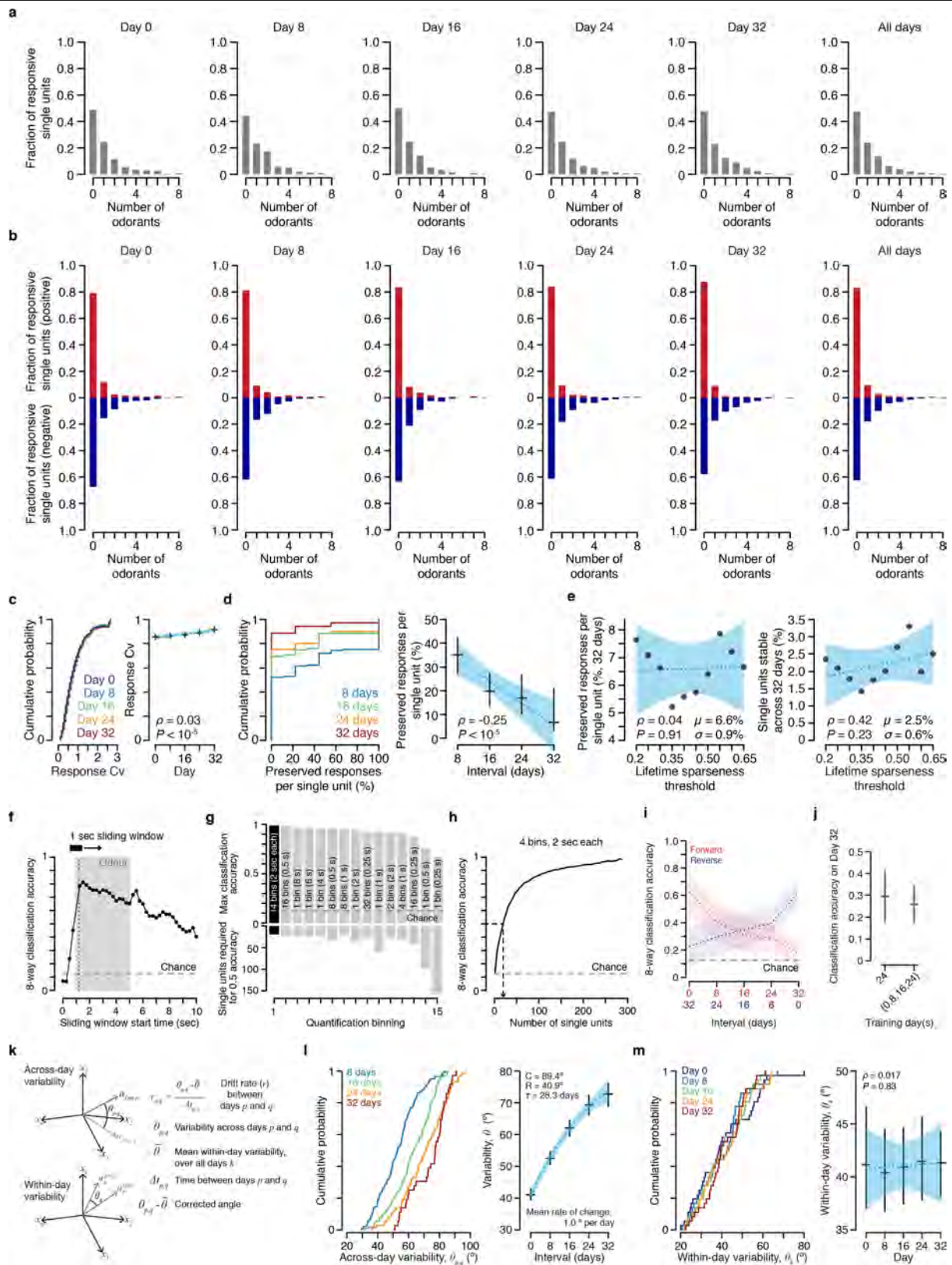
a, Bottom, single unit recorded on day 32 (red, same as shown in Extended Data Fig. 1e) overlaid with the five most similar single units to it recorded on the same day (black), as measured by waveform similarity (Pearson's correlation between waveforms). Top, spike-time cross correlograms between the example single unit and each of the five most similar single units. The absence of a dip in cross correlogram amplitude at the 0-ms time lag (refractory period violations) indicates that the example single unit and each of the five most similar single units correspond to distinct neurons. **b**, Top, for single units held across days, waveform similarity measured on day 0 versus subsequent days (red; within-unit, across-days), and waveform similarity measured between a given single unit and the ten single units most similar to it within a given day (blue; across-units, within-day). Dotted line, median. Shading, top and bottom quartiles. Bottom, cumulative distributions (left, full distribution; inset at right, expanded x-axis) within-unit for day 0 versus day 16 (red), ten most similar single units within a day (blue), one most similar single unit within a day (cyan). Median within-unit waveform similarity between day 0 and day 16, 0.98 (Q1 = 0.97, Q3 = 0.99); median across-unit waveform similarity with the ten most similar, 0.86 (Q1 = 0.80, Q3 = 0.90); median across-unit waveform similarity with the one most similar, 0.94 (Q1 = 0.91, Q3 = 0.95). **c**, **d**, Same analysis as in **b** but for centroid displacement and spike-time ACG distance (Euclidean norm between normalized ACGs), respectively. **c**, Median within-unit displacement between day 0 and day 16, 2.9 μm (Q1 = 1.6 μm , Q3 = 5.0 μm); median across-unit distance from the ten most similar, 11.2 μm (Q1 = 6.9 μm , Q3 = 16.5 μm); median across-unit distance from the one most similar, 6.1 μm (Q1 = 3.6 μm , Q3 = 9.3 μm). **d**, Median within-unit

ACG distance between day 0 and day 16, 0.018 (Q1 = 0.012, Q3 = 0.028); median across-unit ACG distance for the ten most similar, 0.047 (Q1 = 0.031, Q3 = 0.069); median across-unit ACG distance for the one most similar, 0.044 (Q1 = 0.030, Q3 = 0.064). **e-g**, As in **b-d** but for the experiments performed across a 32-day interval. **e**, Median within-unit waveform similarity between day 0 and day 32, 0.97 (Q1 = 0.96, Q3 = 0.98); median across-unit waveform similarity with the ten most similar, 0.87 (Q1 = 0.82, Q3 = 0.91); median across-unit waveform similarity with the one most similar, 0.93 (Q1 = 0.91, Q3 = 0.96). **f**, Median within-unit displacement between day 0 and day 32, 3.5 μm (Q1 = 2.1 μm , Q3 = 5.6 μm); median across-unit distance from ten most similar, 11.2 μm (Q1 = 6.9 μm , Q3 = 16.7 μm); median across-unit distance from one most similar, 7.3 μm (Q1 = 4.6 μm , Q3 = 10.7 μm). **g**, Median within-unit ACG distance between day 0 and day 32, 0.024 (Q1 = 0.016, Q3 = 0.038); median across-unit ACG distance for the ten most similar, 0.048 (Q1 = 0.032, Q3 = 0.072); median ACG distance for the one most similar, 0.050 (Q1 = 0.030, Q3 = 0.077). All within-unit metrics are significantly different from across-unit metrics ($P < 1.4 \times 10^{-60}$, Wilcoxon rank-sum), for both the one most and ten most similar comparisons, across both the 16-day ($n = 690$ single units from 7 mice) and 32-day ($n = 379$ single units from 6 mice) interval experiments. Thus, a single unit is more similar to itself across days than it is even to those single units most similar to it recorded within a given day. **h-j**, Receiver operator characteristic (ROC) curves showing the true positive rate versus the false positive rate for waveform similarity (**h**), centroid displacement (**i**) and ACG distance (**j**) for both 16-day and 32-day intervals for the ten most (blue) and one most (cyan) similar units. When computing the ROC, 'signal' was defined as the distribution with the higher mean. Dashed line corresponds to unity.



Extended Data Fig. 4 | Evoked responses in single units across days.
a, Activity of eight single units across a 32-day interval selected to illustrate the diversity of odour-evoked response profiles across single units within each day and within individual single units across days. Columns separate test odourants (chemical names, top). Spike rasters (rows: 7 trials per day) and PSTHs are colour-coded by day as indicated. Horizontal black bars, 4-s odourant stimulus epochs. The single-unit responses to individual stimuli shown in Fig. 1b are replotted here alongside responses for those single units to the seven other

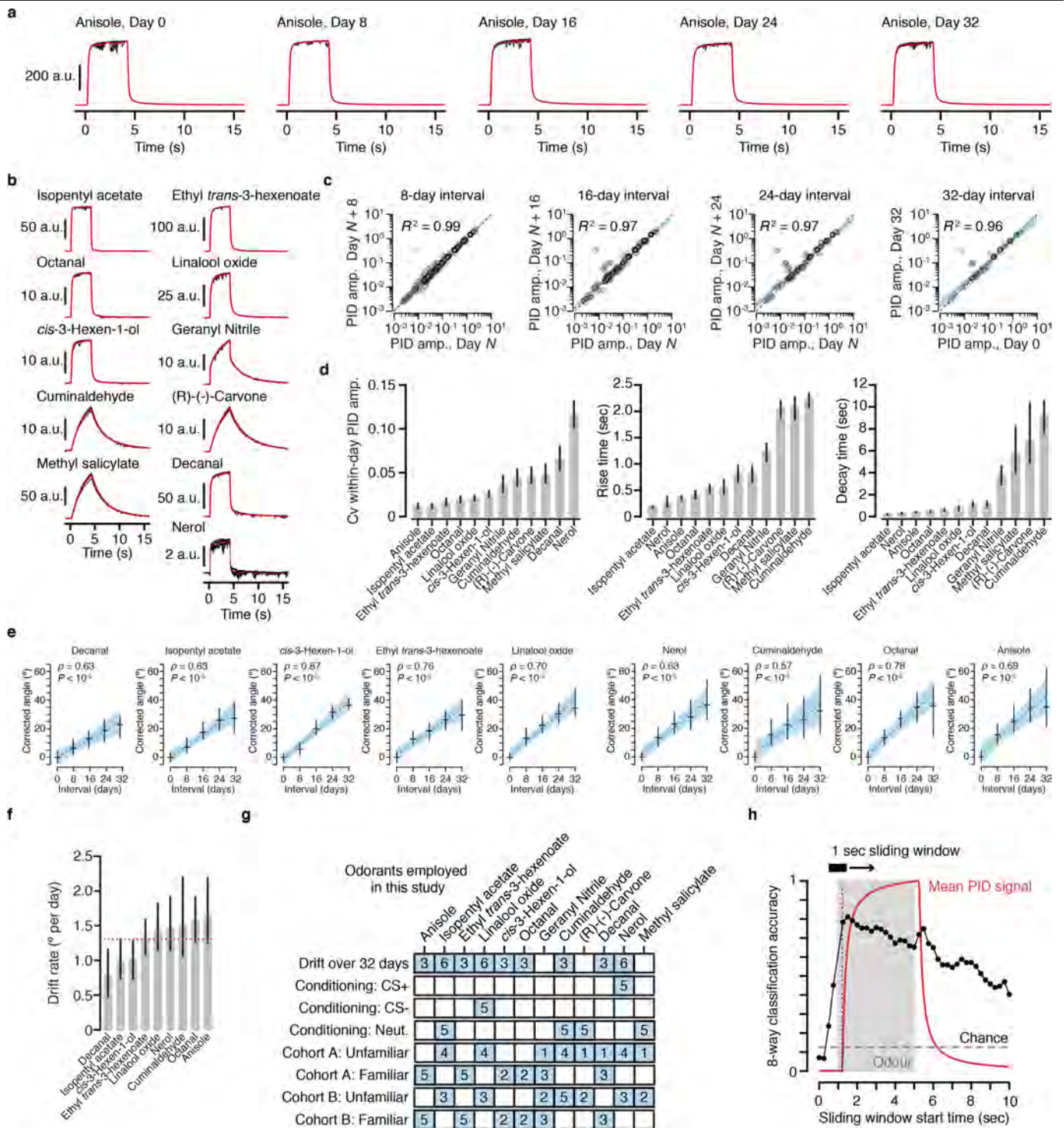
stimuli in the panel: Fig. 1b, left: unit 1, stimulus 7 (linalool oxide); Fig. 1b, middle: unit 2, stimulus 5 (nerol); Fig. 1b, right: unit 3, stimulus 2 (geranyl nitrile). **b**, z-scored odour-evoked activity of 300 randomly selected odour-unit pairs during the 4-s odourant stimulus epoch (black bars), ordered differently in each panel. Left, ordered by odour response magnitude on day 0 (replotted from Fig. 1c); middle, ordered by odour response magnitude on day 32; right, ordered by odour response magnitude individually on each day.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Selective odour responses with stable within-day statistics and across-day drift in piriform cortex. a, b, Fraction of single units that exhibit a significant modulation (**a**) and an increase (red) or decrease (blue) (**b**) in firing rate during the odourant stimulus epoch in response to 0–8 odourants (Wilcoxon rank-sum on firing rate during the odourant stimulus epoch versus spontaneous baseline firing rate, $\alpha = 0.001$). **c**, Cumulative distributions (left) and mean coefficient of variation (c_v ; right) of response magnitude computed on each odour test day across all trials for each odour–unit pair. Mean (95% CI) across days, $c_v = 0.88$ (0.87, 0.89), $n = 19,356$ odour–unit pairs. **d**, Cumulative distributions (left) and mean fraction of responses preserved per responsive single unit (right) across 8–32-day intervals (8 days: 35.0% (27.3%, 42.5%), 16 days: 19.8% (13.5%, 27.5%), 24 days: 16.9% (10.3%, 26.7%), 32 days: 6.6% (1.9%, 17.5%); $\rho = -0.25$, $P = 5 \times 10^{-6}$, $n = 318$ single units). Non-responsive and broadly responsive single units were excluded from the analysis by setting a threshold on lifetime sparseness (0.65). **e**, Left, fraction of preserved responses per single unit across 32 days versus lifetime sparseness threshold. Right, fraction of single units stable across 32 days versus lifetime sparseness threshold. A single unit was considered stable over 32 days if all significantly modulated responses to the odourant panel were preserved. These quantities do not depend on lifetime sparseness threshold (0.2–0.65, 40th–95th percentile across all single units). **f**, Classification accuracy (8-way, SVM, linear kernel, L2 regularization, trained and tested on data stitched across 3 mice, random draws of 231 single unit subsets from 286 total single units to avoid saturation, 1-s sliding window, 250-ms steps). Grey box, 4-s odourant stimulus epoch; vertical dotted line, onset of odour response at odour port (mean time across all stimuli at which the PID signal reached 5% of maximum); horizontal dashed line, chance performance for 8-way classification. **g**, Classification performance for fifteen temporal binnings of the odour response epoch, measured by maximum classification accuracy (top) and number of single units required to reach 50% of maximum accuracy (bottom) ($n = 286$ single units recorded within-day, stitched across 3 mice). Black shading, binning used for all subsequent classification and for computation of pairwise population vector correlations, angles and drift rate, unless otherwise indicated. For classification using single bins, the window start was set to 500 ms after stimulus initiation so the quantification windows did not begin before odourant stimulus onset as measured by the PID signal. **h**, Classification accuracy as a function of number of single units used, using the highest performance binning in **g** (four 2-s bins). Dashed arrow, number (21) of single units required to achieve >50% classification accuracy. **i**, Classification

accuracy for a classifier trained on earlier days and tested on later days ('Forward', replotted from Fig. 2b) compared with a model trained on later days and tested on earlier day ('Reverse'). Dotted lines, mean; shading, s.d.; limit of 41 single units per animal with 100 permutations. **j**, Classification accuracy of a classifier trained on responses on day 24 alone (all 56 trials) and tested on day 32 compared with a model trained on 75 random subsets of 56 trials drawn from days 0–24 and tested on day 32; $P = 2.6 \times 10^{-5}$, Wilcoxon rank-sum, 100 random subsets of 23 single units per mouse. A classifier trained on concatenated data from days 0–24 will assign high weights to single units with stable (less variable) responses across all days and low weights to single units whose responses varied. Thus, if there is a special population of neurons whose responses are informative about stimulus class and are more stable than others, a model trained on a concatenation of days 0 through 24 ought to perform better when tested on day 32 than a model trained on day 24 alone. However, we do not observe this: thus, it is not possible to establish single units that are most informative about odour identity on day 32 based on their responses across days 0–24. This finding argues against the presence of an informative stable subpopulation. **k**, Representational drift between a pair of days can be estimated by measuring the difference in odour-evoked population responses across days after correcting for within-day variability²⁵. Top left, variability across days (across-day drift + within-day variability), estimated by computing the angle ($\theta_{p,q}$) between trial-averaged population vectors \mathbf{u}_p and \mathbf{u}_q for each odour across each pair of days p and q . Bottom left, variability within a day (noise), estimated by measuring the mean of the angle between the trial-averaged population vectors ($\bar{\theta}$) for each odour within each day on odd trials versus even trials (θ_k , over all days k). Right, the drift rate ($r_{p,q}$) is the corrected angle ($\theta_{p,q} - \bar{\theta}$) divided by the time between days p and q ($\Delta t_{p,q}$). **l**, Cumulative distributions (left) and mean angles (right) between trial-averaged population vectors within-day and across 8–32-day intervals ($n = 180$, $n = 144$, $n = 108$, $n = 72$ and $n = 36$ pairs, respectively). Blue dotted line, exponential regression fit with $\theta = C - (C - R) e^{-t/\tau}$, where θ is the variability (angle), C the asymptote, R the intercept at $t = 0$ (within-day variability), and τ the time constant of the exponential in days. The mean rate of change of the exponential fit over the 32-day interval is 1.0° per day. **m**, Cumulative distributions (left) and mean within-day angles (right) between trial-averaged population vectors ($n = 72$ pairs per day). No pair of within-day angles differs significantly ($P \geq 0.56$ for all pairs, Wilcoxon rank-sum). Black crosses, mean \pm 95% CI; blue dotted line, linear regression; blue shading, 95% CI. Classification performed on the three mice presented with an eight-odourant panel. Otherwise, $n = 6$ mice.

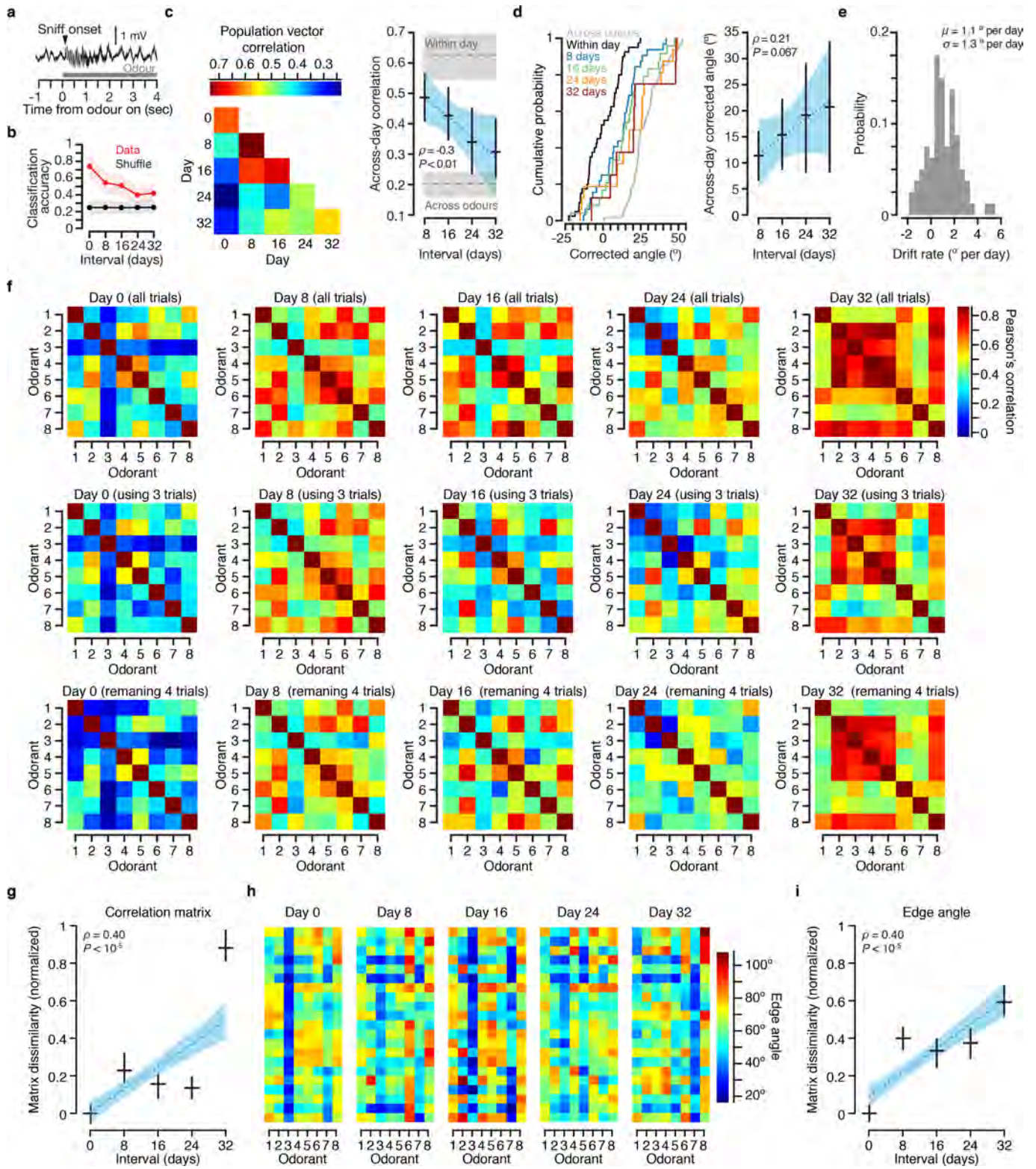


Extended Data Fig. 6 | See next page for caption.

Article

Extended Data Fig. 6 | Odorant stimuli. **a**, PID signals for an example odorant (anisole) across a 32-day interval from a single experiment. Black traces, individual trials; red traces, within-day mean ($n = 7$ trials). **b**, Example PID traces for all other odorant stimuli used in this study. Black traces, individual trials on one day; red, mean ($n = 7$ trials). **c**, Mean PID amplitude across all intervals for experiments in which odorants were presented every 8 days. PID amplitude for a given odorant stimulus is highly correlated across sessions (8-day interval, $\rho = 0.99$, $P = 6.9 \times 10^{-276}$, $n = 288$ comparisons; 16-day interval, $\rho = 0.99$, $P = 7.9 \times 10^{-166}$, $n = 208$ comparisons; 24-day interval, $\rho = 0.98$, $P = 6.2 \times 10^{-94}$, $n = 128$ comparisons; 32-day interval, $\rho = 0.98$, $P = 3.7 \times 10^{-46}$, $n = 64$ comparisons). **d**, Left, coefficient of variation of PID amplitude across all trials for each odorant for experiments in which odorants were presented every 8 days (median across all odorants: $c_v = 0.02$, (Q1 = 0.01, Q3 = 0.04), $n = 400$ mean odorant stimulus pulses computed across 2,800 individual trials for 12 distinct odorants). Middle, rise time (median across all odorants 0.47 s (Q1 = 0.21, Q3 = 1.0 s)). Right, decay time of PID signal (median across odorants 0.51 s (Q1 = 0.21 s, Q3 = 1.3 s)). PID signal onset was defined as the time required to reach 5% of maximum on each trial. PID rise time was defined as the time between onset and 66% of maximum on each trial; PID decay time was defined as the time between 95% and 33% of maximum after stimulus offset. Anisole $n = 350$ trials, 50 days; isopentyl acetate $n = 231$ trials, 33 days; ethyl *trans*-3-hexenoate $n = 350$ trials, 50 days; octanal $n = 245$ trials, 35 days; linalool oxide 231 trials, 33 days; *cis*-3-hexen-1-ol $n = 245$ trials, 35 days; geranyl nitrile $n = 168$ trials, 24 days; cuminaldehyde $n = 168$ trials, 24 days; R(-)-carvone $n = 63$ trials, 9 days; methyl salicylate $n = 63$ trials, 9 days; decanal $n = 231$ trials, 33 days; nerol $n = 231$ trials, 33 days. Grey bars, mean across all experiments by odorant; black bars, 95% CI. **e**, Corrected angle as a function of interval for each odorant

stimulus used in 32-day interval experiments. Black crosses, mean \pm 95% CI. Decanal: corrected angle correlation across intervals, $\rho = 0.63$, $P = 3.5 \times 10^{-6}$ and drift rate, $0.8 (0.5 - 1.1)^\circ$ per day ($n = 45$ population vector pairs from 3 mice). Isopentyl acetate: corrected angle correlation across intervals, $\rho = 0.63$, $P = 2.3 \times 10^{-11}$ and drift rate, $1.0 (0.8 - 1.3)^\circ$ per day ($n = 90$ population vector pairs from 6 mice). *cis*-3-Hexen-1-ol: corrected angle correlation across intervals, $\rho = 0.87$, $P = 1.7 \times 10^{-14}$ and drift rate, $1.0 (0.7 - 1.3)^\circ$ per day ($n = 45$ population vector pairs from 3 mice). Ethyl *trans*-3-hexenoate: corrected angle correlation across intervals, $\rho = 0.766$, $P = 1.3 \times 10^{-9}$ and drift rate, $1.3 (1.1 - 1.6)^\circ$ per day ($n = 45$ population vector pairs from 3 mice). Linalool oxide: corrected angle correlation across intervals, $\rho = 0.70$, $P = 1.7 \times 10^{-14}$ and drift rate, $1.5 (1.2 - 1.8)^\circ$ per day ($n = 90$ population vector pairs from 6 mice). Nerol: corrected angle correlation across intervals, $\rho = 0.78$, $P = 2.6 \times 10^{-10}$ and drift rate, $1.5 (1.1 - 1.9)^\circ$ per day ($n = 90$ population vector pairs from 6 mice). Cuminaldehyde: corrected angle correlation across intervals, $\rho = 0.57$, $P = 4.2 \times 10^{-5}$ and drift rate, $1.5 (1.1 - 2.2)^\circ$ per day ($n = 45$ population vector pairs from 3 mice). Octanal: corrected angle correlation across intervals, $\rho = 0.78$, $P = 2.6 \times 10^{-10}$ and drift rate, $1.6 (1.3 - 1.9)^\circ$ per day ($n = 45$ population vector pairs from 3 mice). Anisole: corrected angle correlation across intervals, $\rho = 0.69$, $P = 1.7 \times 10^{-7}$ and drift rate, $1.6 (1.2 - 2.1)^\circ$ per day ($n = 45$ population vector pairs from 3 mice). **f**, Drift rate for each odorant stimulus used in 32-day interval experiments. Red dotted line, mean drift rate across all experiments (from distribution in Fig. 2e). **g**, Odorants used in this study. Numbers indicate the number of experimental replicates in which each odorant molecule was used for each experiment type. **h**, Mean, normalized PID signal recorded simultaneous to the neural signals that were used to classify odorant stimuli using a linear SVM (superimposed, reproduced from Extended Data Fig. 5f).



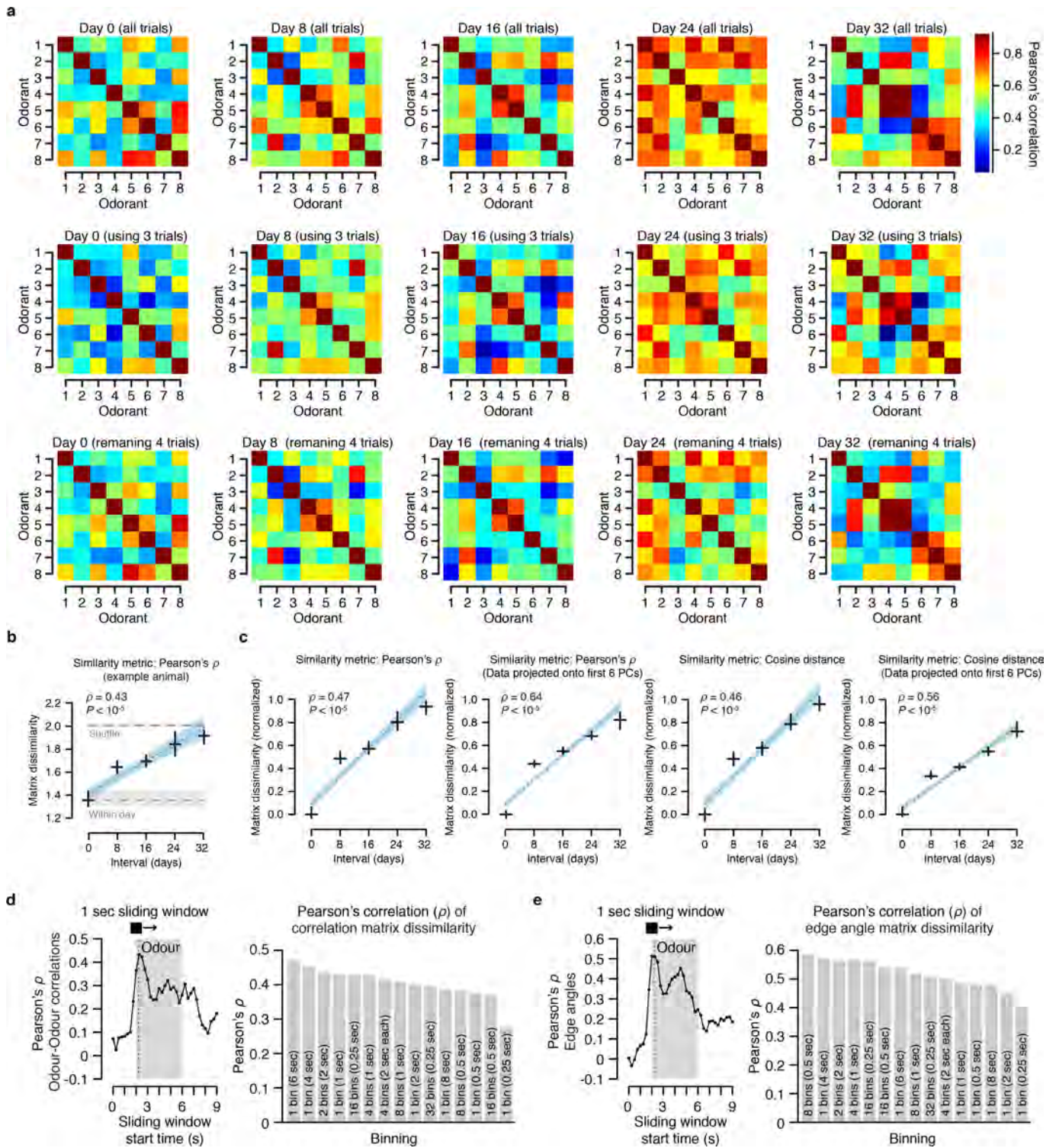
Extended Data Fig. 7 | See next page for caption.

Article

Extended Data Fig. 7 | Drift during the early phase of the odour response.

a, Mean local field potential over all 32 electrodes (filtered 0.1–20 Hz) from an example trial. We estimated the time of first sniff onset following stimulus onset by detecting the first peak of this oscillation on each trial. Only spikes that occurred within the 190 ms after and 10 ms before the detection of sniff onset were analysed (first sniff epoch). Arrowhead, estimated sniff onset. Grey bar, 4-s odorant stimulus epoch. **b**, Red, classification accuracy (4-way, SVM, linear kernel, L2 regularization) of single-trial z-scored population vectors as a function of interval using only the 200-ms window during the first sniff (as estimated by the first peak in the local field potential) in this mouse ($n = 71$ single units). Classification performance on day 0 computed using leave-one-out cross-validation. For all other intervals the model was trained on all

responses from the earlier day and tested on all responses from the later day, as in Fig. 2b. Black, performance with stimulus labels shuffled; mean \pm s.d. **c–e**, As in Fig. 2c–e, but for only the first sniff epoch in this mouse (8-day interval, $n = 32$ pairs of trial-averaged population vectors; 16-day interval, $n = 24$ pairs; 24-day interval, $n = 16$ pairs; 32-day interval, $n = 8$ pairs; within day, all days $n = 40$ pairs). **f**, Top, odour–odour correlation matrices computed on each day during the first sniff epoch in this mouse; middle, bottom, matrices computed using complementary splits of the trials recorded on each day. **g**, Correlation matrix dissimilarity (scaled Frobenius norm; see Methods). **h**, **i**, As in Fig. 3b, c but for only the first sniff epoch in this mouse. **h**, Edge angle matrices; **i**, edge angle matrix dissimilarity (scaled Frobenius norm; see Methods). Black crosses, mean \pm 95% CI; blue dotted line, linear regression; blue shading, 95% CI.

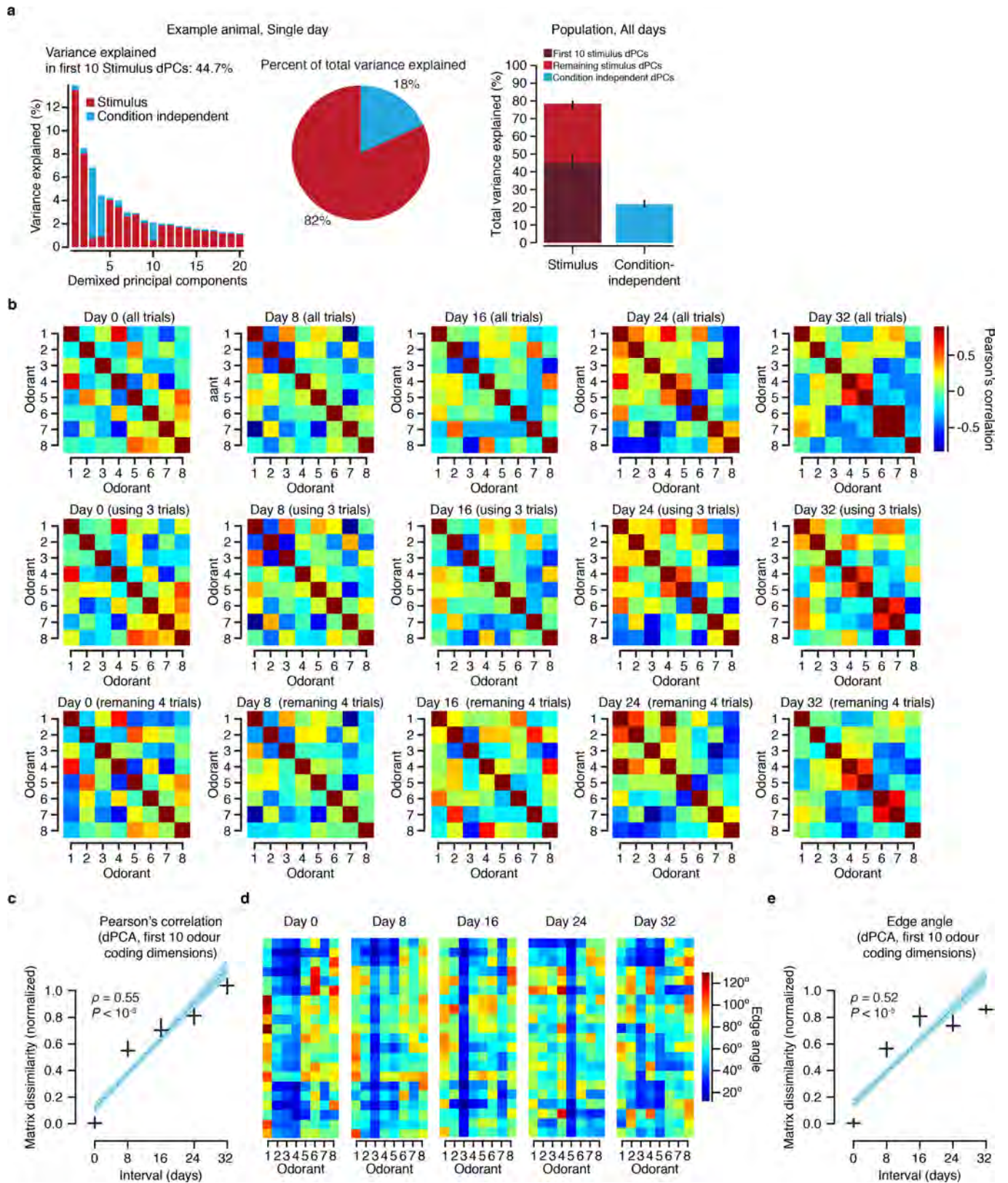


Extended Data Fig. 8 | See next page for caption.

Article

Extended Data Fig. 8 | Drift in response geometry under diverse similarity metrics and temporal binnings. **a**, Top, odour–odour correlation matrices computed within each odour test day using all trials (from same mouse as in Fig. 3b); middle, bottom, computed using splits of three trials per stimulus (middle) and the complementary four trials per stimulus (bottom) recorded on that day. Odour–odour correlation matrices were computed from trial-averaged population response vectors to eight odorant stimuli recorded on each of the five days. **b**, Correlation matrix dissimilarity (Frobenius norm; see Methods) from the same mouse as in **a** (mean \pm 95% CI). Grey dashed line, top, matrix dissimilarity computed using shuffled stimulus identities; bottom, matrix dissimilarity computed within all individual days; grey shading, 95% CI. $\rho = 0.43$, $P = 4.0 \times 10^{-25}$. Both within- and across-day matrix dissimilarity were calculated using odour–odour correlation matrices based on trial-averaged population vectors taken across all combinations of 3-trial/4-trial splits of the data for a given odorant on a given session. Within-day differences in odour response geometry measures (odour–odour correlation matrix dissimilarity) are significantly lower than differences measured across days ($P < 1.2 \times 10^{-34}$ for all measures and intervals from all three mice that were presented with a panel of eight stimuli). **c**, Correlation matrix dissimilarity averaged across all three mice, after scaling results from each subject between 0 (mean within-day

matrix dissimilarity) and 1 (mean shuffle matrix dissimilarity). Blue dotted line, linear regression; blue shading, 95% CI. Within-day, 525 pairs; across days, 1,050 pairs. Correlation matrix dissimilarity increases significantly for: left, raw odour–odour correlations; second from left, odour–odour correlations computed using responses projected onto the data's first six principal components (computed separately on each day); second from right, raw cosine distances; and right, cosine distances computed using responses projected onto the data's first six principal components (computed separately on each day). **d**, Left, correlation coefficients of normalized matrix dissimilarity versus time interval (using odour–odour correlations measured in the full neural space; that is, Pearson's correlation ρ such as that reported in **c**, left), computed at 250-ms steps using a 1-s sliding window along the odour response epoch. Grey box, 4-s odorant stimulus epoch; vertical dotted line, onset of odour response at odour port (mean time across all stimuli at which the PID signal reached 5% of max). Right, correlation coefficients for normalized matrix dissimilarity versus time interval computed using 15 temporal binnings of the odour response epoch. **e**, As in **d** but using edge angles rather than correlations. This effect also holds when using responses projected onto the data's first six principal components, as well as using cosine distance rather than Pearson's correlation (data not shown).

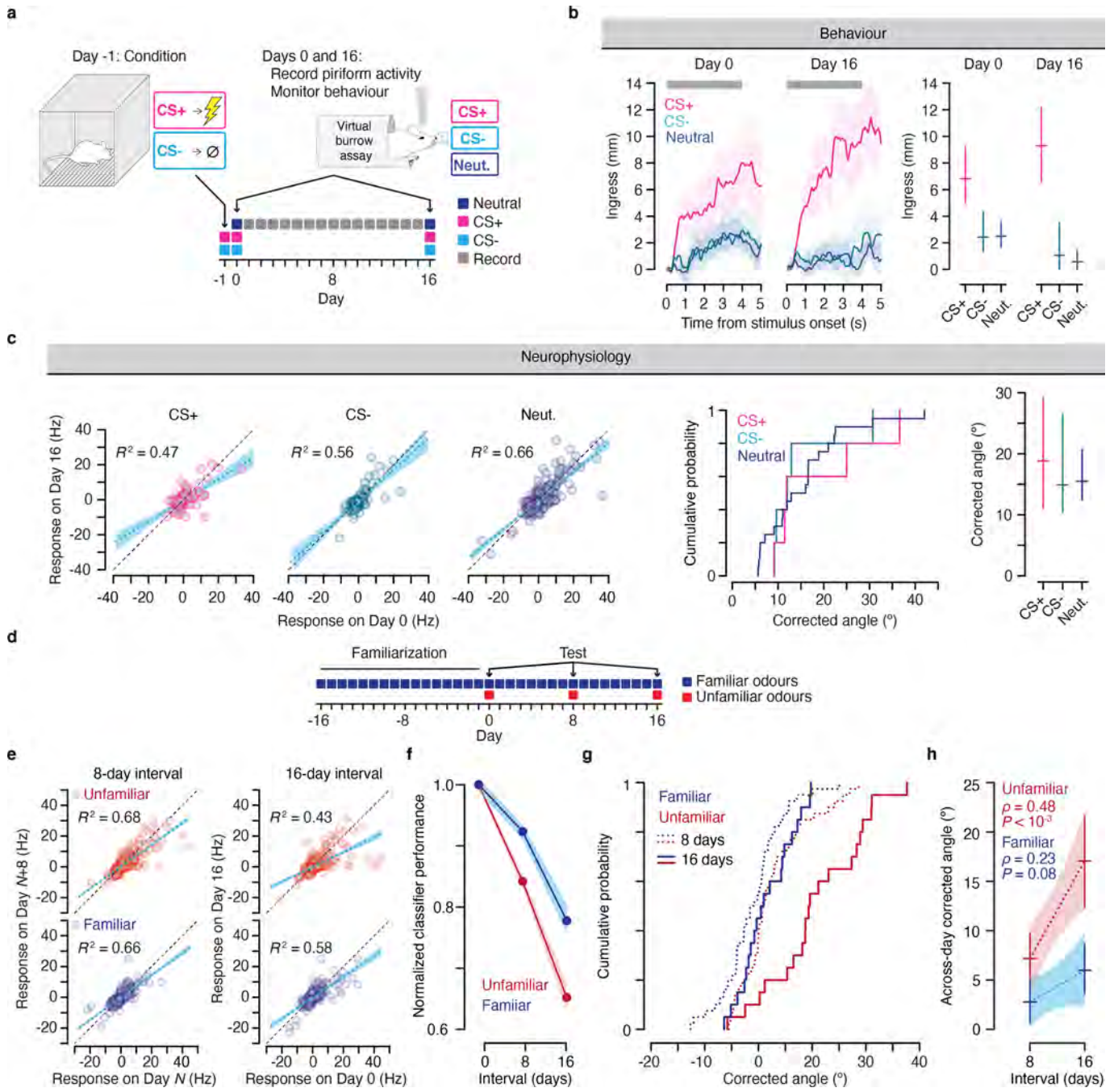


Extended Data Fig. 9 | See next page for caption.

Article

Extended Data Fig. 9 | Drift in response geometry in the odour coding subspace. **a**, Left, percentage of total variance explained by each demixed principal component (dPC)²⁸ from an example mouse; middle, fraction of total stimulus (red) and condition-independent (blue) variance explained for this example mouse; right, total variance explained by stimulus dPCs (red, 78.4% (95% CI 76.0%, 79.8%)) and condition-independent dPCs (blue, 21.6% (95% CI 20.2%, 23.7%)) for the $n = 3$ mice shown a panel of 8 odorant stimuli. Dark red, variance explained by the first ten dPCs that were primarily stimulus coding (45% (41.8%, 48.6%)). Error bars, 95% CI. **b**, Odour-odour correlation matrices from the example mouse computed after projecting responses onto the data's

first ten stimulus-coding dPCs, computed separately on each day (top row) and then (bottom two rows) using complementary splits of the trials recorded on each day. **c**, Correlation matrix dissimilarity (scaled Frobenius norm; see Methods) for all mice, using responses projected onto the data's first ten stimulus-coding dPCs. **d**, Edge angle matrices from the example mouse computed after projecting responses onto the data's first ten stimulus-coding dPCs. **e**, Across-day edge angle matrix dissimilarity (scaled Frobenius norm; see Methods) for all mice, using responses projected onto the data's first ten stimulus-coding dPCs. Black crosses, mean \pm 95% CI; blue dotted line, linear regression; blue shading, 95% CI.



Extended Data Fig. 10 | See next page for caption.

Article

Extended Data Fig. 10 | Effect of fear conditioning and familiarity on drift.

a, Conditioning experiment. Day -1: present one odour paired with shock (CS+) and a second without shock (CS-) in a conditioning chamber. Days 0 and 16: administer conditioned (CS+ and CS-) and four additional neutral odours to head-fixed mouse while recording neural signals and measuring behavioural responses in a virtual burrow assay²⁹. Days 1-15: record neural signals in head-fixed mouse without test odour administration. **b**, Behaviour. Left, trial-averaged ingress amplitude ($n = 5$ mice) across time on days 0 and 16 on trial blocks 2-7 (shading, 95% CI). Grey bar, 4-s odour stimulus epoch. Right, mean \pm 95% CI ingress amplitude during the final second of the odour epoch on blocks 2-7. For days 0 and 16, CS+ versus CS- and CS+ versus neutral, $P < 1.4 \times 10^{-3}$, Wilcoxon rank-sum. **c**, Neurophysiology. Left, scatter plots showing single-unit response magnitude for all three stimulus classes (mean spontaneous baseline-subtracted evoked responses computed during the odour stimulus epoch) of odour-unit pairs on day 0 versus day 16 (CS+: $n = 148$ odour-unit pairs, CS-: $n = 129$ odour-unit pairs, neutral stimuli: $n = 482$ odour-unit pairs, data pooled across 5 mice). Black dashed line, unity; blue dotted line, linear regression; blue shading, 95% CI. Regression was performed across all odour-unit pairs that showed a significantly modulated response on at least one of the two days (Wilcoxon rank-sum, $\alpha = 0.001$). Middle, cumulative distributions; right, mean \pm 95% CI ($n = 5$ mice) of corrected angles for all three classes of stimulus. For all comparisons, $P > 0.05$ (Wilcoxon rank-sum). We note

that classical conditioning reduces within-day variability (unpublished observations). Thus, the odour-unit pair response correlations reported here, which are not corrected for within-day variability, are higher than in other experiments (for example, Fig. 2a), but measures that correct for within-day variability, such as corrected angle or drift rate, are comparable. **d**, Familiarity experiment. Mice were presented with a panel of four neutral odours daily over a 32-day interval (days -16 to 16; familiar). Starting on day 0, a panel of unfamiliar odours was presented at 8-day intervals. **e**, Mean odour-evoked response magnitude (spontaneous baseline-subtracted, computed during the odour stimulus epoch) of odour-unit pairs across intervals of 8 days (left, familiar odours: $n = 741$ odour-unit pairs, unfamiliar odours: $n = 1,137$ odour-unit pairs) and 16 days (right, familiar odours: $n = 371$ odour-unit pairs, unfamiliar odours: $n = 570$ odour-unit pairs), data pooled across 5 mice. Black dashed line, unity; blue dotted line, linear regression; blue shading, 95% CI. Regression was performed across all odour-unit pairs that showed a significantly modulated response on at least one of the two days (Wilcoxon rank-sum, $\alpha = 0.001$). **f**, Across-day classification accuracy (4-way, SVM, linear kernel, L2 regularization, scaled between chance and maximum within-day performance to account for slight differences in within-day performance between the two conditions). Solid lines, mean; shading, 95% CI. **g, h**, Cumulative distribution (**g**) and mean \pm 95% CI (**h**) of corrected angles from $n = 5$ mice. Unfamiliar, $\rho = 0.48$, $P = 1.0 \times 10^{-4}$; familiar, $\rho = 0.23$, $P = 0.08$.

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 our [Editorial Policies](#) 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 | Physiological data were acquired using Blackrock Cerebus Central Suite. Behavioral data were acquired using Python/NI-DAQmx/PyDAQmx and VBMcmd (github.com/goatsofnaxos/VBAcmd). Animal preparation images acquired using the dissection stereo microscope were captured with Leica Application Suite. Histological microphotographs acquired using the epifluorescence microscope were captured with ZEN Imaging Software. |
| Data analysis | Spike sorting was performed using KiloSort (github.com/cortex-lab/KiloSort) and manual curation was performed using Phy 1 (github.com/cortex-lab/phy). Statistical analysis software was written in MATLAB R2016b. The LIBLINEAR 2.3 package (www.csie.ntu.edu.tw/~cjlin/liblinear/) was employed for support vector machine analyses. Auto- and cross-correlograms were computed using spikes (github.com/cortex-lab/spikes). Demixed principal component analysis was performed using dPCA (github.com/machenslab/dPCA). |

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 and 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 data are available upon reasonable request.

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 size was not predetermined. We established that our sample sizes are sufficient based on the size and statistical significance of the effects. |
| Data exclusions | Exclusion criteria are detailed in the Methods. Briefly, templates that fell below criteria for isolation (refractory period violation >2%) or stability (across-day waveform correlation <0.93) were excluded from statistical analyses. Animals whose probe implantations produced either low single unit yield or unstable single unit waveforms were not analyzed. |
| Replication | The effects we report were observed in every subject. |
| Randomization | Randomization is not relevant to this study as comparisons were not made across groups. |
| Blinding | Blinding is not relevant to this study as manual curation of single unit templates was done without knowledge of stimulus responsiveness, and behavior was analyzed automatically without manual scoring. |

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

Methods

| n/a | Involved in the study | n/a | Involved in the study |
|-------------------------------------|---|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies | <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines | <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology | <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |
| <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 | | |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern | | |

Animals and other organisms

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

| | |
|-------------------------|---|
| Laboratory animals | Male C57BL/6J mice, 10-17 week old. |
| Wild animals | This study did not involve wild animals. |
| Field-collected samples | This study did not involve field-collected samples. |
| Ethics oversight | All procedures were approved by the Columbia University Institutional Animal Care and Use Committee (protocol AC-AAAT5466) and were performed in compliance with the ethical regulations of Columbia University as well as the Guide for Animal Care and Use of Laboratory Animals. |

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