Survey of spiking in the mouse visual system reveals functional hierarchy

A full list of authors and affiliations appears at the end of the article.

Abstract

The anatomy of the mammalian visual system, from the retina to the neocortex, is organized hierarchically[1]. However, direct observation of cellular-level functional interactions across this hierarchy is lacking due to the challenge of simultaneously recording activity across numerous regions. Here we describe a large, open dataset—part of the Allen Brain Observatory[2]—that surveys spiking from tens of thousands of units in six cortical and two thalamic regions in the brains of mice responding to a battery of visual stimuli. Using cross-correlation analysis, we reveal that the organization of inter-area functional connectivity during visual stimulation mirrors the anatomical hierarchy from the Allen Mouse Brain Connectivity Atlas[3]. We find that four classical hierarchical measures—response latency, receptive-field size, phase-locking to drifting gratings and response decay timescale—are all correlated with the hierarchy. Moreover, recordings obtained during a visual task reveal that the correlation between neural activity and behavioural choice also increases along the hierarchy. Our study provides a foundation for understanding coding and signal propagation across hierarchically organized cortical and thalamic visual areas.

Mammalian vision is the most widely studied sensory modality. The investigation of its cellular substrate has yielded insights into how the stream of photons that impinge onto the retina leads to conscious perception and visuomotor behaviours. However, much of our knowledge of physiology at the cellular level derives from small-scale studies that are subject to substantial uncontrolled variation, uneven coverage of neurons and selective use

of stimuli. The ability to validate models of visual function has been hampered by the absence of large-scale, standardized and open in vivo physiology datasets[4,5]. To address this shortcoming, we previously developed a two-photon optical physiological pipeline to systematically survey visual responses in genetically defined cell populations[2]. However, this methodology lacks the ability to record simultaneously with high temporal resolution across many cortical and subcortical structures. We therefore built a complementary pipeline that uses Neuropixels probes[6] to measure spiking activity in six cortical visual areas as well as two visual thalamic nuclei: the lateral geniculate nucleus (LGN) and the lateral posterior nucleus (LP), also known as the visual pulvinar.

The concept of hierarchy has informed ideas about the architecture of the mammalian visual system for more than 50 years[7], and has inspired powerful multi-layered computational networks[8-10]. The visual hierarchy has been investigated most extensively in the macaque, from the LGN and the primary visual cortex (V1) into frontal eye fields and beyond[1,11-16]. The existence of such a hierarchy in the mouse, with its far smaller brain and densely connected cortical network[17], is less clear[18-20]. Yet, given the utility of the mouse model, characterizing the presence and extent of such a hierarchy is important.

By analysing anterograde viral tracing with Cre-dependent adeno-associated viruses (AAV) from 1,256 mice, anatomical rules were previously derived to describe projections into and out of 37 cortical and 24 thalamic regions via their layer-specific axonal termination patterns[3]. An optimization algorithm assigned a hierarchy score to every region to reveal a hierarchical ordering of visual areas, with the LGN at the bottom and the higher-order cortical region, antero-medial area (AM), at the top. However, the importance of this anatomical hierarchy is unclear. Functional activity is dynamic and context-dependent, so it is unknown how well the flow of spikes follows the anatomical hierarchy, especially given the presence of all-to-all connectivity[17,19] and branching connections[21]. Therefore, we sought to determine whether the anatomical hierarchy is reflected in the spiking activity of these visual areas, linking hierarchical structure to function.

A survey of visually evoked spiking

We recorded spiking activity across visual cortical and thalamic structures in awake, head-fixed mice viewing diverse visual stimuli, using Neuropixels silicon probes[6] to simultaneously record from hundreds of neurons with high spatial and temporal resolution[22-24]. This dataset of about 100,000 units complements our previously released survey that used optical recordings of calcium-evoked fluorescent activity in 60,000 cortical neurons[2] (see ref. [25] for a comparison of the results from the imaging and electrophysiology datasets). Both datasets are part of the Allen Brain Observatory—a pipeline of animal husbandry, surgical procedures, equipment and standard operating procedures, coupled to strict activity- and operator-independent quality-control measures. All physiological data that passes quality control is made freely and publicly available via the AllenSDK (https://allensdk.readthedocs.io), the DANDI Archive (https://gui.dandiarchive.org) and the AWS Registry of Open Data (https://registry.opendata.aws/allen-brain-observatory/).

Each mouse in this study proceeded through an identical series of steps, carried out by highly trained staff according to a set of standard operating procedures (Fig. 1a, Extended Data Fig. 1a-f; see also http://help.brain-map.org/display/observatory/Documentation). We used cortical area maps derived from intrinsic signal imaging of every mouse to simultaneously target up to six Neuropixels probes to V1 and five higher-order visual cortical areas (latero-medial area (LM), anterol-ateral area (AL), rostro-lateral area (RL), postero-medial area (PM) and AM) (Extended Data Fig. 1g-i). The probes were inserted up to 3.5 mm into the brain to measure responses in the LGN and the LP thalamic areas (Fig. 1b); the hippocampus and other areas traversed by the silicon probes were likewise recorded. This configuration enabled us to sample the mouse visual system with unprecedented coverage, creating cellular-resolution activity maps across up to eight cortical and thalamic visual areas at once (Fig. 1c).

We implemented quality-control procedures to ensure consistent data (Methods, Extended Data Fig. 2), reducing the number of experiments analysed and presented here from 87 to 58. Extracellularly recorded units were sorted via the Kilosort2 algorithm[24,26] and further subjected to quality control (Extended Data Figs. 3, 4). Units were mapped to structures in the Common Coordinate Framework Version 3, a 3D anatomical atlas[27], by imaging fluorescent probe tracks with optical projection tomography (Extended Data Fig. 5). Overall, we recorded 682 ± 144 units per experiment, 119 ± 48 units per probe and 56 ± 30 units per visual area (Extended Data Fig. 1j), sampling 6.1 ± 1.1 visual areas per experiment (Extended Data Fig. 1k).

During each recording session, mice passively viewed a battery of natural and artificial stimuli (Extended Data Fig. 6a-c). Here we focus on a subset of these—including drifting gratings (Fig. 1d), full-field flashes (Fig. 1e) and local Gabor patches (Fig. 1f)—to characterize aspects of hierarchical processing. Units recorded in all eight cortical and thalamic visual areas were highly visually responsive, with 60% displaying significant spatial receptive fields within the boundaries of the monitor used for stimulus presentation (Fig. 1g, Extended Data Fig. 6d, e; categorical χ^2 test, $P < 0.01$). As a control, we searched for significant receptive fields in simultaneously recorded hippocampal regions (CA1, CA3 and dentate gyrus), and found them in only 1.4% of units.

A functional hierarchy of visual areas

A previous anatomical study[3] assigned a hierarchy score to each cortical and thalamic region in the mouse, derived using an optimization algorithm that considers the set of distinct axonal termination patterns of connectivity between areas (deeming each as either a feedforward or a feedback connection), and found the most self-consistent network architecture out of the set of hierarchical area orderings (Fig. 2a). The LGN sits at the bottom of the hierarchy, followed by its major target structure, V1; areas LM, RL, LP and AL reside at intermediate levels, and areas PM and AM occupy the top level of the areas we studied here. The higher-order thalamic area, LP, is interconnected with all visual cortical regions, and resides at an intermediate hierarchical location.

During strong bottom-up, visual stimulation, we anticipated that activity would propagate up this anatomical hierarchy. The directionality of this bottom-up wave of activity should be visible in pairwise leader-follower relationships between connected areas. To test for such a functional hierarchy, we evaluated the directed functional connectivity using spike cross-correlograms (CCG) between units in different areas[28-30] during visual stimulation with drifting gratings. For each pair of recorded units, we examined whether a functional connection was present in the CCG, defined as a 'sharp peak' with a short latency (within ± 10 ms) and a large peak amplitude (more than 7-fold greater than the CCG flank standard deviation; see Methods for details) in the jitter-corrected CCGs (Fig. 2c). Jitter correction removes slow timescale correlations larger than the jitter window (25 ms), yielding 16,119 pairs of units out of 2,089,890 possible pairs within the cortex (Extended Data Fig. 7b; $0.96\% \pm 0.13\%$ per mouse, $n = 25$ mice). These fast-timescale interactions sample the functional hierarchy between areas (see Fig. 2c for an example pair). If spikes in the source area lead spikes in the target area, the distribution of peak offsets will deviate in the positive direction from 0. For example, the peak offset distribution between V1 and LM showed a significant positive delay compared to the V1–V1 distribution (Fig. 2d; $P = 2.6 \times 10^{-8}$, two-sided Wilcoxon rank-sum test), indicating that V1 neurons—on average—lead LM neurons during strong visual drive, and thus are lower in the functional hierarchy.

We computed the distribution of CCG sharp peak time lags for all functionally connected units across each pair of cortical areas in each mouse, and combined the median of peak offset distributions across mice (Extended Data Fig. 7c, d; see Extended Data Fig. 7a for complete peak offset distributions between all areas across all mice). On average, V1 units lead the activity of units in other areas (Extended Data Fig. 7c, left column); by contrast, area AM follows other regions, indicating this area resides at the uppermost levels of the hierarchy (Extended Data Fig. 7c, right column).

To assess leader–follower relationships between areas, we defined a directionality score that quantifies the relative number of positive and negative time lag connections between any two areas (see Methods). The matrix of pairwise directionality scores (Fig. 2e) between areas was very similar to the matrix of anatomical hierarchy score differences (Fig. 2b) (Pearson's $r = 0.74$, $P = 1 \times 10^{-4}$; see also Extended Data Fig. 7c). The spatial layout of areas could not account for this correlation in terms of average physical distance (Extended Data Fig. 7e). Furthermore, this organization was absent during spontaneous activity (Extended Data Fig. 7f, g), suggesting that the functional hierarchy we identified reflects population activity driven by bottom-up input. Network simulations of simple architectures ranging from completely parallel to purely hierarchical organizations suggest that our empirical CCG observations are most consistent with a ladder-like hierarchy with abundant feedback (Extended Data Fig. 8).

We next assessed how this ordering of areas correlated with four classical measures of functional hierarchy[11,12,15]. First, we quantified the temporal latency of evoked responses to full-field flashes. Although units in each visual area have broadly distributed onsets (Fig. 3a, b), which is consistent with results in primates[15], the mean visual latency of each area was correlated with its anatomical hierarchy score (Fig. 3c; Pearson's $r = 0.95$, $P = 0.00025$). Statistical testing revealed significantly different latencies for all pairs of areas,

except for LGN–V1, RL–LP, LP–AL and AM–PM (Extended Data Fig. 9a). Differences in spontaneous firing rates do not account for these differences in latency (Extended Data Fig. 9b, c).

Second, the size of spatial receptive fields typically increases when ascending the visual processing stream[20,31-33], which is probably due to the pooling of convergent inputs from lower regions. We measured receptive fields using a localized Gabor stimulus (Fig. 3d), and found a systematic increase in receptive field size with anatomical hierarchy score (Fig. 3d-f; Pearson's $r = 0.97$, $P = 8.3 \times 10^{-5}$). Statistical testing revealed significantly different receptive field sizes for all pairs of areas, except for LM–RL (Extended Data Fig. 9d).

Third, the fraction of cells with phase-dependent grating responses is a useful measure of hierarchical level because it mirrors receptive field complexity[34]. We quantified this with a modulation index that reflects phase-dependent responses to drifting gratings[34,35]. The modulation index was highest in the LGN, whereas higher areas showed gradually less phase-dependent modulation (Fig. 3g-i, Pearson's $r = -0.89$, $P = 0.003$). Statistical tests revealed significantly different modulation indices for all pairs of areas, except for RL–AL and AM–PM (Extended Data Fig. 9e).

Finally, previous work in primate and mouse brains demonstrated that the 'timescale' of neural activity increases in the upper echelons of the hierarchy[12,13,36]. We assessed intrinsic timescale by fitting an exponential decay function to the spontaneous spike-count autocorrelation of each unit during grey screen periods between stimulus presentations. Whereas this mean intrinsic timescale for each area was not correlated with the visual hierarchy (Extended Data Fig. 9f, g) (Pearson's $r = -0.24$, $P = 0.57$), the response decay timescale[12], which is quantified by fitting an exponential decay function to the spike-count autocorrelation of individual units during the evoked response to the full-field flash stimulus (Fig. 3j), was. Higher-order areas had a longer response decay timescale, and therefore maintain stimulus-evoked activity over longer temporal windows, than lower stages—an important signature of multi-layer processing (Fig. 3j-l; Pearson's $r = 0.86$, $P = 0.007$). Statistical testing revealed significantly distinct response decay timescales for all pairs of areas, except for LM–AL and AM–PM (Extended Data Fig. 9h).

Together, these four response metrics—along with our cross-correlation analysis—support the existence of a functional hierarchy that spans the cortical and thalamic visual system. These metrics are not dependent on overall firing rate, which does not correlate with hierarchy score (Extended Data Fig. 9i, j). Because we densely sampled units across all cortical layers in each area, we were able to assess the layer-dependence of each of these metrics and found similar results (Extended Data Fig. 10a, b). Analysis of layer-wise CCG interactions indicated that superficial layers (2/3 and 4) were hierarchically lower compared to deep layers (5, 6) in the same area (Extended Data Fig. 10c-e).

The role of this hierarchy should ultimately be related to the behavioural and cognitive operations implemented by the system, because higher levels are better positioned to integrate sensory input with behavioural goals. To test whether the hierarchy we found correlates with behaviourally relevant processing, we measured spiking activity during a

visual change detection task ($n = 4{,}057$ units from 12 mice). In this go/no-go task, mice see briefly presented natural scenes (250 ms stimulus presentations, separated by 500 ms grey screen) (Fig. 4a, left). In each trial, a repeating 'reference' image changes identity after a random number of presentations, and mice are rewarded for detecting the change by licking a spout[37,38]. To assess hierarchical processing during active behaviour compared with passive stimulation, we separated each recording session into two blocks: first, the mice performed the behavioural task for 60 min; second, the lick spout was retracted and the same sequence of visual stimuli were presented to the mice under these passive viewing conditions (Fig. 4a, right).

Mice performed with high hit and low false alarm rates (mean hit rate = 0.78, mean false alarm rate = 0.13, and mean detection sensitivity (d') = 2.0 ± 0.1, in 12 mice, 21 sessions; Fig. 4b). Units recorded during the task had clear visually evoked spiking responses to the images and showed greater evoked spike rates when the stimulus changed identity (from A to B at $t = 0$ in Fig. 4c, d). Consistent with results described above for full-field flashes, the first spike latency for image responses during behaviour was correlated with the anatomical hierarchy score (Fig. 4e).

When humans and rats detect changes in a stream of stimuli, change detection signals increase at higher cortical levels (the oddball P300)[39-41]. To detect such mismatch signals in the mice in this experiment, we computed a 'change modulation index' (CMI) that captures the differential response to the same natural image when it was the reference (pre-change) compared with when it was the change image (see Methods). During active behaviour, CMI was positive for each area, which indicates that a change in image identity elicits stronger responses compared with presenting the same image repeatedly. More importantly, CMI systematically increased along the hierarchy from the LGN to the AM (Fig. 4f; Pearson's $r = 0.83$, $P = 0.011$). Consistent with a role in change perception, we found a significant correlation of CMI with hierarchy on hit but not on miss trials (Extended Data Fig. 9k; hit trials, Pearson's $r = 0.85$, $P = 0.007$; miss trials, Pearson's $r = 0.51$, $P = 0.2$). Moreover, CMI values were larger during active behaviour compared with passive viewing of the same stimulus sequence, which indicates that change signals cannot be explained solely by passive effects, such as adaptation to the repeated reference image (Fig. 4f). Other aspects of neural activity during the task—including the baseline firing rate, response to the pre-change image and the response to the change image—were not correlated with hierarchy score (Extended Data Fig. 9l-n).

To assess whether activity at higher levels more closely covaries with the decisions made by the mouse, we used random forest decoders trained on the spiking activity of units within individual areas to predict when the image either did or did not change (Fig. 4g, Extended Data Fig. 9o). Decoders were separately trained for 20 units within each area and performed significantly better than chance across trials, indicating that change trials could be read out from all 8 areas (Extended Data Fig. 9p). Notably, however, we found a strong increase in trial-wise decoder-behaviour covariation at higher levels of the hierarchy, starting with no correlation at the level of the LGN (Fig. 4h; Pearson's $r = 0.88$, $P = 0.004$; see Methods). In other words, change-related signals are amplified at higher levels of the visual hierarchy, and

spiking activity at these stages is more correlated with behavioural choices, which suggests that hierarchical processing is relevant for behaviour.

Discussion

One long-term goal of the Allen Institute is to systematically survey neuronal activity in a way that is minimally biased, maximally reproducible and freely accessible to all[42]. Here we add to our Allen Brain Observatory database with a survey of spiking activity from approximately 100,000 units recorded by Neuropixels probes. In this first report on our survey, we used CCG time-lag analysis to uncover a marked correspondence between the anatomical and functional network organization of mouse cortical visual areas during sensory drive (Fig. 2f). Four popular measures of hierarchical processing—response latency, receptive field size, degree of phase modulation by a drifting grating, and response decay timescale—all changed systematically across the eight cortical and thalamic visual regions we examined (Fig. 3), as did change detection signals (Fig. 4f), especially during active behaviour (compared to passive viewing) and on trials in which the mouse correctly perceived a stimulus change (Extended Data Fig. 9k). This suggests that unexpected stimuli are amplified by successive levels of the hierarchy[39-41,43], a result consistent with general theories of hierarchical predictive processing[44]. Moreover, the behavioural importance of the hierarchy is further supported by our finding that higher levels have stronger trial-wise covariation with mouse behaviour than do lower ones (Fig. 4h).

Correlating functional metrics with a single anatomical variable—the hierarchy score—only serves as a crude, first-order characterization. Although, to our knowledge, we recorded spiking activity simultaneously from more mouse visual areas than any previous study, we sampled only 6 of the 16 extant cortical visual areas[45]. The primate visual system is organized into distinct processing streams[46,47]; there is also anatomical and functional evidence for parallel streams in mice[19,48-50]. The cortex also displays additional levels of organization, including functional sub-modules and parallel processing streams[19,21,33,47]. These diverse aspects must be incorporated to establish a more complete mapping between cortical structure and function.

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Mice

Mice were maintained in the Allen Institute for Brain Science animal facility and used in accordance with protocols approved by the Allen Institute's Institutional Animal Care and Use Committee. The bulk of experiments used C57BL/6J wild-type mice ($n = 30$), supplemented by recordings in three transgenic lines ($n = 8$ Pvalb-IRES-Cre × Ai32, $n = 12$ Sst-IRES-Cre × Ai32, and $n = 8$ Vip-IRES-Cre × Ai32), to facilitate the identification of genetically defined inhibitory cell types via opto-tagging[53].

Wild-type C57BL/6J mice were purchased from Jackson Laboratories at postnatal day (P)25–50. For experiments involving opto-tagging of inhibitory cells, Pvalb-IRES-Cre, Vip-IRES-Cre and Sst-IRES-Cre mice were bred in-house and crossed with an Ai32 channelrhodopsin reporter line[54]. Pvalb-IRES-Cre;Ai32 breeding sets (pairs and trios) consisted of heterozygous Pvalb-IRES-Cre mice crossed with either heterozygous or homozygous Ai32(RCL-ChR2(H134R)_EYFP) mice. Pvalb-IRES-Cre is expressed in the male germline. To avoid germline deletion of the stop codon in the *loxP*-STOP-*loxP* cassette, Pvalb-IRES-Cre;Ai32 mice were not used as breeders. Sst-IRES-Cre;Ai32 breeding sets (pairs and trios) consisted of heterozygous Sst-IRES-Cre mice crossed with either heterozygous or homozygous Ai32(RCL-ChR2(H134R)_EYFP) mice. Vip-IRES-Cre;Ai32 breeding sets (pairs and trios) consisted of heterozygous Vip-IRES-Cre mice crossed with either heterozygous or homozygous Ai32(RCL-ChR2(H134R)_EYFP) mice. Cre$^+$ cells from Ai32 lines are highly photosensitive, owing to the expression[55] of Channelrhodopsin-2.

After surgery, all mice were single-housed and maintained on a reverse 12-h light cycle in a shared facility with room temperatures between 20 and 22 °C and humidity between 30 and 70%. All experiments were performed during the dark cycle. For passive viewing experiments, mice were given ad libitum access to food and water. For behavioural experiments, mice were given an amount of water required to maintain 85% of their initial body weight, with ad libitum access to food.

Surgery

Headframe design.—To enable co-registration across surgical, intrinsic signal imaging, and electrophysiology rigs, each mouse was implanted with a grade 5 titanium headframe that provides access to the brain via a cranial window and permits head fixation in a reproducible configuration[2]. The cranial window angle was at 23° of roll and 6° of pitch, referenced to a plane passing through lambda and bregma and the mediolateral axis. Use of this headframe allowed the 5 mm craniotomy to be repeatability centred at $x = -2.8$ mm and $y = 1.3$ mm (origin at lambda).

The headframe was glued to a black acrylic photopolymer well that served four functions: (1) shielding the craniotomy and probes during the experiment, (2) providing a surface for precisely aligning the insertion window, (3) routeing the animal ground to an exposed gold pin, and (4) holding threads for a plastic cap that protects the craniotomy before and after the experiment.

Surgical procedures.—A pre-operative injection of dexamethasone (3.2 mg kg^{-1}, subcutaneously (s.c.)) was administered 1 h before surgery to reduce swelling and postoperative pain by decreasing inflammation. Mice were initially anesthetized with 5% isoflurane (1–3 min) and placed in a stereotaxic frame (Model 1900, Kopf). Isoflurane levels were maintained at 1.5–2.5% for the duration of the surgery. Body temperature was maintained at 37.5 °C. Carprofen was administered for pain management (5–10 mg kg^{-1}, s.c.) and atropine was administered to suppress bronchial secretions and regulate hearth rhythm (0.02–0.05 mg mg kg^{-1}, s.c.). An incision was made to remove skin, and the exposed skull was levelled with respect to pitch (bregma–lambda level), roll and yaw.

The headframe was placed on the skull and fixed in place with White C&B Metabond (Parkell). Once the Metabond was dry, the mouse was placed in a custom clamp to position the skull at a rotated angle of 20°, to facilitate creation of the craniotomy over the visual cortex. A circular piece of skull 5 mm in diameter was removed, and a durotomy was performed. The brain was covered by a 5-mm-diameter circular glass coverslip, with a 1-mm lip extending over the intact skull. The bottom of the coverslip was coated with a layer of polydimethylsiloxane (SYLGARD 184, Sigma-Aldrich) to reduce to reduce adhesion to the brain surface. The coverslip was secured to the skull with Vetbond (Patterson Veterinary)[56]. Kwik-Cast (World Precision Instruments) was added around the coverslip to further seal the implant, and Metabond bridges between the coverslip and the headframe well were created to hold the Kwik-Cast in place. At the end of the procedure, but before recovery from anesthesia, the mouse was transferred to a photodocumentation station to capture a spatially registered image of the cranial window (Extended Data Fig. 1a).

Surgery quality control.—In cases of excessive bleeding or other complications, the surgical procedure was aborted and the mouse was euthanized. Mice that completed surgery entered a 7–10 day recovery period that included regular checks for overall health, cranial window clarity and brain health. If mice failed the first health check, they received another one the following week. Mice that exhibited signs of deteriorating health or damaged brain surface vasculature were not passed on to the next step. Out of 105 mice entering the surgery step, 4 were removed from the pipeline due to quality control failures at this stage (Extended Data Fig. 2a).

Intrinsic signal imaging

Intrinsic signal imaging (ISI) measures the haemodynamic response of the cortex to visual stimulation across the entire field of view. This technique can be used to obtain retinotopic maps representing the spatial relationship of the visual field (or, in this case, coordinate position on the stimulus monitor) to locations within each cortical area. This mapping procedure was used to delineate functionally defined visual area boundaries to enable targeting of Neuropixels probes to retinotopically defined locations in primary and secondary visual areas[57].

Data acquisition.—Mice were lightly anesthetized with 1–1.4% isoflurane administered with a SomnoSuite (model 715; Kent Scientific) and vital signs were monitored with a PhysioSuite (model PS-MSTAT-RT). Eye drops (Lacri99 Lube Lubricant Eye Ointment; Refresh) were applied to maintain hydration and clarity of eyes during anesthesia. Imaging sessions began with a vasculature image acquired under green illumination (527-nm LEDs; Cree., C503B-GCN-CY0C0791). Next, the imaging plane was defocused between 500 µm and 1,500 µm along the optical axis, to match our established retinotopic mapping procedure[2]. The haemodynamic response to a visual stimulus was imaged under red light (635-nm LEDs; Avago Technologies, HLMP-EG08-Y2000) with an Andor Zyla 5.5 10 tap sCMOS camera. The stimulus consisted of an alternating checkerboard pattern (20° wide bar, 25° square size) moving across a mean luminance grey background. On each trial, the stimulus bar was swept across the four cardinal axes 10 times in each direction at a rate[58] of 0.1 Hz. Up to 10 trials were performed on each mouse.

Data processing.—A minimum of three trials were averaged to produce altitude and azimuth phase maps, calculated from the discrete Fourier transform of each pixel. A 'sign map' was produced from the phase maps by taking the sine of the angle between the altitude and azimuth map gradients. In the sign maps, each cortical visual area appears as a contiguous red or blue region[59]. These maps are used to confirm the cortical area identity of each probe insertion, using the vasculature as fiducial markers (Extended Data Fig. 1b, h, i).

The altitude and azimuth maps were also used to create a map of eccentricity from the centre of visual space (the intersection of 0° altitude and 0° azimuth). Because the actual centre of gaze will vary from mouse to mouse, the eccentricity map was shifted to align with the screen coordinates at the centre of V1 (which maps to the centre of the retina). This V1-aligned eccentricity map was used for probe targeting, to ensure that recorded neurons represent a consistent region on the retina, approximately at the centre of the right visual hemifield.

ISI quality control.—The quality control process for the ISI-derived maps included four distinct inspection steps:

1. The brain surface and vasculature images were inspected post-acquisition for clarity, focus, and position of the cranial window within the field of view.

2. Individual trials were inspected for visual coverage range and continuity of phase maps, localization of the signal from the amplitude maps and stereotypical organization of sign maps. Only trials respecting these criteria were included in the final average, and a minimum of three trials were required.

3. Visual area boundaries were delineated using automated segmentation, and maps were curated on the basis of stringent criteria to ensure data quality. The automated segmentation and identification of a minimum of six visual areas including V1, LM, RL, AL, AM and PM was required. A maximum of three manual adjustments were permitted to compensate for algorithm inefficiency.

4. Each processed retinotopic map was inspected for coverage range (35–60° altitude and 60–100° azimuth), bias (absolute value of the difference between max and min of altitude or azimuth range; $<10°$), alignment of the centre of retinotopic eccentricity with the centroid of V1 ($<15°$ apart), and the area size of V1 (>2.8 cm^2).

If quality control was not passed after the first round of ISI mapping, the procedure was repeated up to two more times to obtain a passing map. In addition to the quality control procedures carried out on the ISI-derived maps, the vasculature images were also examined for the presence of white artefacts on the brain surface. White artefacts, an indicator of potential brain damage, were grounds for failing the mouse out of the pipeline. Out of 101 mice entering ISI, 9 did not pass onto habituation owing to quality control failures during this step (Extended Data Fig. 2b).

Habituation and behaviour training

Habituation for passive viewing experiments.—Mice underwent two weeks of habituation in sound-attenuated training boxes containing a headframe holder, running wheel and stimulus monitor (Extended Data Fig. 1c). Each mouse was trained by the same operator throughout the two-week period. During the first week, the operator gently handles the mice, introduces them to the running wheel, and head-fixes them with progressively longer durations each day. During the second week, mice run freely on the wheel and are exposed to visual stimuli for 10 to 50 min per day. The following week, mice undergo habituation sessions of 75 min and 100 min on the recording rig, in which they view a truncated version of the same stimulus that will be shown during the experiment.

Behaviour training.—A subset of mice were trained to perform a change detection task in which one of 8 natural images was continuously flashed (250-ms image presentation followed by 500-ms grey screen) and mice were rewarded for licking when the image identity changed (Fig. 4a). The change detection task has been described in detail previously[37]. In brief, for each trial the time of image change was drawn from an exponential distribution with a minimum of 5 image flashes (3.75 s) and a maximum of 11 flashes (8.25 s). Licking before the image change restarted the trial. Trials in which the mouse licked within 750 ms of image change were 'hits', whereas licks within 750 ms of non-change catch trials (occurring at the same distribution of times since the last change as change trials) were classified as false alarms (Fig. 4b). Mice must perform the task with a d' of greater than 1 and have at least 100 contingent (non-aborted) trials for 3 consecutive days before moving to the recording rig.

Habituation quality control.—Upon completion of the second week of habituation, mice received an assessment of overall stress levels that reflected observations made by the trainer, including coat appearance, components of the mouse grimace scale and overall body movements. Out of 92 mice entering habituation for passive viewing experiments, 2 did not pass on to the insertion window implant step (Extended Data Fig. 2c).

Insertion window implant

Window generation.—After the completion of a successful ISI map, a custom insertion window was generated for each mouse. First, six insertion targets were manually drawn on the V1-aligned eccentricity map using a web-based annotation tool. Targets were positioned at the centre of retinotopy of V1, LM, AL, AM and PM; because the retinotopic centre of RL often lies on the boundary between RL and S1 barrel cortex, the target location was adjusted to be closer to the geometric centre of this area. The coordinates of each target were used to automatically generate the outlines of the insertion window, which was subsequently laser-cut out of 0.5 mm clear PETG plastic (Ponoko). When seated in the headframe well, the window facilitates access to the brain via holes over each of the six visual areas. A solidified agarose/ACSF mixture injected between the brain and the window stabilizes the brain during the recording.

Surgical procedure.—On the day of recording, the cranial coverslip was removed and replaced with an insertion window containing holes aligned to six cortical visual areas. First,

the mouse was anesthetized with isoflurane (3–5% induction and 1.5% maintenance, 100% O_2) and eyes were protected with ocular lubricant (I Drop, VetPLUS). Body temperature was maintained at 37.5 °C (TC-1000 temperature controller, CWE, Incorporated). Metabond bridges were removed from the glass cranial window, followed by the sealing layer of Kwik-Cast. Using a 2-mm silicone suction cup, the cranial window was gently lifted to expose the brain. The insertion window was then placed in the headframe well and sealed with Metabond. An agarose mixture was injected underneath the window and allowed to solidify. The mixture consisted of 0.4 g high EEO Agarose (Sigma-Aldrich), 0.42 g Certified Low-Melt Agarose (BioRad), and 20.5 ml ACSF (135.0 mM NaCl, 5.4 mM KCl, 1.0 mM $MgCl_2$, 1.8 mM $CaCl_2$, 5.0 mM HEPES). This mixture was optimized to be firm enough to stabilize the brain with minimal probe drift, but pliable enough to allow the probes to pass through without bending. A layer of silicone oil (30,000 cSt, Aldrich) was added over the holes in the insertion window to prevent the agarose from drying (Extended Data Fig. 1d). A 3D-printed plastic cap was screwed into the headframe well to keep out cage debris. At the end of this procedure, mice were returned to their home cages for 1–2 h.

Insertion window implant quality control.—Three out of 90 mice did not pass through to the recording step owing to procedure failures during implantation of the insertion window. These failures were caused by the headframe coming loose from the skull or excessive bleeding after removal of the cranial window, after which the mice were euthanized (Extended Data Fig. 2d).

Neuropixels recordings

Probes.—All neural recordings were carried out with Neuropixels probes[6]. Each probe contains 960 recording sites, a subset of 374 ('Neuropixels 3a') or 383 ('Neuropixels 1.0') of which can be configured for recording at any given time. The electrodes closest to the tip were always used, providing a maximum of 3.84 mm of tissue coverage. The sites are oriented in a checkerboard pattern on a 70 μm wide × 10 mm long shank. Neural signals are routed to an integrated base containing amplification, digitization and multiplexing circuitry. The signals from each recording site are split in hardware into a spike band (30-kHz sampling rate, 500-Hz high-pass filter) and an LFP band (2.5-kHz sampling rate, 1,000-Hz low-pass filter). Owing to their dense site configuration (20-μm vertical separation along the entire length of the shank), each probe has the capacity to record hundreds of neurons at the same time. Our goal was to insert six probes per mouse. Overall, we achieved a penetration success of 5.7 probes per mouse, with failures due to dura regrowth, collisions with the protective cone or opto-tagging fibre optic cable, or probe breakage during manipulation.

The base of each probe contains 32 10-bit analogue-to-digital converters (ADCs), each of which are connected to 12 spike-band channels and 12 LFP-band channels via multiplexers. A full cycle of digitization requires 156 samples: 12 samples from each of 12 spike-band channels, and 1 sample from each of 12 LFP-band channels. Each ADC serves a contiguous bank of odd or even channels, so ADC 1 digitizes channels [1,3,5,…,23], ADC 2 digitizes channels [2,4,6,…,24], ADC 3 digitizes channels [25,27,29,…,47], etc. Because of the need for interleaved sampling, common-mode noise will be shared across all channels that are acquired simultaneously, for example, [1,2,25,26,49,50,…,361,362].

Experimental rig.—The experimental rig (Extended Data Fig. 1g) was designed to allow six Neuropixels probes to penetrate the brain approximately perpendicular to the surface of the visual cortex. Each probe is mounted on a 3-axis micromanipulator (New Scale Technologies), which are in turn mounted on a solid aluminium plate, known as the probe cartridge. The cartridge can be removed from the rig using a pair of pneumatic tool-changers, to facilitate probe replacement and maintenance.

Workflow sequencing engine.—The experimental procedure was guided by a work sequencing engine (WSE), a custom graphical user interface (GUI) written in Python. This software ensured that all experimental steps were carried out in the correct order, reducing trial-to-trial variability and optimizing operator efficiency. The GUI logged the operator ID, mouse ID and session ID, and ensured that all hardware and software were properly configured. The WSE was also used to start and stop the visual stimulus, the body- and eye-tracking cameras, and Neuropixels data acquisition.

Probe alignment.—The tip of each probe was aligned to its associated opening in the insertion window using a coordinate transformation obtained via a previous calibration procedure. The XY locations of the six visual area targets were supplied by the WSE, and these values were translated into XYZ coordinates for each 3-axis manipulator using a custom Python script. The operator then moved each probe into place with a joystick, with the probes fully retracted along the insertion axis.

Application of CM-DiI.—CM-DiI (1 mM in ethanol; Thermo Fisher, V22888) was used to localize probes during the ex vivo imaging step because its fluorescence is maintained after brain clearing, and it has a limited diffusion radius. The probes were coated with CM-DiI before each recording by immersing them one by one into a well filled with dye, for approximately 1 min each.

Head fixation.—The mouse was placed on the running wheel and fixed to the headframe clamp with three set screws. Next, the plastic cap was removed from the headframe well and an aluminium cone with 3D-printed wings was lowered to prevent the mouse's tail from contacting the probes. An infrared dichroic mirror was placed in front of the right eye to allow the eye-tracking camera to operate without interference from the visual stimulus. A black curtain was then lowered over the front of the rig, placing the mouse in complete darkness except for the visual stimulus monitor.

Grounding.—A 32 AWG silver wire (A-M Systems) was epoxied to the headframe before the initial headframe/cranial window surgery. This wire becomes electrically conductive with the brain surface after the application of the ACSF/agarose mixture beneath the insertion window. The wire was pre-soldered to a gold pin embedded in the headframe well, which mates with a second gold pin on the protective cone. The cone pin was soldered to 22 AWG hook-up wire (SparkFun Electronics), which was connected to both the behaviour stage and the probe ground. Before the experiment, the brain-to-probe ground path was checked using a multimeter.

The reference connection on the Neuropixels probes was permanently soldered to ground using a silver wire, and all recordings were made using an external reference configuration. The headstage grounds (which are contiguous with the Neuropixels probe grounds) were connected together with 36 AWG copper wire (Phoenix Wire). For Neuropixels 3a, two probes had a direct path to animal ground, and the others were wired up serially. All probes were also connected to the main ground via the data cable (a dual coaxial cable). For Neuropixels 1.0, all probes were connected in parallel to animal ground, and were not connected to the main ground through the data cable (a single twisted pair cable).

Probe insertion.—The probe cartridge was initially held approximately 30 cm above the mouse. After the mouse was secured in the headframe, the cartridge was lowered so the probe tips were approximately 2.5-mm above the brain surface. The probes were then manually lowered one by one to the brain surface until spikes were visible on the electrodes closest to the tip. After the probes penetrated the brain to a depth of around 100 μm, they were inserted automatically at a rate of 200 μm min^{-1} (total of 3.5 mm or less in the brain) to avoid damage caused by rapid insertion[60]. After the probes reached their targets, they were allowed to settle for 5–10 min. Photo-documentation was taken with the probes fully retracted, after the probes reached the brain surface (Extended Data Fig. 1e), and again after the probes were fully inserted.

Data acquisition and synchronization.—Neuropixels data was acquired at 30 kHz (spike band) and 2.5 kHz (LFP band) using the Open Ephys GUI[61]. Gain settings of 500× and 250× were used for the spike band and LFP band, respectively. Each probe was either connected to a dedicated FPGA streaming data over Ethernet (Neuropixels 3a) or a PXIe card inside a National Instruments chassis (Neuropixels 1.0). Raw neural data was streamed to a compressed format for archiving, which was extracted before analysis.

Videos of the eye and body were acquired at 30 Hz. The angular velocity of the running wheel was recorded at the time of each stimulus frame, at approximately 60 Hz. Synchronization signals for each frame were acquired by a dedicated computer with a National Instruments card acquiring digital inputs at 100 kHz, which was considered the master clock. A 32-bit digital 'barcode' was sent with an Arduino Uno (SparkFun DEV-11021) every 30 s to synchronize all devices with the neural data. Each Neuropixels probe has an independent sample rate between 29,999.90 Hz and 30,000.31 Hz, making it necessary to align the samples offline to achieve precise synchronization. The synchronization procedure used the first matching barcode between each probe and the master clock to determine the clock offset, and the last matching barcode to determine the clock scaling factor. If probe data acquisition was interrupted at any point during the experiment, each contiguous chunk of data was aligned separately. Because one LFP band sample was always acquired after every 12th spike band sample, these data streams could be synchronized automatically once the spike band clock rate has been determined.

To synchronize the visual stimulus to the master clock, a silicon photodiode (PDA36A, Thorlabs) was placed on the stimulus monitor above a 'sync square' that flips from black to white every 60 frames. The analogue photodiode signal was thresholded and recorded

as a digital event by the sync computer. Individual frame times were reconstructed by interpolating between the photodiode on/off events.

Stimulus monitor.—Visual stimuli were generated using custom scripts based on PsychoPy[62] and were displayed using an ASUS PA248Q LCD monitor, with $1,920 \times 1,200$ pixels (55.7 cm wide, 60 Hz refresh rate). Stimuli were presented monocularly, and the monitor was positioned 15 cm from the right eye of the mouse and spanned $120° \times 95°$ of visual space before stimulus warping. Each monitor was gamma corrected and had a mean luminance of 50 cd m^{-2}. To account for the close viewing angle of the mouse, a spherical warping was applied to all stimuli to ensure that the apparent size, speed and spatial frequency were constant across the monitor as seen from the mouse's perspective.

Stimuli for passive viewing experiments.—All experiments began with a receptive field mapping stimulus consisting of 2 Hz, 0.04 cycles per degree drifting gratings with a 20° circular mask. These Gabor patches randomly appeared at one of 81 locations on the screen (9×9 grid) for 250 ms at a time, with no blank interval. The receptive field mapping stimulus was followed by a series of dark or light full-field flashes, lasting 250 ms each and separated by a 2-s inter-trial interval.

Next, mice were shown one of two possible stimulus sets. The first, called 'Brain Observatory 1.1' is a concatenation of two sessions from the Two-Photon Imaging Brain Observatory[2] (Extended Data Fig. 6b). Drifting gratings were shown with a spatial frequency of 0.04 cycles per degree, 80% contrast, 8 directions (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°, clockwise from 0° = right-to-left) and 5 temporal frequencies (1, 2, 4, 8 and 15 Hz), with 15 repeats per condition. Static gratings were shown at 6 different orientations (0°, 30°, 60°, 90°, 120°, 150°, clockwise from 0° = vertical), 5 spatial frequencies (0.02, 0.04, 0.08, 0.16, 0.32 cycles per degree) and 4 phases (0, 0.25, 0.5, 0.75); they are presented for 0.25 s, with no intervening grey period. The Natural Images stimulus consisted of 118 natural images taken from the Berkeley Segmentation Dataset[63], the van Hateren Natural Image Dataset[51] and the McGill Calibrated Colour Image Database[52]. The images were presented in greyscale and were contrast-normalized and resized to $1,174 \times 918$ pixels. The images were presented in a random order for 0.25 s each, with no intervening grey period. Two natural movie clips were taken from the opening scene of the movie Touch of Evil[64]. Natural Movie One was a 30-s clip repeated 20 times (2 blocks of 10), while Natural Movie Three was a 120-s clip repeated 10 times (2 blocks of 5). All clips were contrast-normalized and were presented in greyscale at 30 fps.

The second stimulus set, called 'Functional Connectivity', consisted of a subset of the stimuli from the Brain Observatory 1.1 set shown with a higher number of repeats (Extended Data Fig. 6c). Drifting gratings were presented at 4 directions and one temporal frequency (2 Hz) with 75 repeats. A contrast-tuning stimulus consisting of drifting gratings at 4 directions (0°, 45°, 90°, 135°, clockwise from 0° = left-to-right) and 9 contrasts (0.01, 0.02, 0.04, 0.08, 0.13, 0.2, 0.35, 0.6, 1.0) was also shown. The Natural Movie One stimulus was presented a total of 60 times, with an additional 20 repeats of a temporally shuffled version. Last, a dot motion stimulus consisting of approximately 200 1.5°-radius white dots on a mean-luminance grey background moving at one of 7 speeds (0° s^{-1}, 16° s^{-1}, 32° s^{-1},

$64° \text{ s}^{-1}$, $128° \text{ s}^{-1}$, $256° \text{ s}^{-1}$, $512° \text{ s}^{-1}$) in four different directions (−45°, 0°, 45°, 90°; + = clockwise; 0° = left-to-right) at 90% coherence was shown.

Stimuli for behavioural experiments.—Mice carried out one hour of a change detection task[37]. After the behaviour session, the lick spout was retracted and receptive field mapping stimuli and full-field flashes were presented for 25 min, with the same parameters as those used in the passive viewing experiments. Finally, the exact sequence and timing of images viewed during the behavioural task were re-played (one hour). All other aspects of the rig—including the running wheel, stimulus monitor, and electrophysiological recordings —were the same as for the passive viewing experiments.

Probe removal and cleaning.—When the stimulus set was over, probes were retracted from the brain at a rate of 1 mm s^{-1}, after which the probe cartridge was raised to its full height. The protective cap was screwed into the headframe well, then mice were removed from head fixation and returned to their home cages overnight. Probes were immersed in a well of 1% Tergazyme for around 12 h, which was sufficient to remove tissue and silicone oil before the next recording session.

Quality control for the Neuropixels recording session.—Neuropixels recording sessions were subjected to the following quality control criteria (Extended Data Fig. 2e):

Eye foam.: If white build-up around the eye obscured the pupil, the experiment was cancelled and the session was failed (8 mice).

Bleeding.: If bleeding resulting from the window implant or the probe insertion obscured the vasculature, the session was failed (4 mice).

Probe insertion.: If fewer than four probes successfully entered the brain, the session was failed (1 mouse).

Dropped frames.: If the stimulus monitor photodiode measured more than 60 delayed frames, the session was failed (1 mouse).

Missing files.: If any critical files were overwritten, the session was failed (2 mice).

Noise levels.: If high root mean square noise levels in the spike band persisted after median subtraction, the session was failed (4 mice).

Probe drift.: If one or more probes exhibited more than 80 μm of drift over the course of the experiment, the session was failed (6 mice). Typical drift levels were around 40 μm, and drift levels were highly correlated across probes.

In total, out of 87 mice entering the recording step, 61 passed session-level quality control.

Ex vivo imaging

Tissue clearing.—Mice were perfused with 4% paraformaldehyde (PFA) (after induction with 5% isoflurane and $1\,l\,min^{-1}$ of O_2). The brains were preserved in 4% PFA, rinsed with $1\times$ PBS the next morning, and stored at 4 °C in PBS. Next, brains were run through a tissue clearing process based on the iDISCO method[65]. This procedure uses different solvents that dehydrate and delipidate the tissue. The first day, the brains were immersed in different concentrations of methanol (20, 40, 60%) for an hour each, then overnight in 80% methanol. On the second day, they were dipped into 100% methanol (twice for one hour) and then into a mixture of 1/3 methanol and 2/3 dichloromethane overnight. On the third day, the brains were moved from pure dichloromethane (2×20 min) to pure dibenzyl ether, where they remained for 2–3 days until clearing was complete (Extended Data Fig. 5a).

Optical projection tomography.—Whole-brain 3D imaging was accomplished with optical projection tomography (OPT)[66-68]. The OPT instrument consisted of collimated light sources for transmitted illumination (on-axis white LED, Thorlabs MNWHL4 with Thorlabs SM2F32-A lens and Thorlabs DG20-600 diffuser) or fluorescence excitation (off-axis Thorlabs M530L3, with Thorlabs ACL2520U-DG6-A lens and Chroma ET535/70m-2P diffuser), a 0.5× telecentric lens (Edmund Optics 62-932) with emission filter (575 nm LP, Edmund Optics 64-635), and a camera (IDS UI-3280CP). The specimen was mounted on a rotating magnetic chuck attached to a stepper motor, which positioned the specimen on the optical axis and within a glass cuvette filled with dibenzyl ether. The stepper motor and illumination triggering were controlled with an Arduino Uno (SparkFun DEV-11021) and custom shield including a Big Easy Driver (SparkFun ROB-12859). Instrument communication and image capture was accomplished with MicroManager[69].

A series of 400 images were captured with transmitted LED illumination with each image captured with the specimen rotated 0.9° relative to the previous position. This series of 400 images was repeated with the fluorescence excitation LED. Each channel was stored as a separate OME-TIFF dataset before extracting individual planes and metadata required for reconstruction using a custom Python script (Extended Data Fig. 5b).

Isotropic 3D volumes were reconstructed from these projection images using NRecon (Bruker). The rotation axis offset and region-of-interest bounds were set for each image series pair using the transmitted channel dataset, then the same values applied to the fluorescence channel dataset. A smoothing level of 3 using a Gaussian kernel was applied to all images. Reconstructions were exported as single-plane 16-bit TIFF images taken along the rotation axis with final voxel size of 7.9 μm per side (Extended Data Fig. 5c).

Registering probes to the common coordinate framework.—Reconstructed brains were downsampled to 10 μm per voxel and roughly aligned to the Allen Institute Common Coordinate Framework (CCFv3) template brain using an affine transform. The volume was then cropped to a size of $1{,}023 \times 1{,}024 \times 1{,}024$ and converted to Drishti format (https://github.com/nci/drishti). Next, 6–54 registration points were marked in up to 14 coronal slices of the individual brain by comparing to the CCFv3 template brain[70] (Extended Data Fig. 5c). Fluorescent probe tracks were manually labelled in coronal slices of the individual

brain, and the best-fit line was found using singular value decomposition (Extended Data Fig. 5e). The registration points were used to define a 3D nonlinear transform (VTK thinPlateSplineTransform), which was used to translate each point along the probe track into the CCFv3 coordinate space. Each CCFv3 coordinate corresponds to a unique brain region, identified by its structure acronym (for example, CA3, LP, VISp, etc.). A list of CCFv3 structure acronyms along each track was compared to the physiological features measured by each probe (for example, unit density, LFP theta power; Extended Data Fig. 5f). The locations of major structural boundaries were manually identified to align the CCFv3 labels with the physiology data; the most important features were the decrease in unit density at the cortical surface and L6-hippocampus boundary, and the decrease in theta power at the hippocampus-thalamus boundary. After the manual alignment procedure, each recording channel (and its associated units) was assigned to a unique CCFv3 structure (Extended Data Fig. 5g). White matter structures were not included; any units mapped to a white matter structure inherited the grey matter structure label that was immediately ventral along the probe axis.

Identification of cortical visual area targets.—To confirm the identity of the cortical visual areas, images of the probes taken during the experiment were compared to images of the brain surface vasculature taken during the ISI session. Vasculature patterns were used to overlay the visual area map on an image of the brain surface with the probes inserted. When done in custom software, key points were selected along the vasculature on both images and a perspective transform (OpenCV) was performed to warp the insertion image to the retinotopic map. When done manually, the overlap of both images was done in Photoshop or Illustrator (Adobe Suite). In both cases, the probe entry points were manually annotated. Finally, an area was assigned to each probe. Overall, successful targeting of the 6 target visual areas occurred at the following rates: 89% for AM, 72% for PM, 98% for V1, 85% for LM, 79% for AL and 90% for RL. A small subset of penetrations were mapped to LI, MMA or MMP[45]. Penetration points that could not be unambiguously associated with a particular visual area were classified as 'VIS'. If the cortical area label obtained via CCFv3 registration did not match the area identified in the insertion image overlay, the insertion image overlay took precedence.

Cortical depth and layer labels.—For cortical units registered to the CCFv3, we used 'cortical streamlines' to extract their relative depths (Extended Data Fig. 10a; 0 = surface, 1 = white matter). Each point in the cortex is mapped to a unique depth along a path orthogonal to the equi-potential fields between the brain surface and white matter (based on the solution to Laplace's equation in three dimensions). This method yields normalized depth estimates even for regions of extreme cortical curvature, such as the prefrontal cortex. Streamlines are preferable to using distance along the probe axis, as they account for differences in insertion angle across areas.

In addition, CCFv3 coordinates were used as indices into the template volume in order to extract layer labels for each cortical unit (L1, L2/3, L4, L5, or L6). The relative thickness of each layer, which can vary both within and across areas, is based on the average of the 1,675 individual brains used to create the template volume.

Ex vivo imaging quality control.—Quality control was performed on a probe-by-probe, rather than a mouse-by-mouse, basis. Some probes were not visible in the OPT images due to faint CM-DiI signal or reconstruction artefacts caused by air bubbles in the tissue (Extended Data Fig. 2f). In total, 284 out of 332 probes were mapped to the CCFv3. Probes that failed the ex vivo imaging step were not excluded from further analysis, but only included structure labels for channels in the cortex (with the bottom of the cortex identified on the basis of the drop in unit density between the cortex and the hippocampus).

Spike sorting

Data pre-processing.—Data was written to disk in a format containing the original 10-bit samples from each ADC. These files were backed up to a tape drive, then extracted to a new set of files that represent each sample as a 16-bit integer, scaled to account for the gain settings on each channel. Separate data files were generated for the LFP band and the spike band, along with additional files containing the times of synchronization events. The extracted files consume approximately 36% more disk space than the originals.

Before spike sorting, the spike-band data passed through four steps: offset removal, median subtraction, filtering and whitening. First, the median value of each channel was subtracted to centre the signals around zero. Next, the median across channels was subtracted to remove common-mode noise. Although Neuropixels have been measured to have a spike-band RMS noise levels of 5.1 µV in saline[6], this cannot be achieved in practice when recording in vivo. The signals become contaminated by background noise in neural tissue; movement artefacts associated with mouse locomotion, whisking and grooming; and electrical noise introduced by the additional wiring required to support several probes on one rig. To remove noise sources that are shared across channels, the median was calculated across channels that are sampled simultaneously, leaving out adjacent (even/odd) channels that are probably measuring the same spike waveforms, as well as reference channels that contain no signal. For each sample, the median value of channels n:24:384, where n = [1,2,3,…,24], was calculated, and this value was subtracted from the same set of channels. This method rejects high-frequency noise more effectively than subtracting the median of all channels, at the cost of leaving a residual of around 2 µV for large spikes, visible in the mean waveforms. Given that this value is well below the RMS noise level of the Neuropixels probes under ideal conditions, it should not affect spike sorting. The original data are overwritten with the median-subtracted version, with the median value of each block of 16 channels saved separately, to enable reconstruction of the original signal if necessary. The median-subtracted data file is sent to the Kilosort2 MATLAB package (https://github.com/mouseland/Kilosort2, commit 2fba667359dbd-dbb0e52e67fa848f197e44cf5ef; 8 April 2019), which applies a 150-Hz high-pass filter, followed by whitening in blocks of 32 channels. The filtered, whitened data are saved to a separate file for the spike-sorting step.

Kilosort2.—Kilosort2 was used to identify spike times and assign spikes to individual units[24]. Traditional spike sorting techniques extract snippets of the original signal and perform a clustering operation after projecting these snippets into a lower-dimensional feature space. By contrast, Kilosort2 attempts to model the complete dataset as a sum of spike 'templates'. The shape and locations of each template is iteratively refined until the

data can be accurately reconstructed from a set of N templates at M spike times, with each individual template scaled by an amplitude, a. A critical feature of Kilosort2 is that it allows templates to change their shape over time, to account for the motion of neurons relative to the probe over the course of the experiment. Stabilizing the brain using an agarose-filled plastic window has almost eliminated probe motion associated with mice running, but slow drift of the probe over approximately 3-h experiments is still observed. Kilosort2 is able to accurately track units as they move along the probe axis, eliminating the need for the manual merging step that was required with the original version of Kilosort[26]. The spike-sorting step runs in approximately real time (around 3 h per session) using a dual-processor Intel 4-core, 2.6-GHz workstation with an NVIDIA GTX 1070 GPU.

Removing putative double-counted spikes.—The Kilosort2 algorithm will occasionally fit a template to the residual left behind after another template has been subtracted from the original data, resulting in double-counted spikes. This can create the appearance of an artificially high number of ISI violations for one unit or artificially high zero-time-lag synchrony between nearby units. To eliminate the possibility that this artificial synchrony will contaminate data analysis, the outputs of Kilosort2 are post-processed to remove spikes with peak times within 5 samples (0.16 ms) and peak waveforms within 5 channels (around 50 μm). This process removes more than 10 within-unit overlapping spikes from $2.5 \pm 1.8\%$ of units per session. It removes $2.05 \pm 0.65\%$ of spikes in total, after accounting for between-unit overlapping spikes.

Removing units with artefactual waveforms.—Kilosort2 generates templates of a fixed length (2 ms) that matches the time course of an extracellularly detected spike waveform. However, there are no constraints on template shape, which means that the algorithm often fits templates to voltage fluctuations with characteristics that could not physically result from the current flow associated with an action potential. The units associated with these templates are considered 'noise', and are automatically filtered out on the basis of three criteria: spread (single channel, or more than 25 channels), shape (no peak and trough, based on wavelet decomposition), or multiple spatial peaks (waveforms are non-localized along the probe axis). The automated algorithm removed 94% of noise units, or 26% of total units. A final manual inspection step was used to remove an additional 2,140 noise units across all experiments (Extended Data Fig. 3).

Spike-sorting quality control.—All units not classified as noise are packaged into Neurodata Without Borders (NWB) files for potential further analysis. Because different analyses may require different quality thresholds for defining inclusion criteria, we calculate a variety of metrics that can be used to filter units. These metrics are based on both the physical characteristics of the units' waveforms[71], or their isolation with respect to other units from the same recording (Extended Data Fig. 4a).

Firing rate: n/T, where $n =$ number of spikes in the complete session and $T =$ total time of the recording session in seconds.

Presence ratio: The session was divided into 100 equal-sized blocks; the presence ratio is defined as the fraction of blocks that include one or more spikes from a particular unit.

Units with a low presence ratio are likely to have drifted out of the recording, or could not be tracked by Kilosort2 for the duration of the experiment.

Maximum drift: To compute the maximum drift for one unit, the peak channel was calculated from the top principal components of every spike. Next, the peak channel values are binned in 51-s intervals, and the median value is calculated across all spikes in each bin (assuming at least 10 spikes per bin). The maximum drift is defined as the difference between the maximum peak channel and the minimum peak channel across all bins. The average maximum drift across all units is used to identify sessions with a high amount of probe motion relative to the brain.

Waveform amplitude: The difference (in microvolts) between the peak and trough of the waveform on a single channel.

Waveform spread: Spatial extent (in μm) of channels in which the waveform amplitude exceeds 12% of the peak amplitude.

Waveform duration: Difference (in ms) of the time of the waveform peak and trough on the channel with maximum amplitude.

ISI violations: This metric searches for refractory period violations that indicate a unit contains spikes from multiple neurons. The ISI violations metric represents the relative firing rate of contaminating spikes. It is calculated by counting the number of violations of less than 1.5 ms, dividing by the amount of time for potential violations surrounding each spike, and normalizing by the overall spike rate. It is always positive (or 0), but has no upper bound. See ref. [72] for more details.

Signal-to-noise ratio: After selecting 1,000 individual spike waveforms on the channel with maximum amplitude, the mean waveform on that channel was subtracted. The signal-to-noise ratio (SNR) is defined as the ratio between the waveform amplitude and 2× the standard deviation of the residual waveforms[73]. Because this definition of SNR assumes that waveforms remain stable over time, changes in a unit's waveform as a result of probe motion will cause this metric to be inaccurate. In addition, because it is only calculated for the peak channel, this metric does not necessarily reflect the overall isolation quality of a unit when taking into account all available information.

Isolation distance: The square of the Mahalanobis distance required to find the same number of 'other' spikes as the total number of spikes for the unit in principal component space[74]. Similarly to SNR, isolation distance is not tolerant to electrode drift, and changes in waveform shape over time can reduce the isolation distance calculated over the entire session.

d': Linear discriminant analysis is used to find the line of maximum separation in principal component space. d' indicates the separability of the unit of interest from all other units. See ref. [72] for more information. This metric is not tolerant to electrode drift, and changes in waveform shape over time can reduce the value of d' calculated over the entire session.

Amplitude cutoff: This metric provides an approximation of a unit's false negative rate. First, a histogram of spike amplitudes is created, and the height of the histogram at the minimum amplitude is extracted. The percentage of spikes above the equivalent amplitude on the opposite side of the histogram peak is then calculated. If the minimum amplitude is equivalent to the histogram peak, the amplitude cutoff is set to 0.5 (indicating a high likelihood that more than 50% of spikes are missing). This metric assumes a symmetrical distribution of amplitudes and no drift, so it will not necessarily reflect the true false negative rate.

Nearest neighbours hit rate: For each spike belonging to the unit of interest, the four nearest spikes in principal-component space are identified. The 'hit rate' is defined as the fraction of these spikes that belong to the unit of interest. This metric is based on the 'isolation' metric from ref. [75]. Again, electrode drift that alters waveform shape can negatively affect this metric without necessarily changing the isolation quality of a unit at any given time point.

Filtering of units on the basis of quality metrics and other criteria is illustrated in Extended Data Fig. 4b.

Data analysis

Receptive field analysis.—The receptive field for one unit is defined as the 2D histogram of spike counts (quantified during the 250-ms stimulus presention) at each of 81 locations of the Gabor stimulus (9×9 pixels, $10°$ separation between pixel centres, Extended Data Fig. 6d).

A chi-square test for independence was used to assess the presence of a significant receptive field. A chi-square test statistic was computed $\chi^2 = \sum_{i=0}^{n} \frac{(E_i - O_i)^2}{E_i}$, where $O_i = \frac{1}{m_i} \sum_{j=0}^{m_i} R_{i,j}$ is the observed average response (R) of the unit over m presentations of the Gabor stimulus at location i, and $E_i = \frac{\sum_i^n \sum_j^{m_i} R_{i,j}}{\sum_i^n m_i}$ is the expected (grand average) response per stimulus presentation. A P value was then calculated for each unit by comparing the test statistic against a null distribution of 1,000 test statistics, each computed from the unit's responses after shuffling the locations across all presentations.

To compute the receptive field area and centre location, each receptive field was first smoothed using a Gaussian filter ($\sigma = 1.0$). The smoothed receptive field (RF) was thresholded at max(RF) – std(RF), a value that provided good agreement with the qualitative receptive field boundaries. The receptive field centre location was calculated on the basis of the centre of mass of the largest contiguous area above threshold, and its area was equivalent to its pixel-wise area multiplied by 100 degrees^2 (Extended Data Fig. 6e).

Cross-correlation analysis.—We measured functional interactions between pairs of units using CCGs[28,29,76]. CCGs were calculated for periods of full-field drifting grating stimuli (2-s stimulus presentation interleaved with 1-s grey period; orientations = [0, 45, 90, 135] degrees, temporal frequency = 2 cycles per second, spatial frequency = 0.04 cycles per

degree, contrast = 0.8) for units with mean firing rate greater than 2 Hz between 50 ms and 500 ms after stimulus onset.

The CCG is defined as:

$$\mathrm{CCG}(\tau) = \frac{\frac{1}{M}\sum_{i=1}^{M}\sum_{t=1}^{N} x_1^i(t) x_2^i(t+\tau)}{\theta(\tau)\sqrt{\lambda_1 \lambda_2}}$$

where M is the number of trials, N is the number of bins in the trial, x_1^i and x_2^i are the spike trains of the two units on trial i, τ is the time lag relative to reference spikes, and λ_1 and λ_2 are the mean firing rates of the two units. The CCG is essentially a sliding dot product between two spike trains. $\theta(\tau)$ is the triangular function which corrects for the overlap time bins caused by the sliding window. To correct for firing-rate dependence, we normalized the CCG by the geometric mean spike rate. An individually normalized CCG is computed separately for each drifting grating orientation (75 repeats per orientation) then averaged across 4 orientations to obtain the CCG for each pair of units.

Ajitter-correction method was used to remove stimulus-locked correlations and slow temporal correlations from the original CCG.

$$\mathrm{CCG}_{\text{jitter\_corrected}} = \mathrm{CCG}_{\text{original}} - \mathrm{CCG}_{\text{jittered}}$$

The jitter-corrected CCG was created by subtracting the expected value of CCGs produced from a resampled version of the original dataset with spike times randomly perturbed (jittered) within the jitter window[28,29]. The correction term ($\mathrm{CCG}_{\text{jittered}}$) is the true expected value which reflects the average over all possible resamples of the original dataset. $\mathrm{CCG}_{\text{jittered}}$ is normalized by the geometric mean rate before subtracting from $\mathrm{CCG}_{\text{original}}$. The analytical formula used to create a probability distribution of resampled spikes is provided in ref. [77]. This method disrupted the temporal correlation within the jitter window, while maintaining the number of spikes in each jitter window and the shape of the peristimulus time histogram (PSTH) averaged across trials. For our measurement, a 25-ms jitter window was chosen on the basis of previous studies[28,30]. This jitter-correction method removes both the stimulus-locked component of the response, as well as slow fluctuations larger than the jitter window.

A sharp peak was deemed significant if the maximum of jitter-corrected CCG amplitude within a ±10 ms window had a magnitude larger than sevenfold of the standard deviation of the CCG flanks (between ±50–100 ms from zero). All subsequent analysis was based on significant CCG sharp peaks.

A Wilcoxon rank-sum test was used to compare the distribution of CCG peak offsets between neighbouring areas (defined by the anatomical hierarchical score) and the distribution of CCG peak offset within an area. The significance test was performed within each mouse, and the P values were combined across 25 mice using Fisher's method. V1–LM vs V1–V1, $P = 0$; LM–RL vs LM–LM, $P = 1.9 \times 10^{-5}$; RL–AL vs RL–RL, $P =$

2.4×10^{-5}; AL–PM vs AL–AL, $P = 0.081$; PM–AM vs PM–PM, $P = 3.2 \times 10^{-4}$. All between-area distributions are significantly different from the within-area distributions at the 5% confidence level, except for AL–PM.

Response latency.—Response latency was calculated as the time to first spike (TFS). TFS was estimated in each trial by looking for the time of first spike 30 ms after stimulus onset. If no spike was detected within 250 ms after stimulus onset, that trial was not included. The overall latency for each unit was defined as the median TFS across trials.

As a control, we calculated TFS using the same procedure, but during the 1-s pre-stimulus interval with a mean-luminance grey screen. Under these conditions, TFS for individual areas was not correlated with anatomical hierarchy score (Pearson's $r = 0.57$, $P = 0.14$), but was strongly negatively correlated with their baseline firing rates (Pearson's $r = -0.98$, $P = 0.00001$; Extended Data Fig. 9b, c).

Modulation index.—The stimulus modulation index (MI) reflects how spiking activity of each unit is modulated by the temporal frequency of the drifting grating stimulus[34,35]. It is defined as:

$$MI = \left| \frac{PS(f_{\text{pref}}) - \langle PS \rangle_{\text{f}}}{\sqrt{\langle PS^2 \rangle_{\text{f}} - \langle PS \rangle_{\text{f}}^2}} \right|$$

where PS indicates the power spectral density of the PSTH, and denotes the averaged power over all frequencies; f_{pref} is the preferred temporal frequency of the unit. This metric quantifies the difference between spiking response power at each unit's preferred stimulus frequency ($PS(f_{\text{pref}})$) versus its averaged response power across frequencies ($\langle PS \rangle_{\text{f}}$). The power spectrum was computed using Welch's method on the 10 ms-binned PSTH for each unit's preferred condition. MI values greater than 3 correspond to strong modulation of spiking at the stimulus frequency (indicative of simple-cell–like responses), whereas smaller MI values indicate less modulation by stimulus temporal frequency (indicative of complex-cell–like responses)[78].

Intrinsic timescale and response decay timescale.—We calculated intrinsic timescale using a method similar to that described previously[13]. We first extracted spike times for each unit during the 1 s pre-stimulus period before the onset of each full-field flash, and binned them in 10-ms intervals. We then calculated the Pearson correlation between spike counts at each of 100 possible offsets, to fill the upper triangle of a 100×100 correlation matrix. We averaged the correlations along the diagonals of this matrix, and fit an exponential decay function to the first 50 points (500 ms), with the decay timescale bounded between 1 and 1,000 ms. Units were only included in the overall average if the standard deviation of the estimated timescale parameter was less than 100, and at least 100 spikes were used for fitting. The distribution of intrinsic timescales for each visual area is shown in Extended Data Fig. 9f.

We calculated the response decay timescale for each unit on the basis of binned spike counts during the 250-ms presentation period of the full-field flash stimulus, with 10 ms temporal

resolution. Using this data, we calculated a 2D autocorrelation matrix (scipy.signal.correlate) and averaged this matrix across trials. An exponential decay function was fit to the result, with the decay timescale bounded between 1 and 1,000 ms. Units were only included in the overall average if the standard deviation of the estimated timescale parameter was less than 20, and at least 50 spikes were used for fitting. The distribution of response decay timescales for each visual area is shown in Fig. 3k.

Directionality score.—We quantified the relative proportion of positive and negative CCG time lags with a 'directionality score' (DS). DS is defined as:

$$DS = \frac{C_{\text{positive}} - C_{\text{negative}}}{C_{\text{positive}} + C_{\text{negative}}}$$

where C_{positive} represents the number of functional connections (that is, number of pairwise significant CCG sharp peaks) with positive time lag from source to target area, and C_{negative} represents the number of functional connections with negative time lag. The DS is bounded between −1 and 1. A positive value indicates that temporally leading connections predominate from source to target area, whereas a negative value indicates that lagging connections are more common from source to target area. We calculated a DS for the peak offset distribution between all pairs of areas, visualized as a matrix (Fig. 2e). Note that this metric alone does not make any specific assumption of feedforward or feedback connections, it only quantifies the relative number of positive and negative time lag connections between two areas. To complement this measurement, we also quantified the asymmetry of the between-area time lags using the median of these distributions (Extended Data Fig. 7c, d). However, the median of the CCG time lag distribution in principle cannot reflect the shape of distribution, which influences the relative hierarchy (see simulation in Extended Data Fig. 8e, f).

Analysis of neural responses during the change detection task.—For each unit, spike density functions (SDFs) were calculated by convolving spike times relative to each image change or the image flash preceding image change ('pre-change') with a causal exponential filter (decay time constant = 5 ms). The firing rate during a baseline window 250 ms immediately preceding image presentation was subtracted from each SDF. Mean SDFs were then calculated by averaging across all image change or pre-change presentations. Units were included in further analysis if their mean firing rate was greater than 0.1 spikes per second and the peak of the mean SDF after image change was greater than 5 times the standard deviation of the mean SDF during the baseline window.

Responses to image change and pre-change were calculated as the mean baseline-subtracted firing rate during the response window. We defined the change modulation index (Fig. 4f) for each unit as the difference between the mean response to each image on change and pre-change presentations divided by their sum, and took the average of this value across all eight images. This analysis was repeated for data collected during a 'passive' session during which the lick spout was retracted and the exact sequence and timing of images viewed during the behavioural task were replayed.

For comparison of decoder predictions and mouse behaviour (Fig. 4g, h), we trained random forest classifiers with fivefold cross validation to distinguish population activity associated with change or pre-change image presentations. The input to the decoder for each trial was a vector of length neurons × time samples, formed by concatenating the SDFs of each neuron. Using these features, the decoder is trained to predict whether each trial was a change or pre-change image presentation. For each brain region and task session we used activity from subsamples of 20 neurons, beyond which decoder accuracy improvements were minimal (Extended Data Fig. 9o). The number of subsamples varied depending on the number of neurons recorded such that there was a greater than 99% chance that each neuron was included in at least one subsample. The result for each experiment was the median parameter value (for example, decoder accuracy) across subsamples. The average output of the ensemble of random forest classifiers ($n = 100$) results in a probability predicting whether a given trial was an image change. We compared these values to the response of the mouse on each trial (hit or miss) using a Pearson correlation, and then averaged across experiments for each region to generate Fig. 4h.

Eye and pupil tracking.—A single, universal eye tracking model was trained in DeepLabCut[79], a ResNET-50 based network, to recognize up to 12 tracking points each around the perimeter of the eye, the pupil, and the corneal reflection. A published numerical routine[80] was used to fit ellipses to each set of tracking points. For each ellipse, the following parameters were calculated: centre coordinates, half-axes and rotation angle. Fits were performed on each frame if there were at least six tracked points and a confidence of $l > 0.8$ as reported by the output of DeepLabCut. For frames in which there were fewer than 6 tracked points above the confidence threshold, the ellipse parameters were set to not-a-number (NaN).

The training dataset contained two sources of hand-annotated data: (1) Three frames from each of 40 randomly selected movies. On each frame, eight points were annotated around the eye and pupil. The centre of the corneal reflection was annotated with a single point. (2) 4,150 frames with the pupil and corneal reflections annotated with ellipses.

Across 50 mice with processed eye-tracking videos, we used the gaze_mapping module of the AllenSDK to translate pupil position into screen coordinates (in units of degrees). On average, 95% of gaze locations fell within $6.4 \pm 2.1°$ of the mean, with a maximum of $13.6°$.

Anatomical hierarchy analysis.—A detailed description of the unsupervised construction of a data-driven anatomical hierarchy is available in ref. [3]. Here we provide a summary of how the anatomical hierarchy of the six visual cortical areas (V1, LM, AL, RL, PM and AM) and two thalamic nuclei (LGN and LP) was constructed on the basis of the anatomical connectivity. Specifically, the anatomical hierarchy was uncovered on the basis of cortical lamination patterns of the structural connections among the cortical and thalamic regions of interest, obtained from Cre-dependent viral tracing experiments.

To classify laminar patterns of cortico-cortical (CC) and thalamo-cortical (TC) connections and to assign a direction to each cluster of laminar patterns, we used a large-scale dataset on cell class-specific connectivity among all 37 cortical areas and 24 thalamic nuclei defined

using 15 Cre driver transgenic lines (849 cortical and 81 thalamic experiments; 7,063 unique source-target-Cre line combinations), available in ref. [3]. For each transgenic line, the strength and layer termination pattern of the connections were quantified on the basis of relative layer density, the fraction of the total projection signal in each layer scaled by the relative layer volumes in that target. For the connections above a threshold ($10^{-1.5}$), unsupervised clustering of the layer termination patterns was performed, yielding nine clusters of distinct cortical layer termination patterns of CC and TC connections. See figure 5a, b of ref. [3] for a schematic of the nine types of cortical target lamination pattern.

Following the classification of the nine clusters of the laminar patterns, an unsupervised method was used to simultaneously assign a direction to a cluster type and to construct a hierarchy by maximizing the self-consistency of the obtained hierarchy. The mapping function M_{CC} maps a type of CC connection cluster ($C_{T_{i,j}} \in \{1, \dots 9\}$, where $C_{T_{i,j}}$ denotes the layer termination pattern of the connection from area j to area i for Cre-line T) to either feedforward ($M_{CC} = 1$) or feedback ($M_{CC} = -1$) type, that is, $M_{CC}: \{1, \dots, 9\} \rightarrow \{-1, 1\}$. Similarly, the mapping function M_{TC} of the thalamocortical layer termination types to either direction is defined as $M_{TC}: \{1, \dots, 9\} \rightarrow \{-1, 1\}$. By constructing the hierarchy of all 37 cortical areas and 24 thalamic nuclei, the optimal mapping function that maximizes the self-consistency measured by the *global hierarchy score* was found[3] (refer to equations 5 and 10 of ref. [3] to see how the global hierarchy score was defined for CC and TC connections, respectively.). Specifically, the optimal mapping for CC connections assigns connections of cluster 2, 6 and 9 to one direction (feedback) and 1, 3, 4, 5, 7 and 8 to the opposite direction (feedforward). For TC connections, the most self-consistent hierarchy that maximizes the global hierarchy score is obtained when connections of cluster 2 and 6 correspond to feedback and the rest to feedforward patterns (figure 6a of ref. [3]).

With these mapping functions M_{CC} and M_{TC} obtained from the construction of the all-area hierarchy (figure 6a of ref. [3]), the hierarchical organization of the six visual cortical areas (V1, LM, AL, RL, PM and AM) and the two thalamic nuclei (LGN and LP) was constructed using only the connections among these eight regions. We first uncovered the cortical hierarchy using the intra-cortical connections among the six cortical areas: V1, LM, AL, RL, PM, and AM (240 unique 'source-target-Cre line' combinations). The initial hierarchical position of a cortical area is defined as:

$$H_i^0 = \frac{1}{2}(\langle M_{CC}(C_{T_{i,j}}) \cdot \mathrm{conf}(T)\rangle_j - \langle M_{CC}(C_{T_{j,i}}) \cdot \mathrm{conf}(T)\rangle_j), \tag{1}$$

where the first term describes the average direction of connections to area i, and thus represents the hierarchical position of the area as a target. The second term, on the other hand, represents the average direction of connections from area i, depicting the hierarchical position of the area as a source. To account for the Cre-line-specific bias, the Cre-dependent confidence measure, $\mathrm{conf}(T) = 1 - |\langle M_{CC}(C_{T_{i,j}})\rangle_{i,j}|$ is included. The initial hierarchy score (H_i^0) of each area i then is iterated using a two-step iterative scheme until the fixed point is reached:

$$H_i^{n-1/2} = \frac{1}{2}\{\langle H_j^{n-1} + M_{\text{cc}}(C_{T_{i,j}})\rangle_j - \langle -H_j^{n-1} + M_{\text{cc}}(C_{T_{j,i}})\rangle_j\} \tag{2}$$

$$H_i^n = H_i^{n-\frac{1}{2}} - \left\langle H_j^{n-\frac{1}{2}} \right\rangle_j \tag{3}$$

where *n* refers to iterative steps.

After hierarchical positions of cortical areas are found based on CC connections, the hierarchical positions of the LGN and LP relative to the cortical areas were computed by including TC connections from the LGN and LP to the six visual cortical areas (25 unique 'source-target-Cre line' combinations). Because thalamic areas are always the source in TC connections, the initial hierarchy score of each thalamic area *i* is defined by the average direction of connections from the area:

$$H_i^0 = -\left\langle M_{\text{TC}}(C_{T_{j,i}}) \cdot \frac{\min(N_{\text{ff}} \cdot N_{\text{fb}})}{N_{\text{ff}} + N_{\text{fb}}} \right\rangle_j \tag{4}$$

The parameters N_{ff} and N_{fb} refer to the numbers of feedforward and feedback thalamocortical connections, respectively. Once the initial positions of the thalamic areas in the hierarchy are obtained using equation (4), hierarchy scores of thalamic and cortical areas are iterated until the fixed points are reached, using a full mapping function $M_{\text{CC+TC}}$ that combines M_{CC} and M_{TC}, as done with the cortical hierarchy based on CC connections only (equations (2) and (3)).

To test the significance of the hierarchy levels of these areas, we generated 100 sampled connectivity data of the same size via bootstrapping, and computed the hierarchy scores of the eight regions using the bootstrapped connectivity data. We performed Wilcoxon paired signed rank-sum tests on these scores, showing that hierarchy levels of LM and RL cannot be meaningfully distinguished ($P = 0.08$) but the rest of the areas are at significantly distinct hierarchical positions, with the 5% confidence level.

Network model simulation.—To quantitatively evaluate the degree of 'hierarchy' of our measured functional network and to compare it to parallel network architectures, we performed a series of model simulations. We examined how the functional connectivity matrix would change with different network structures and calculated a 'total hierarchy score' (THS) to reflect the degree of hierarchy. The model is a simple graph model that assumes each area is a node, and the connection strength and directionality between nodes (feedforward and feedback connections) are defined by a simulated distribution of CCG peak offsets between the two areas. The peak offset distributions are approximated by Gaussian distributions, because most of the distributions of between area peak offsets are Gaussian-like (normality test with scipy.stats.normaltest; with $P > 0.05$). The distribution of offsets in the actual data has a mean of 1.1 ± 0.4 ms ($n = 5$ pairs of neighbouring areas) and a standard deviation of 3.7 ± 0.2 ms ($n = 15$ pairs of areas; Extended Data Fig. 7a). Inspired by the data, we simulated peak offset distributions between neighbouring levels

using Gaussian distributions with $\sigma = 4$ ms and $\mu = 1$ ms ($\mu = L_i - L_j$ between hierarchical levels i and j). The mean and standard deviation of the Gaussians define the directionality score (DS), which reflects the relative proportion of measured feedforward connections between two areas. When the Gaussian has a mean of 0, the DS is 0, which means the two nodes reside at the same level of the graph, while a DS of 1 indicates unidirectional information flow from the lower to the higher node.

In Extended Data Fig. 8a, we first quantified DS on the basis of the peak offset distributions from the experimental data. The left panel shows the distribution of peak offset from V1 to LM, between the cut-off times of ±10 ms that we impose to minimize multi-synaptic connections. In the middle is the functional connectivity matrix with the DS between areas (values range from −1 to 1). On the right is the mean of DS from each source area to all target areas. The mean DS gradually decreased along the anatomical hierarchy. The maximum difference of this trajectory (here between V1 and AM) is defined as the total hierarchy score (THS) of the network, which is 0.89 for the measured functional network in our data. For our simulations, we first tested a fully recurrent network in which all nodes have unbiased reciprocal connections (Extended Data Fig. 8b). This network has a DS connectivity matrix with all zeroes and a THS of 0. Then, we simulated a two-level, one-to-all network that models parallel feedforward projections from V1 to all other areas (Extended Data Fig. 8c), but with all other areas recurrently connected to each other in an unbiased way. This network generated a THS of 0.15. Next, we simulated a three-level network, assuming V1 is at the lowest level, RL, LM and AL are at the second level, AM and PM are at the top level (Extended Data Fig. 8d). This network generated a THS of 0.38. Next, we simulated a ladder hierarchical network, in which the mean of peak offset distribution between two areas is determined by their position difference in the hierarchy ($L_i - L_j$) (such that the mean time lag between neighbouring levels is 1 ms). We first tested a network with parameters constrained by our data ($\sigma = 4$ ms and $\mu = L_i - L_j$ ms) (Extended Data Fig. 8e). The resulting DS matrix showed a gradient very similar to the real data, with a THS equal to 0.88. To push the network to an extreme, we tested a strongly feedforward network by defining a narrow Gaussian ($\sigma = 1$) to produce fewer feedback connections (Extended Data Fig. 8f). We found the DS matrix was more saturated and the THS value is 1.54. In theory, the maximum THS is 2. Therefore, our measured network is more hierarchical than a 'one-to-all' network and less hierarchical than a purely feedforward, hierarchical network.

Other statistical methods.—To quantify the correlation between the mean value of each metric and the anatomical hierarchy score, both the Pearson correlation coefficient (scipy.stats.pearsonr) and Spearman's rank correlation coefficient (scipy.stats.spearmanr) were used.

To test for significant differences between pairs of areas, a Wilcoxon rank-sum statistic was used (scipy.stats.ranksum), with each unit considered an independent sample. Correction for multiple comparisons was performed using the Benjamini–Hochberg false discovery rate (statsmodels.stats.multitest.multipletests).

Data processing pipeline

Data for each session were uploaded to the Allen Institute Laboratory Information Management System (LIMS). Each dataset was run through the same series of processing steps using a set of project-specific workflows. Out of 61 sessions entering the processing pipeline, 58 resulted in successful NWB file generation. The three processing failures were due to mismatches in session identifiers or expected file structures that prevented the workflow from completing.

Reporting summary

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

Extended Data
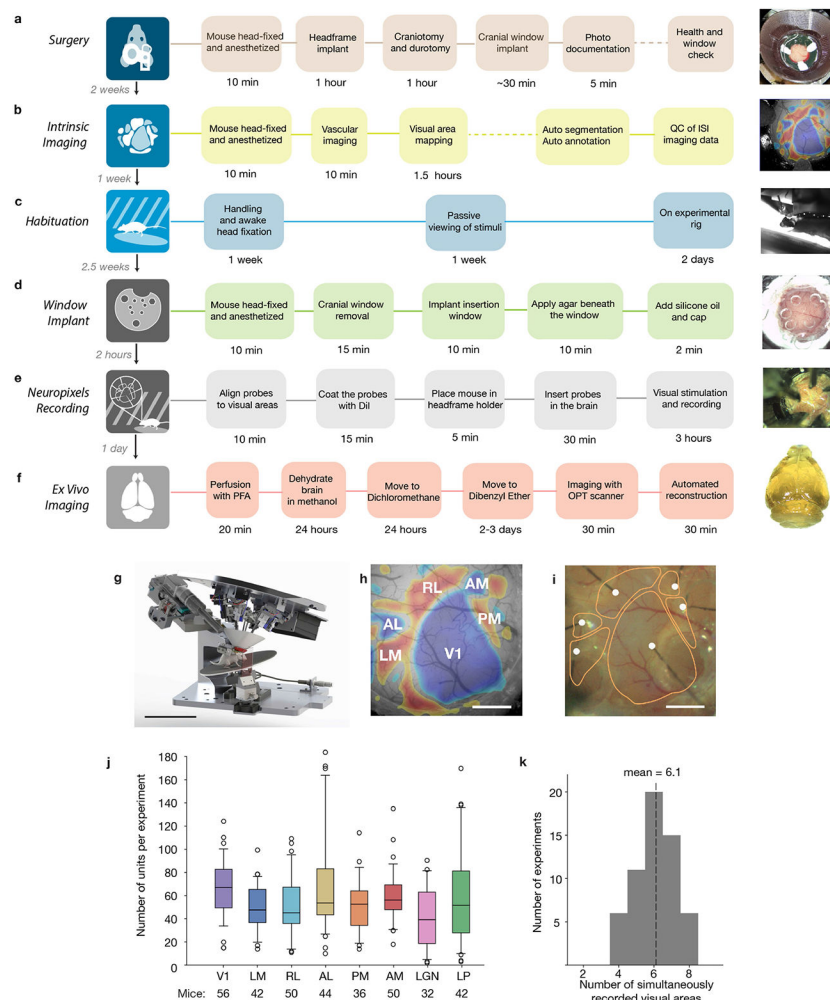
Extended Data Fig. 1 |. Pipeline procedures.
a–f, Summary of procedures involved in each step of the pipeline. **g**, Rig for parallel recording from six Neuropixels probes. Scale bar, 10 cm. **h**, Example retinotopic map used

for targeting probes to six cortical visual areas. Scale bar, 1 mm. **i**, Image of Neuropixels probes during an experiment, with area boundaries from **h** overlaid in orange. Probe tips are marked with white dots. Scale bar, 1 mm. **j**, Box plot of the number of units recorded per area per experiment, after filtering based on ISI violations (<0.5), amplitude cutoff (<0.1) and presence ratio (>0.95) (see Methods and Extended Data Fig. 4 for quality metric definitions and distributions). Box plot edges represent upper and lower quartiles; centre line represents the median; whiskers represent 5th to 95th percentile range; open circles represent any data points beyond the edge of the whiskers. **k**, Histogram of the number of simultaneously recorded cortical and thalamic visual areas per experiment ($n = 58$ experiments).
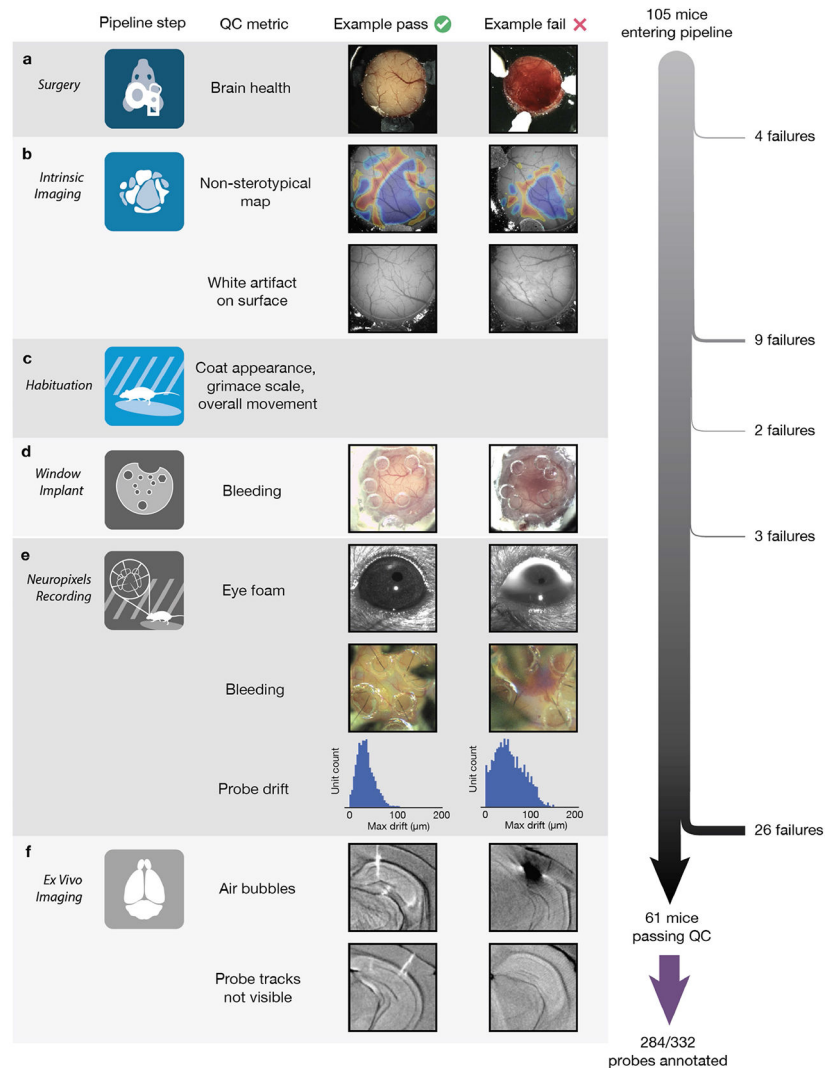
Extended Data Fig. 2 |. Pipeline quality control.
a–f, Major quality control metrics for each pipeline step, with examples of passing and failing experiments. The number of mice failing quality control at each stage is shown on the right.
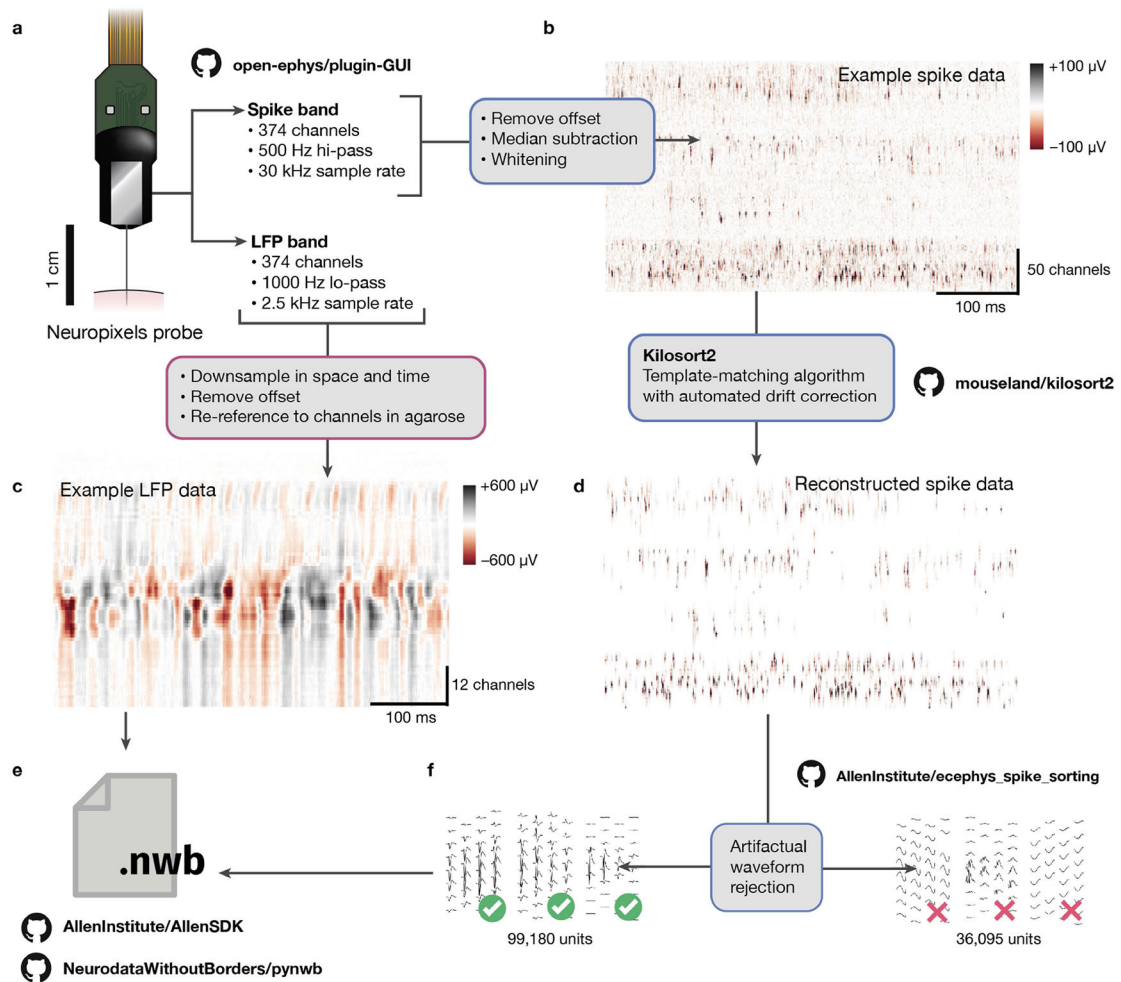
Extended Data Fig. 3 |. Data processing steps.
a, Data from the Neuropixels probe is split at the hardware level into two separate streams for each electrode: spike band and LFP band. **b**, The spike band passes through offset subtraction, median subtraction and whitening steps before sorting. The resulting data can be viewed as an image, with dimensions of time and channels, and colours corresponding to voltage levels. **c**, The LFP data are down sampled to 1.25 kHz and 40 μm channel spacing before packaging. **d**, We use the Kilosort2 to match spike templates to the raw data. The output of this algorithm can be used to reconstruct the original data using information about template shape, times and amplitudes. **e**, The spike and LFP data are packaged into Neurodata Without Borders (NWB) 2.0 files. **f**, The outputs of Kilosort2 are passed through a semi-automated quality control procedure to remove units with artefactual waveforms. Only units with obvious spike-like characteristics are used for further analysis.
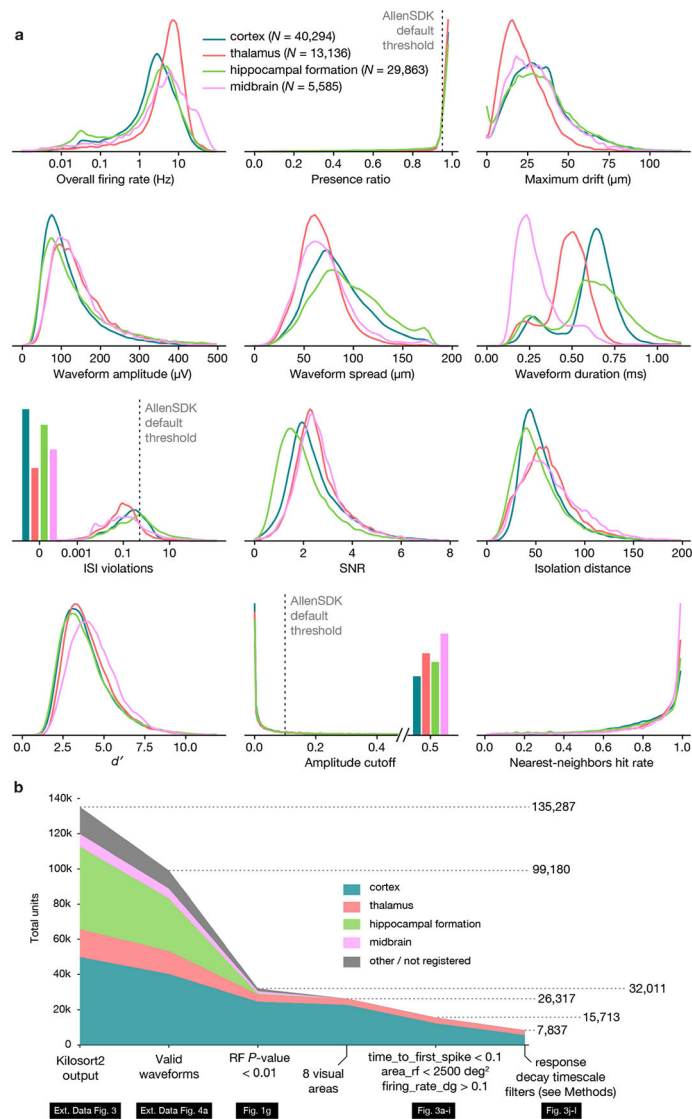
Extended Data Fig. 4 |. Unit quality metrics.

a, Density functions for twelve quality control metrics, plotted for units in cortex, hippocampus, thalamus and midbrain, aggregated across experiments. Default AllenSDK thresholds are shown as dotted lines. **b**, Unit selection flowchart for generating manuscript figures. Note that we do not use the default AllenSDK filters in this work, but instead use a receptive field *P* value of 0.01 as the primary metric for selecting units for analysis. CCFv3 structure labels used for region identification are as follows: cortex (VISp, VISl, VISrl, VISam, VISpm, VISal, VISmma, VISmmp, VISli, VIS), thalamus (LGd, LD, LP, VPM, TH, MGm, MGv, MGd, PO, LGv, VL, VPL, POL, Eth, PoT, PP, PIL, IntG, IGL, SGN, VPL, PF, RT), hippocampal formation (CA1, CA2, CA3, DG, SUB, POST, PRE, ProS, HPF), midbrain (MB, SCig, SCiw, SCsg, SCzo, SCop, PPT, APN, NOT, MRN, OP, LT, RPF), other/nonregistered (CP, ZI, grey).
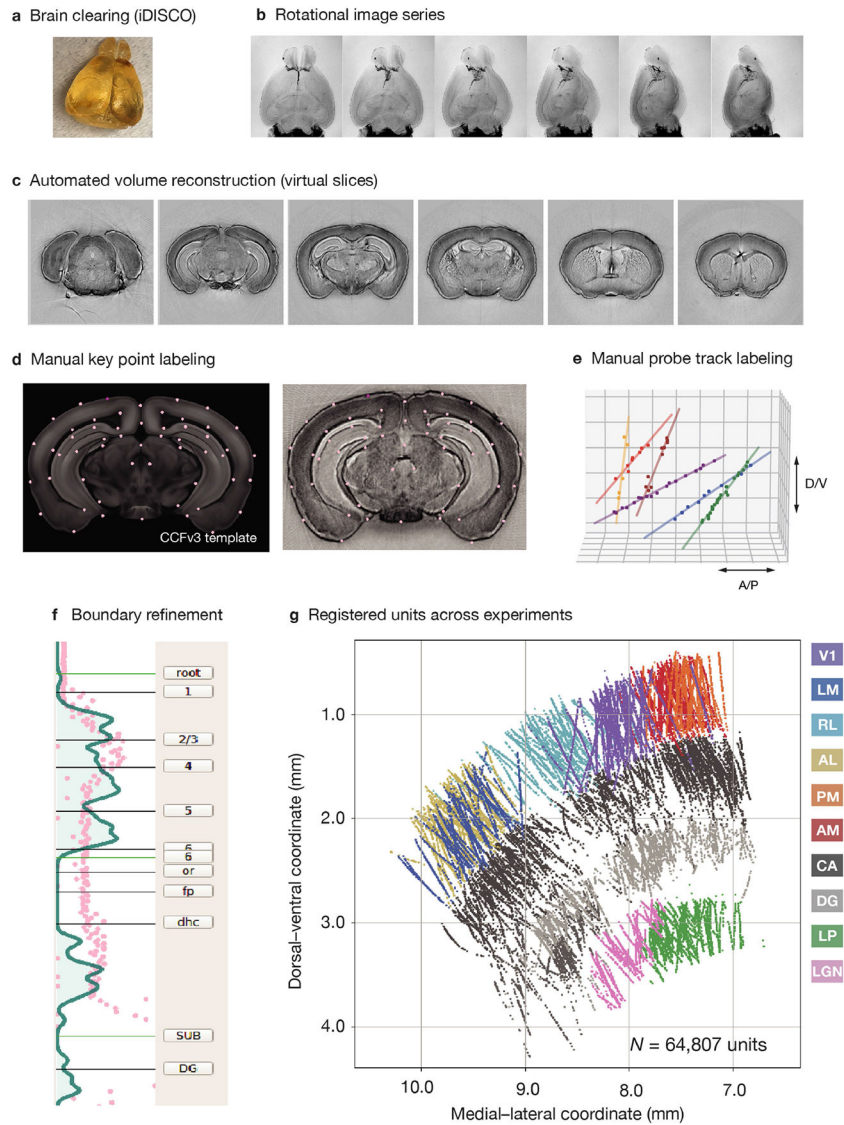
Extended Data Fig. 5 |. Aligning units with the Common Coordinate Framework (CCFv3).
a, After each experiment, the brain is removed and cleared using a variant of the iDISCO method. **b**, The cleared brain is imaged at 400 rotational angles using a custom-built optical projection tomography microscope. **c**, We generated an isotropic 3D volume from rotational images using a computational tomography algorithm. **d**, Key points from the CCFv3 template brain are manually identified in each individual brain. **e**, Points along each fluorescently labelled probe track are manually identified in the volume. Using the key points from **d**, we define a warping function to translate points along the probe axis into the Common Coordinate Framework. **f**, We then align the regional boundaries to boundaries in the physiological data, primarily the decrease in unit density at the border between the cortex and hippocampus, and between the hippocampus and thalamus. The shaded area represents unit density on each recording site, and pink dots represent low-frequency LFP power (<10 Hz) along the probe axis. **g**, Finally, units in the database are mapped to a 3D location in the CCFv3 and are assigned a structure label. Units in cortex are also assigned a

relative depth (0, surface; 1, white matter) and a layer label (L1, L2/3, L4, L5 or L6), on the basis of the annotation of the CCFv3 template volume (10-μm resolution).
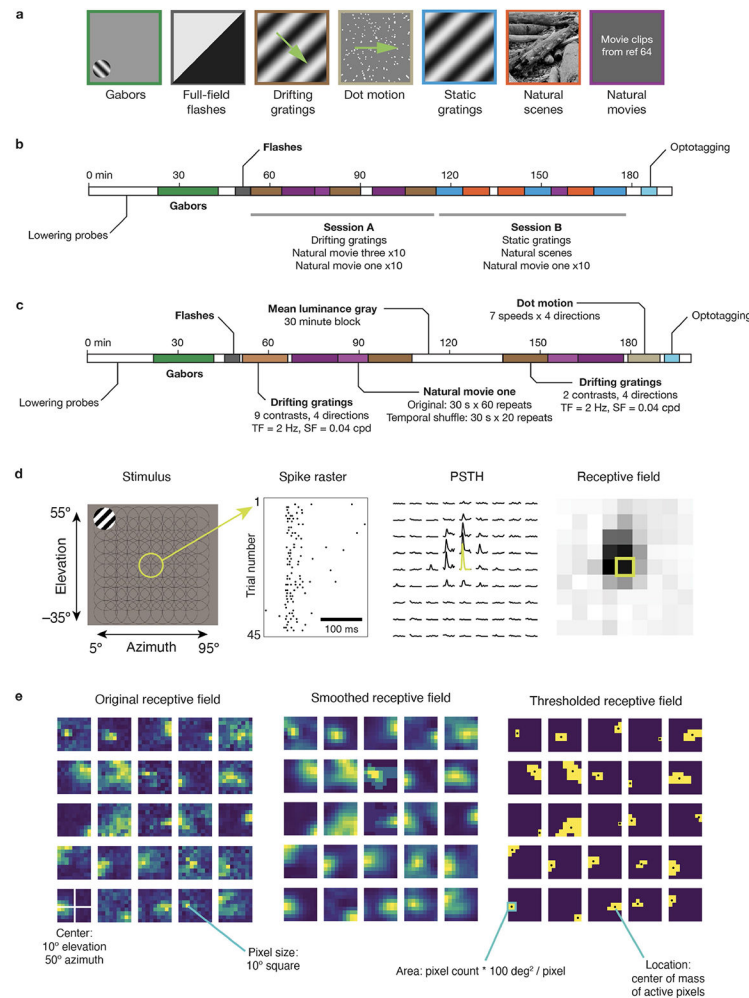
Extended Data Fig. 6 |. Details of the visual stimulus set and receptive field mapping procedure.
a, Example frames from each type of stimulus. Green arrows indicate direction of motion. The natural scene image is shown illustrative purposes. The natural scene images shown to the mice are from refs. [51] and [52]. **b**, Timing diagram for visual stimulus set #1, known as 'Brain Observatory 1.1'. **c**, Timing diagram for visual stimulus set #2, known as 'Functional Connectivity'. **d**, Receptive field mapping used 20° diameter drifting gratings flashed for 250 ms in each of 81 randomized locations on the screen. A spike raster for one unit shows the timing of spikes on each of 45 trials with the stimulus at a particular location. Collapsing over trials yields a peristimulus time histogram for each location. Collapsing over time yields a spike count for each spatial bin. A matrix of spike counts represents the receptive field for this unit. **e**, To calculate receptive field properties, the receptive field is first smoothed with a Gaussian filter, and all pixels above a threshold value are selected. The centre of mass of the above-threshold pixels indicates the receptive field location, while the total number of above-threshold pixels indicates the area. These processing steps are shown for 25 receptive fields randomly chosen from one experiment.
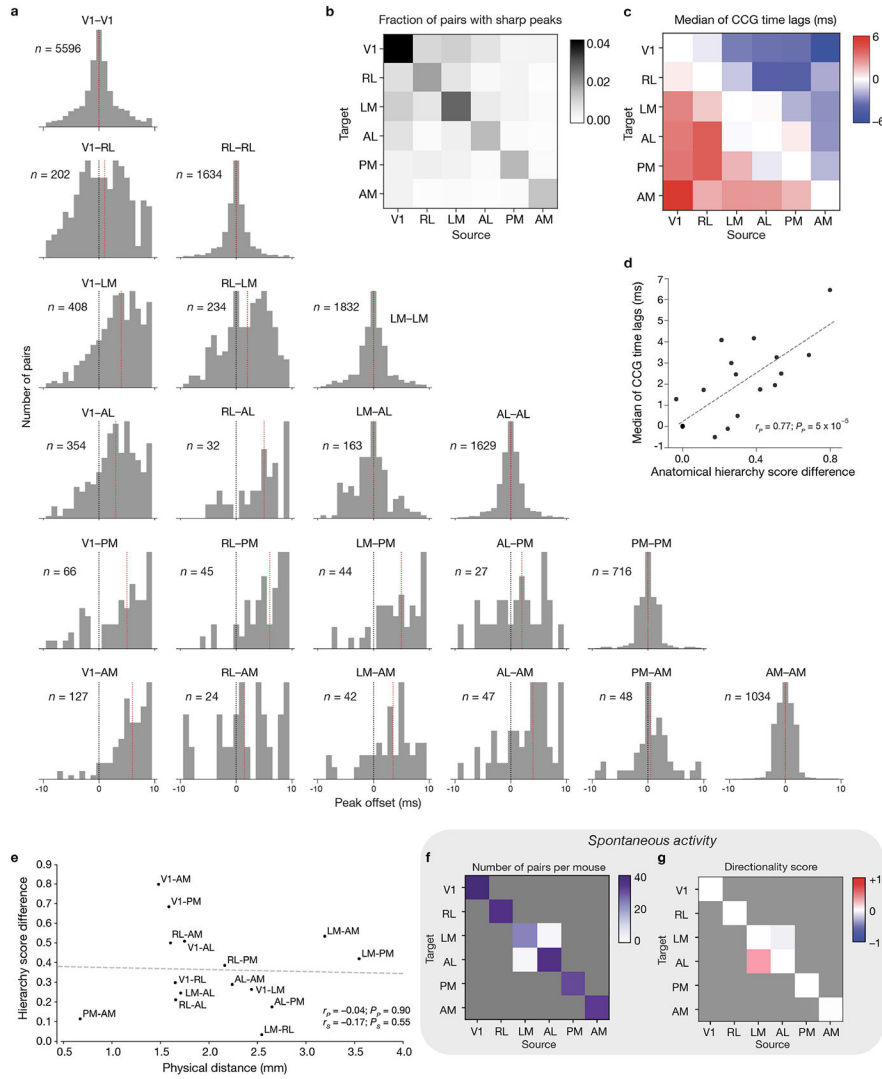
Extended Data Fig. 7 |. Functional connections between visual cortical areas.
a, Peak offset distributions aggregated across 25 mice for each area combination, during drifting gratings presentation. The total number of pairs (*n*) is labelled in each sub-panel. Dashed black line indicates zero time lag. Dashed red line indicates the median of the distribution. **b**, Fraction of within- and between-area unit pairs exhibiting sharp peaks, out of all simultaneously recorded pairs. **c**, Combined median of peak offsets across mice (averaged across mice; *n* = 25 mice in total) for each pair of cortical areas. **d**, Correlation between the median peak offset and the difference in hierarchy scores among 21 pairs (lower triangle and diagonal of the matrix). **e**, Relationship between average 3D Euclidean distance between units simultaneously recorded in each pair of areas (following registration to the CCFv3) and their hierarchy score difference. r_P, Pearson correlation coefficient; r_S, Spearman's rank correlation coefficient. **f**, Average number of sharp peak connections per mouse for jitter-corrected CCGs calculated during spontaneous activity (30 min grey screen period). Pixels masked with grey indicate no sharp peaks were detected. **g**, Average directionality score across mice during spontaneous activity.
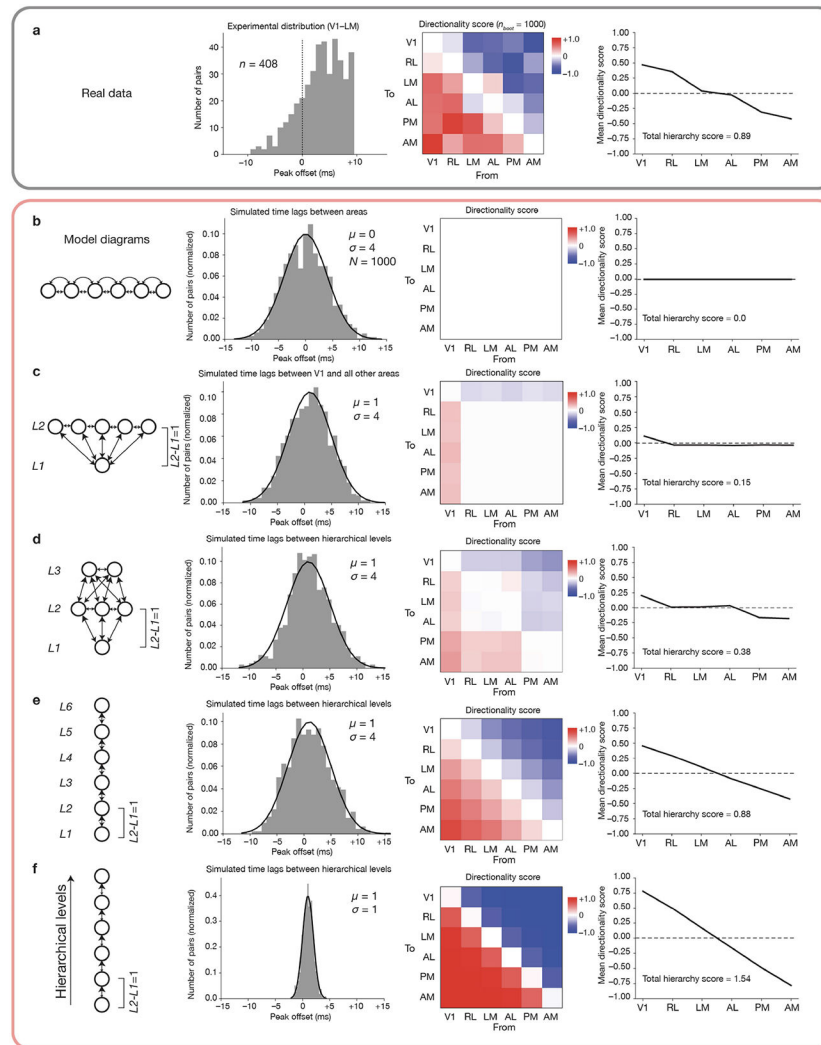
Extended Data Fig. 8 |. Simulation of functional connectivity profiles for different network structures.

a, Directionality score (DS) and total hierarchy score calculated from actual data. Left, an example distribution of peak offsets between V1 (source) and LM (target); middle, DS matrix for all area combinations; right, mean DS for each source area to all target areas, which gradually decreases along the hierarchy. The maximum difference of the mean DS across areas represents the total hierarchy score for the real network. **b–f**, Simulations based on different hypothetical network structures. Because the standard deviation of peak offset distribution in our measured CCG time lag distribution is 3.7 ± 0.2 ms and the median CCG time lag of neighbouring areas is 1.1 ± 0.4 ms, we simulated Gaussian distributions of the model peak offsets with $\sigma = 4$ and $\mu = 1$ for neighbouring hierarchical levels ($\mu = L_i - L_j$ between hierarchical levels i and j). See Methods for additional details of this simulation. **b**, A fully recurrent network where all nodes (areas) are at the same hierarchical level and have unbiased reciprocal connections ($\mu = 0$). **c**, A two-level, one-to-all network that models parallel feedforward projections from V1, with all other areas recurrently connected with one another in an unbiased way. **d**, A three-level network, assuming V1 at the lowest level,

RL, LM and AL at the second level, AM and PM at the top level. **e**, A six-level hierarchical network with each area at a distinct hierarchical level. Network parameters were constrained by real data ($\sigma = 4$ and $\mu = 1$ for neighbouring hierarchical levels, and $\mu = L_i - L_j$ between any hierarchical levels i and j). **f**, A six-level hierarchical network with a narrow distribution of peak offsets ($\sigma = 1$) that simulates a paucity of feedback connections.
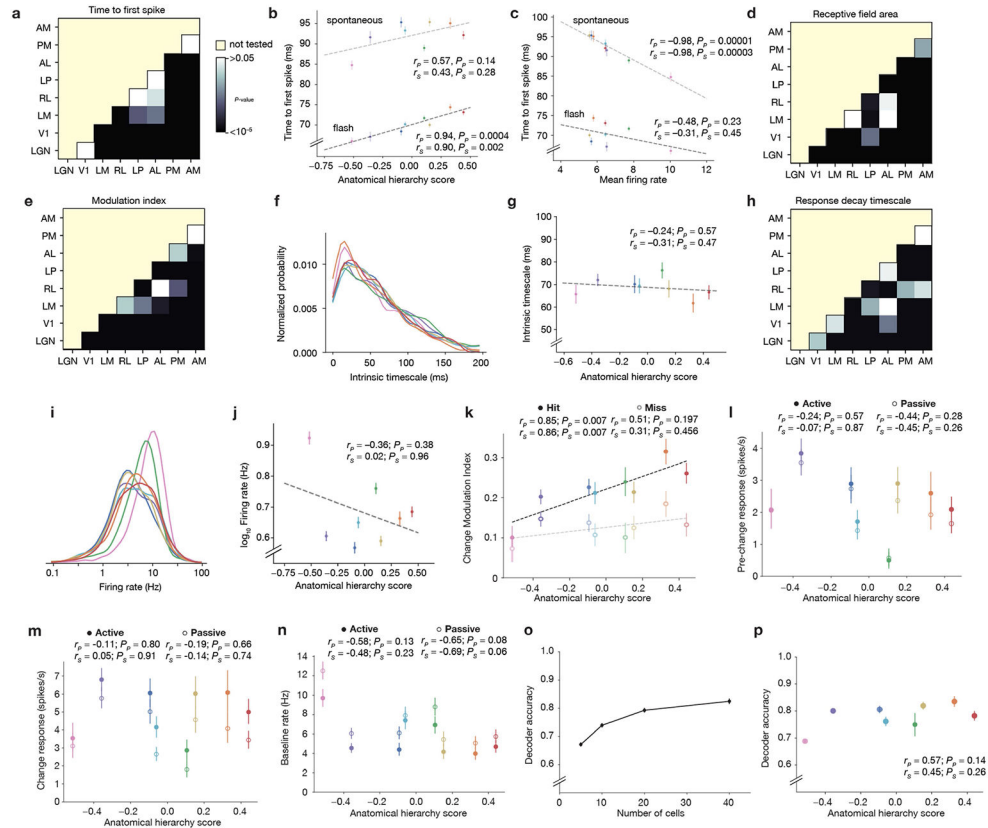
Extended Data Fig. 9 |. Statistics and additional analysis of hierarchy measures.
a, P values for pairwise comparisons of time to first spike between areas (two-sided Wilcoxon rank–sum test with Benjamini-Hochberg false discovery rate correction). **b**, Comparison between time-to-first-spike measured in response to the onset of the flash stimulus ('flash') versus during the inter-stimulus interval which corresponds to spontaneous firing ('spontaneous'). The colour scheme is the same as in Fig. 4; error bars represent mean $\pm$ 95% confidence intervals; $n = 15{,}713$ units from 58 mice. **c**, Relationship between time-to-first spike and mean firing rate for a given area, either in response to the flash stimulus, or during the inter-trial interval ('spontaneous'). **d**, P values for pairwise comparisons of receptive field size between areas. Colour scale is the same as in **a**. **e**, P values for pairwise comparisons of modulation index between area. Colour scale is the same as in **a**. **f**, Distribution of intrinsic timescale across units in each of 8 areas. **g**, Correlation between mean intrinsic timescale and anatomical hierarchy score. The absence of a significant correlation is inconsistent with the findings from ref. [13], in which it was shown that intrinsic timescale increases with hierarchical level in primates. This discrepancy may stem from differences between mouse and primate neocortex, or the fact that the areas we have

recorded do not span the full range of the mouse cortical hierarchy. In addition, it is known that standard exponential fitting procedures produce biased and unreliable timescale estimates, which may account for the null result we observed[91]. **h**, P-values for pairwise comparisons of response decay timescales between areas. Colour scale is the same as in **a**. **i**, Distribution of overall firing rates for all units in each area. **j**, Correlation between mean firing rate and anatomical hierarchy score. **k**, Relationship between change modulation index and anatomical hierarchy score, grouped by hit and miss trials. **l**, Relationship between pre-change response and anatomical hierarchy score, grouped by active and passive trials. **m**, Relationship between change response and anatomical hierarchy score, grouped by active and passive trials. **n**, Relationship between baseline firing rate and anatomical hierarchy score, grouped by active and passive trials. **o**, Decoder accuracy as a function of number of neurons used for decoding, averaged across all brain regions and behaviour sessions. **p**, Decoder accuracy for each brain region (mean ± s.e.m., averaged across sessions) is not correlated with the anatomical hierarchy score. r_P, Pearson correlation coefficient; r_S, Spearman's rank correlation coefficient.
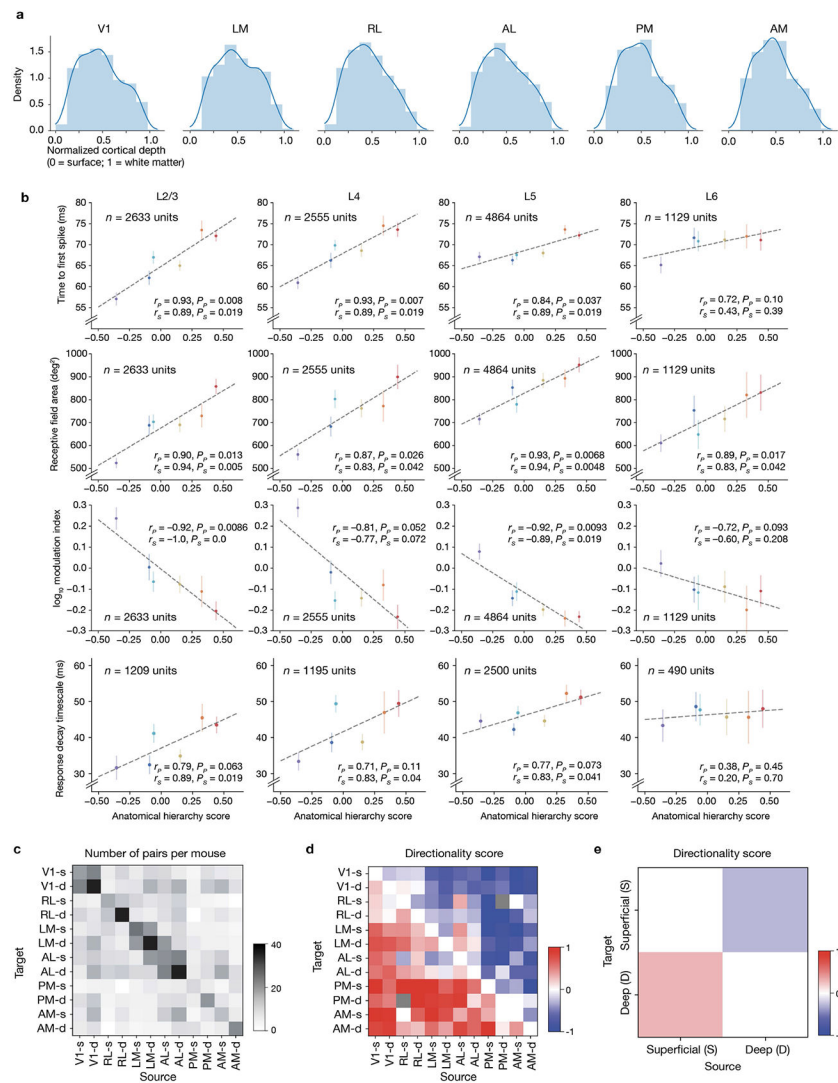
Extended Data Fig. 10 |. Layer-wise analysis.
a, Distribution of unit depths by area. 0 = surface, 1 = white matter. Normalized depth
is measured along lines normal to the cortical surface ('cortical streamlines'), rather than
distance along the probe. **b**, Time-to-first-spike, receptive field area, modulation index, and
response decay timescale analysed separately for each cortical layer. Colours are the same
as those used in Fig. 4. Error bars represent mean ± 95% bootstrap confidence intervals. On
average, in comparison to deep layers (5 and 6), superficial layers (2/3 and 4) had an earlier
time to first spike (2.59 ms difference, $P = 2.7 \times 10^{-19}$, two-sided Wilcoxon rank-sum test),
smaller receptive fields (109° difference, $P = 1.1 \times 10^{-33}$), higher modulation index (0.09
MI difference, $P = 5.0 \times 10^{-23}$), and faster response decay timescale (6.6 ms difference, $P
= 3.8 \times 10^{-33}$). The presence of slightly earlier spikes in L2/3 than L4 of V1 is probably
due to the existence of direct connections from LGN to L2/3 of this area[92]. r_P, Pearson
correlation coefficient; r_S, Spearman's rank correlation coefficient. **c**, Average number of
sharp peak pairs for each area and layer combination. Units in each area are bi-partitioned
into superficial (layers 2–4) and deep layers (layers 5–6). **d**, Directionality score (averaged

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

across mice) as an indicator of feedforward and feedback asymmetry. Areas ordered by hierarchy and layers arranged from superficial to deep. **e**, Directionality score based on average within-layer and between-layer distributions in **d**; superficial layers tend to drive deep layers within a cortical area.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Joshua H. Siegle[1,6], Xiaoxuan Jia[1,6], Séverine Durand[1], Sam Gale[1], Corbett Bennett[1], Nile Graddis[1], Greggory Heller[1], Tamina K. Ramirez[1], Hannah Choi[1,2], Jennifer A. Luviano[1], Peter A. Groblewski[1], Ruweida Ahmed[1], Anton Arkhipov[1], Amy Bernard[1], Yazan N. Billeh[1], Dillan Brown[1], Michael A. Buice[1], Nicolas Cain[1], Shiella Caldejon[1], Linzy Casal[1], Andrew Cho[1], Maggie Chvilicek[1], Timothy C. Cox[3], Kael Dai[1], Daniel J. Denman[1,4], Saskia E. J. de Vries[1], Roald Dietzman[1], Luke Esposito[1], Colin Farrell[1], David Feng[1], John Galbraith[1], Marina Garrett[1], Emily C. Gelfand[1], Nicole Hancock[1], Julie A. Harris[1], Robert Howard[1], Brian Hu[1], Ross Hytnen[1], Ramakrishnan Iyer[1], Erika Jessett[1], Katelyn Johnson[1], India Kato[1], Justin Kiggins[1], Sophie Lambert[1], Jerome Lecoq[1], Peter Ledochowitsch[1], Jung Hoon Lee[1], Arielle Leon[1], Yang Li[1], Elizabeth Liang[1], Fuhui Long[1], Kyla Mace[1], Jose Melchior[1], Daniel Millman[1], Tyler Mollenkopf[1], Chelsea Nayan[1], Lydia Ng[1], Kiet Ngo[1], Thuyahn Nguyen[1], Philip R. Nicovich[1], Kat North[1], Gabriel Koch Ocker[1], Doug Ollerenshaw[1], Michael Oliver[1], Marius Pachitariu[5], Jed Perkins[1], Melissa Reding[1], David Reid[1], Miranda Robertson[1], Kara Ronellenfitch[1], Sam Seid[1], Cliff Slaughterbeck[1], Michelle Stoecklin[1], David Sullivan[1], Ben Sutton[1], Jackie Swapp[1], Carol Thompson[1], Kristen Turner[1], Wayne Wakeman[1], Jennifer D. Whitesell[1], Derric Williams[1], Ali Williford[1], Rob Young[1], Hongkui Zeng[1], Sarah Naylor[1], John W. Phillips[1], R. Clay Reid[1], Stefan Mihalas[1], Shawn R. Olsen[1,7], Christof Koch[1,7]

Affiliations

[1]Allen Institute for Brain Science, Seattle, WA, USA.

[2]Department of Applied Mathematics, University of Washington, Seattle, WA, USA.

[3]University of Missouri-Kansas City School of Dentistry, Kansas City, MO, USA.

[4]The University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA.

[5]Janelia Research Campus, Ashburn, VA, USA.

[6]These authors contributed equally: Joshua H. Siegle, Xiaoxuan Jia.

[7]These authors jointly supervised this work: Shawn R. Olsen, Christof Koch.

Acknowledgements

Data availability

The data from all 58 passive viewing experiments used to generate main text Figs. 1-3 is available for download in Neurodata Without Borders (NWB) format via the AllenSDK. Example Jupyter Notebooks for accessing the data can be found at https://allensdk.readthedocs.io/en/latest/visual_coding_neuropixels.html.

The Neurodata Without Borders files are also available on the DANDI Archive (https://gui.dandiarchive.org/#/dandiset/000021; https://gui.dandiarchive.org/#/dandiset/000022)) and as an AWS public dataset (https://registry.opendata.aws/allen-brain-observatory/).

The metrics table used to generate Fig. 4e-h (active behaviour experiments) is available in the GitHub repository for this manuscript (https://github.com/AllenInstitute/neuropixels_platform_paper).

Code availability

Code for the following purposes are available from these repositories: generating manuscript figures, https://github.com/AllenIn-stitute/neuropixels_platform_paper; data pre-processing and unit metrics, https://github.com/AllenInstitute/ecephys_spike_sort-ing; spike-sorting, https://github.com/mouseland/Kilosort2; OPT post-processing, https://github.com/AllenInstitute/AIBSOPT; calculating stimulus metrics, https://github.com/AllenInstitute/AllenSDK; data acquisition, https://github.com/open-ephys/plugin-GUI, github.com/open-ephys-plugins/neuropixels-3a, https://github.com/open-ephys-plugins/neuropixels-PXI.

The following open-source software was used: NumPy[81], SciPy[82], IPython[83], Matplotlib[84], Pandas[85], xarray[86], scikit-learn[87], VTK[88], DeepLabCut[79,89], statsmodels[90], allenCCF[70], tifffile (https://pypi.org/project/tifffile/), Jupyter (https://jupyter.org/), pynwb (https://pynwb.readthedocs.io/en/stable/).

References

1. Felleman DJ & Van Essen DC Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47 (1991). [PubMed: 1822724]

2. de Vries SEJ et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. Nat. Neurosci 23, 138–151 (2020). [PubMed: 31844315]

3. Harris JA et al. Hierarchical organization of cortical and thalamic connectivity. Nature 575, 195–202 (2019). [PubMed: 31666704]

4. Carandini M et al. Do we know what the early visual system does? J. Neurosci 25, 10577–10597 (2005). [PubMed: 16291931]

5. Olshausen BA & Field DJ How close are we to understanding V1? Neural Comput. 17, 1665–1699 (2005). [PubMed: 15969914]

6. Jun JJ et al. Fully integrated silicon probes for high-density recording of neural activity. Nature 551, 232–236 (2017). [PubMed: 29120427]

7. Hubel DH & Wiesel TN Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (Lond.) 160, 106–154 (1962). [PubMed: 14449617]

8. Fukushima K Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern 36, 193–202 (1980). [PubMed: 7370364]

9. Krizhevsky A, Sutskever I & Hinton GE ImageNet classification with deep convolutional neural networks. in Proc. 25th International Conference on Neural Information Processing Systems (eds Pereira F et al.) 1097–1105 (NeurIPS, 2012).

10. Riesenhuber M & Poggio T Hierarchical models of object recognition in cortex. Nat. Neurosci 2, 1019–1025 (1999). [PubMed: 10526343]

11. Bullier J Integrated model of visual processing. Brain Res. Rev 36, 96–107 (2001). [PubMed: 11690606]

12. Chaudhuri R, Knoblauch K, Gariel M-A, Kennedy H & Wang X-J A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. Neuron 88, 419–431 (2015). [PubMed: 26439530]

13. Murray JD et al. A hierarchy of intrinsic timescales across primate cortex. Nat. Neurosci 17, 1661–1663 (2014). [PubMed: 25383900]

14. Rockland KS & Pandya DN Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res. 179, 3–20 (1979). [PubMed: 116716]

15. Schmolesky MT et al. Signal timing across the macaque visual system. J. Neurophysiol 79, 3272–3278 (1998). [PubMed: 9636126]

16. Yamins DLK & DiCarlo JJ Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci 19, 356–365 (2016). [PubMed: 26906502]

17. G m nu R et al. The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. Neuron 97, 698–715. e10 (2018). [PubMed: 29420935]

18. Glickfeld LL & Olsen SR Higher-order areas of the mouse visual cortex. Annu. Rev. Vis. Sci 3, 251–273 (2017). [PubMed: 28746815]

19. Wang Q, Sporns O & Burkhalter A Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. J. Neurosci 32, 4386–4399 (2012). [PubMed: 22457489]

20. Wang Q & Burkhalter A Area map of mouse visual cortex. J. Comp. Neurol 502, 339–357 (2007). [PubMed: 17366604]

21. Han Y et al. The logic of single-cell projections from visual cortex. Nature 556, 51–56 (2018). [PubMed: 29590093]

22. Allen WE et al. Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. Science 364, 253 (2019). [PubMed: 30948440]

23. Steinmetz NA, Zatka-Haas P, Carandini M & Harris KD Distributed coding of choice, action and engagement across the mouse brain. Nature 576, 266–273 (2019). [PubMed: 31776518]

24. Stringer C et al. Spontaneous behaviors drive multidimensional, brainwide activity. Science 364, 255 (2019). [PubMed: 31000656]

25. Siegle JH et al. Reconciling functional differences in populations of neurons recorded with two-photon imaging and electrophysiology. Preprint at 10.1101/2020.08.10.244723 (2020).

26. Pachitariu M, Steinmetz NA, Kadir SN, Carandini M & Harris KD Fast and accurate spike sorting of high-channel count probes with KiloSort. In Advances in Neural Information Processing Systems 29 (eds Lee D et al.) 4448–4456 (NeurIPS, 2016).

27. Wang Q et al. The Allen Mouse Brain Common Coordinate Framework: a 3D reference atlas. Cell 181, 936–953.e20 (2020). [PubMed: 32386544]

28. Jia X, Tanabe S & Kohn A γ and the coordination of spiking activity in early visual cortex. Neuron 77, 762–774 (2013). [PubMed: 23439127]

29. Smith MA & Kohn A Spatial and temporal scales of neuronal correlation in primary visual cortex. J. Neurosci 28, 12591–12603 (2008). [PubMed: 19036953]

30. Zandvakili A & Kohn A Coordinated neuronal activity enhances corticocortical communication. Neuron 87, 827–839 (2015). [PubMed: 26291164]

31. Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP & Movshon JA A functional and perceptual signature of the second visual area in primates. Nat. Neurosci 16, 974–981 (2013). [PubMed: 23685719]

32. Hubel D Eye, Brain, and Vision Vol. 22 (Scientific American Press, 1988).

33. Lennie P Single units and visual cortical organization. Perception 27, 889–935 (1998). [PubMed: 10209632]

34. Matteucci G, Bellacosa Marotti R, Riggi M, Rosselli FB & Zoccolan D Nonlinear processing of shape information in rat lateral extrastriate cortex. J. Neurosci 39, 1649–1670 (2019). [PubMed: 30617210]

35. Wypych M et al. Standardized F1: a consistent measure of strength of modulation of visual responses to sine-wave drifting gratings. Vision Res. 72, 14–33 (2012). [PubMed: 23000273]

36. Runyan CA, Piasini E, Panzeri S & Harvey CD Distinct timescales of population coding across cortex. Nature 548, 92–96 (2017). [PubMed: 28723889]

37. Garrett M et al. Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. eLife 9, e50340 (2020). [PubMed: 32101169]

38. Groblewski PA et al. Characterization of learning, motivation, and visual perception in five transgenic mouse lines expressing GCaMP in distinct cell populations. Front. Behav. Neurosci 14, 104 (2020). [PubMed: 32655383]

39. Grimm S, Escera C, Slabu L & Costa-Faidella J Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. Psychophysiology 48, 377–384 (2011). [PubMed: 20636288]

40. Dürschmid S et al. Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. Proc. Natl Acad. Sci. USA 113, 6755–6760 (2016). [PubMed: 27247381]

41. Vinken K, Vogels R & Op de Beeck H Recent visual experience shapes visual processing in rats through stimulus-specific adaptation and response enhancement. Curr. Biol 27, 914–919 (2017). [PubMed: 28262485]

42. Koch C & Reid RC Observatories of the mind. Nature 483, 397–398 (2012). [PubMed: 22437592]

43. Issa EB, Cadieu CF & DiCarlo JJ Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. eLife 7, e42870 (2018). [PubMed: 30484773]

44. Keller GB & Mrsic-Flogel TD Predictive processing: a canonical cortical computation. Neuron 100, 424–435 (2018). [PubMed: 30359606]

45. Zhuang J et al. An extended retinotopic map of mouse cortex. eLife 6, e18372 (2017). [PubMed: 28059700]

46. Maunsell JHR Functional visual streams. Curr. Opin. Neurobiol 2, 506–510 (1992). [PubMed: 1525550]

47. Ungerleider L & Mishkin M in Analysis of Visual Behavior (eds Ingle DJ, Goodale MA & Mansfield RJW) 549–586 (MIT Press, 1982).

48. D'Souza RD et al. Canonical and noncanonical features of the mouse visual cortical hierarchy. Preprint at 10.1101/2020.03.30.016303 (2020).

49. Murakami T, Matsui T & Ohki K Functional segregation and development of mouse higher visual areas. J. Neurosci 37, 9424–9437 (2017). [PubMed: 28847805]

50. Smith IT, Townsend LB, Huh R, Zhu H & Smith SL Stream-dependent development of higher visual cortical areas. Nat. Neurosci 20, 200–208 (2017). [PubMed: 28067905]

51. van Hateren JH & van der Schaaf A Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. Lond. B 265, 359–366 (1998).

52. Olmos A & Kingdom FAA A biologically inspired algorithm for the recovery of shading and reflectance images. Perception 33, 1463–1473 (2004). [PubMed: 15729913]

References

53. Lima SQ, Hromádka T, Znamenskiy P & Zador AM PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. PLoS ONE 4, e6099 (2009). [PubMed: 19584920]

54. Madisen L et al. A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. Nat. Neurosci 15, 793–802 (2012). [PubMed: 22446880]

55. Zhang F, Wang L-P, Boyden ES & Deisseroth K Channelrhodopsin-2 and optical control of excitable cells. Nat. Methods 3, 785–792 (2006). [PubMed: 16990810]

56. Goldey GJ et al. Removable cranial windows for long-term imaging in awake mice. Nat. Protoc 9, 2515–2538 (2014). [PubMed: 25275789]

57. Juavinett AL, Nauhaus I, Garrett ME, Zhuang J & Callaway EM Automated identification of mouse visual areas with intrinsic signal imaging. Nat. Protoc 12, 32–43 (2017). [PubMed: 27906169]

58. Kalatsky VA & Stryker MP New paradigm for optical imaging: temporally encoded maps of intrinsic signal. Neuron 38, 529–545 (2003). [PubMed: 12765606]

59. Garrett ME, Nauhaus I, Marshel JH & Callaway EM Topography and areal organization of mouse visual cortex. J. Neurosci 34, 12587–12600 (2014). [PubMed: 25209296]

60. Fiáth R et al. Slow insertion of silicon probes improves the quality of acute neuronal recordings. Sci. Rep 9, 111 (2019). [PubMed: 30643182]

61. Siegle JH et al. Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology. J. Neural Eng 14, 045003 (2017). [PubMed: 28169219]

62. Peirce JW PsychoPy—Psychophysics software in Python. J. Neurosci. Methods 162, 8–13 (2007). [PubMed: 17254636]

63. Martin D, Fowlkes C, Tal D & Malik J A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proc. Eighth IEEE International Conference on Computational Vision 416–423 (IEEE, 2001).

64. Welles O Touch of Evil (Universal - International, 1958).

65. Renier N et al. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. Cell 159, 896–910 (2014). [PubMed: 25417164]

66. Nguyen D et al. Optical projection tomography for rapid whole mouse brain imaging. Biomed. Opt. Express 8, 5637–5650 (2017). [PubMed: 29296493]

67. Sharpe J et al. Optical projection tomography as a tool for 3D microscopy and gene expression studies. Science 296, 541–545 (2002). [PubMed: 11964482]

68. Wong MD, Dazai J, Walls JR, Gale NW & Henkelman RM Design and implementation of a custom built optical projection tomography system. PLoS ONE 8, e73491 (2013). [PubMed: 24023880]

69. Edelstein AD et al. Advanced methods of microscope control using μManager software. J. Biol. Methods 1, 10 (2014).

70. Shamash P, Carandini M, Harris KD & Steinmetz NA A tool for analyzing electrode tracks from slice histology. Preprint at 10.1101/447995 (2018).

71. Jia X et al. High-density extracellular probes reveal dendritic backpropagation and facilitate neuron classification. J. Neurophysiol 121, 1831–1847 (2019). [PubMed: 30840526]

72. Hill DN, Mehta SB & Kleinfeld D Quality metrics to accompany spike sorting of extracellular signals. J. Neurosci 31, 8699–8705 (2011). [PubMed: 21677152]

73. Suner S, Fellows MR, Vargas-Irwin C, Nakata GK & Donoghue JP Reliability of signals from a chronically implanted, silicon-based electrode array in non-human primate primary motor cortex. IEEE Trans. Neural Syst. Rehabil. Eng 13, 524–541 (2005). [PubMed: 16425835]

74. Schmitzer-Torbert N, Jackson J, Henze D, Harris K & Redish AD Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience 131, 1–11 (2005). [PubMed: 15680687]

75. Chung JE et al. A fully automated approach to spike sorting. Neuron 95, 1381–1394.e6 (2017). [PubMed: 28910621]

76. Gerstein GL & Perkel DH Mutual temporal relationships among neural spike trains. Biophys. J 12, 453–473 (1972). [PubMed: 5030560]

77. Harrison MT & Geman S A rate and history-preserving resampling algorithm for neural spike trains. Neural Comput. 21, 1244–1258 (2009). [PubMed: 19018703]

78. Matteucci G, Bellacosa Marotti R, Riggi M, Rosselli FB & Zoccolan D Nonlinear processing of shape information in rat lateral extrastriate cortex. J. Neurosci 39, 1649–1670 (2019). [PubMed: 30617210]

79. Mathis A et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci 21, 1281–1289 (2018). [PubMed: 30127430]

80. Halir R & Flusser J Numerically stable direct least squares fitting of ellipses. In Proc. Sixth International Conference in Central Europe on Computer Graphics and Visualization (WSCG, 1998).

81. van der Walt S, Colbert SC & Varoquaux G The NumPy array: a structure for efficient numerical computation. Comput. Sci. Eng 13, 22–30 (2011).

82. Virtanen P et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17, 261–272 (2020). [PubMed: 32015543]

83. Pérez F & Granger BE IPython: a system for interactive scientific computing. Comput. Sci. Eng 9, 21–29 (2007).

84. Hunter JD Matplotlib: a 2D graphics environment. Comput. Sci. Eng 9, 90–95 (2007).

85. McKinney W Data structures for statistical computing in Python. In Proc. 9th Python in Science Conference (eds van der Walt S & Millman J) 51–56 (SciPy, 2010).

86. Hoyer S & Hamman J xarray: N–D labeled arrays and datasets in Python. J. Open Res. Softw 5, 10 (2017).

87. Pedregosa F et al. Scikit-learn: machine learning in Python. J. Mach. Learn. Res 12, 2825–2830 (2011).

88. Schroeder W, Martin K & Lorensen B The Visualization Toolkit 4th edn (Kitware, 2006).

89. Nath T et al. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. Nat. Protoc 14, 2152–2176 (2019). [PubMed: 31227823]

90. Seabold S & Perktold J Statsmodels: Econometric and statistical modeling with Python. In Proc. 9th Python in Science Conference (eds van der Walt S & Millman J) 92–96 (SciPy, 2010).

91. Zeraati R, Engel TA & Levina A Estimation of autocorrelation timescales with approximate Bayesian computations. Preprint at 10.1101/2020.08.11.245944 (2020).

92. Morgenstern NA, Bourg J & Petreanu L Multilaminar networks of cortical neurons integrate common inputs from sensory thalamus. Nat. Neurosci 19, 1034–1040 (2016). [PubMed: 27376765]
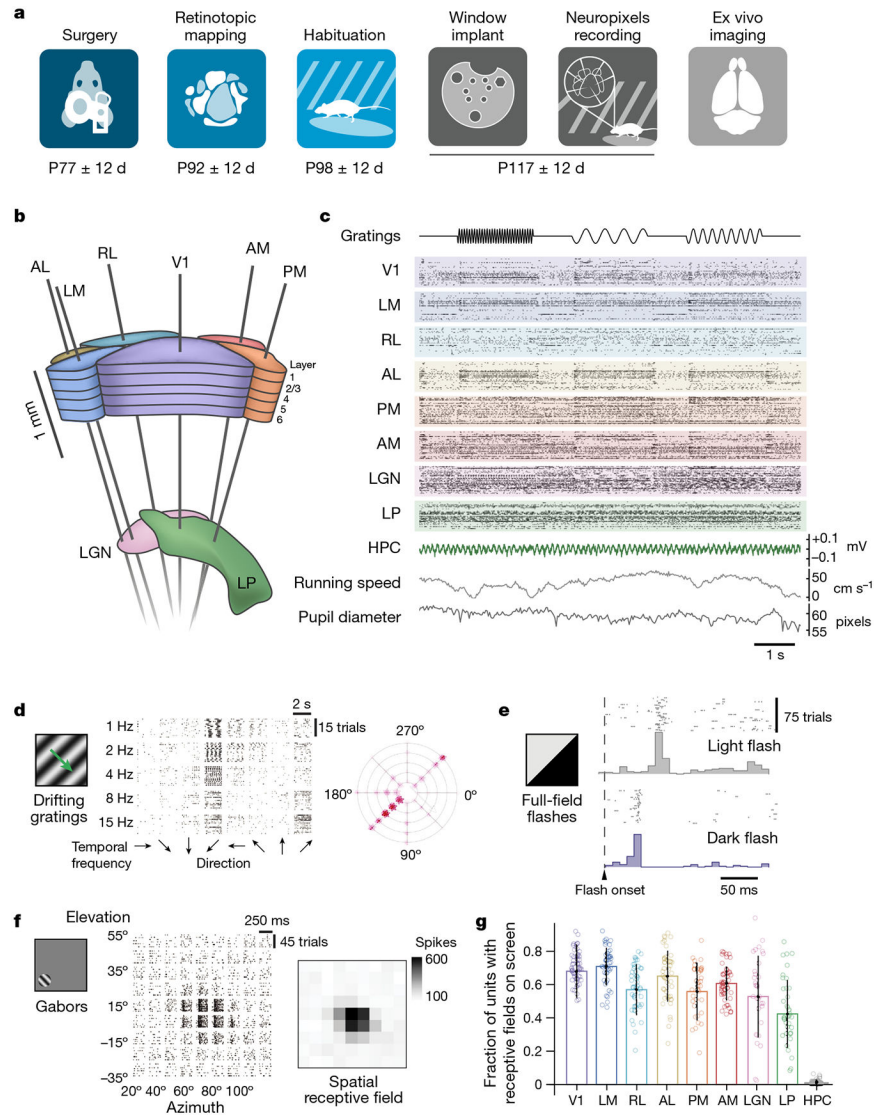
Fig. 1 |. A standardized pipeline for electrophysiology in the mouse visual system.
a, Data collection pipeline, with the average age of mice (in days) indicated below. **b**,
Schematic of probe insertion trajectories through visual cortical (V1, LM, AL, RL, AM,
PM) and thalamic (LGN, LP) areas. **c**, Example raster plot of 405 simultaneously recorded
units from 8 visual areas during drifting grating stimuli (15 Hz, 2 Hz or 4 Hz), with
hippocampal (HPC) local field potential, mouse running speed and pupil diameter shown
below. **d**, Raster plots of spike times for different drifting grating stimuli from an exemplar
V1 unit. Single-trial responses are represented by a star plot (right), in which stimulus
orientation and temporal frequency are indicated by angle and radius, respectively, and
firing rate is indicated by the intensity of the pink blob. **e**, Raster plots and peri-stimulus
time histograms of the full-field flash stimulus, for the same unit as in **d**. **f**, Raster plot
of spike times for 81 conditions of the Gabor stimulus for the same unit. Summing the
spike counts across 45 trials at each location produces a spatial receptive field, shown on
the right. Spike count is quantified over a 250-ms window. **g**, Mean fraction of units with

significant receptive field across eight visual areas, with hippocampus included as a control (see Methods). Data are mean ± s.d., and dots show the results of individual sessions.
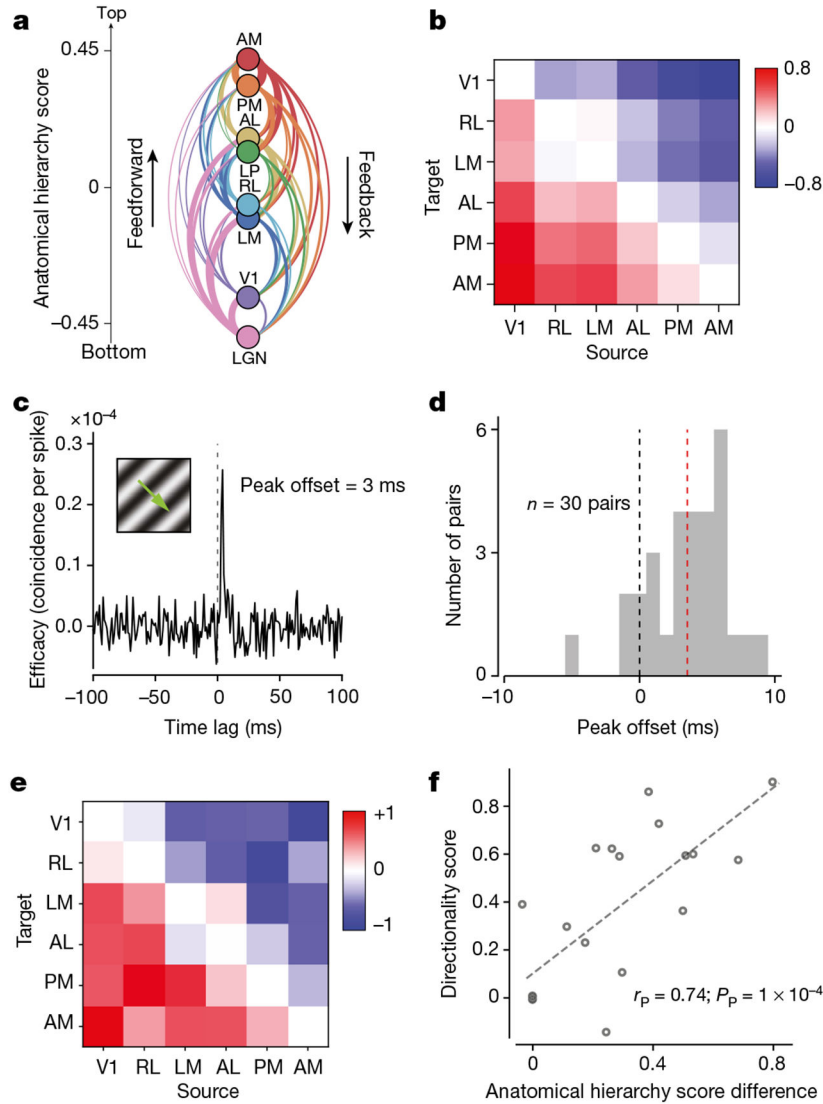
Fig. 2 |. Functional connectivity recapitulates the anatomical hierarchy.

a, Anatomical hierarchy scores of the eight areas of Interest recomputed from ref. [3].
b, Replotting the anatomical hierarchy scores from **a**, showing the difference in score between all cortical areas. All areas have significantly different anatomical hierarchy scores, except for RL and LM (Wilcoxon rank-sum test, $P = 0.08$; see Methods). **c**, An example cross-area 'sharp peak' spiking interaction between a pair of units in V1 and LM. **d**, Distribution of CCG peak time lags between V1 and LM in one example mouse. The median (3.9 ms) is shown by the red line. **e**, Directionality scores calculated from peak offset distributions across 25 mice for each pair of cortical areas. Statistical testing (two-sided Wilcoxon rank-sum test) revealed that the peak offset distributions of neighbouring areas were significantly different from within-area distributions, except for AL–PM ($P = 0.08$). **f**, Correlation between directionality score and anatomical hierarchy score difference ($n = 21$ pairs; lower triangle and diagonal of the matrices in **b** and **e**), indicating a link between structure and function. r_P, Pearson correlation coefficient.
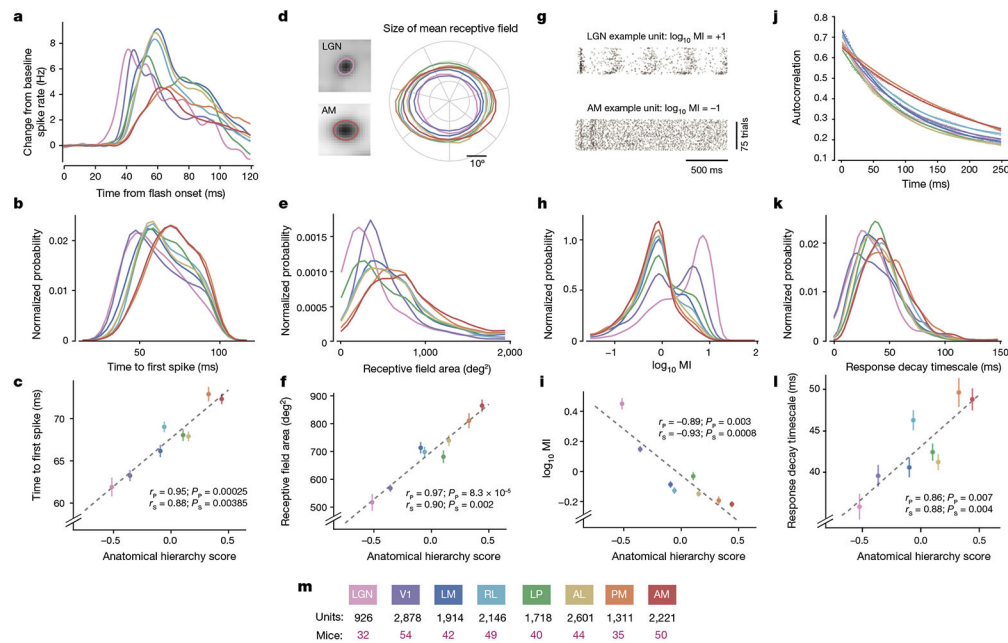
Fig. 3 |. Four measures of hierarchical processing applied to the mouse visual system.
a, Mean response (baseline-subtracted) to a full-field flash stimulus for units in eight visual regions. **b**, Distribution of time to first spike in response to the flash stimulus across all units in each of eight areas. **c**, Correlation between mean time to first spike and hierarchy score obtained from anatomical tracing studies. **d**, Outlines of the extent of the mean receptive field for each area, at 50% of the peak firing rate. Example mean receptive fields for the LGN and the AM are shown on the left. **e**, Distribution of receptive field sizes across all units. **f**, Correlation between mean receptive field size and anatomical hierarchy score. **g**, Raster plots showing the response of exemplar LGN and AM units to a 2-Hz drifting grating stimulus, with corresponding modulation index (MI). **h**, Distribution of modulation index across all units. **i**, Correlation between mean modulation index and anatomical hierarchy score. **j**, Mean autocorrelation averaged across all units in each area in the 250-ms period following the onset of a full-field flash stimulus. **k**, Distribution of response decay timescales across all units in each area. **l**, Correlation between mean response decay timescales and anatomical hierarchy score; $n = 7,837$ units from 58 mice. **m**, Key indicating the colour code used in the graphs, the number of units per area and the total number of mice per area. See Extended Data Fig. 4b for unit selection criteria. Data are mean ± 95% bootstrap confidence intervals. $n = 15,713$ units from 58 mice unless otherwise specified. r_S, Spearman correlation coefficient.
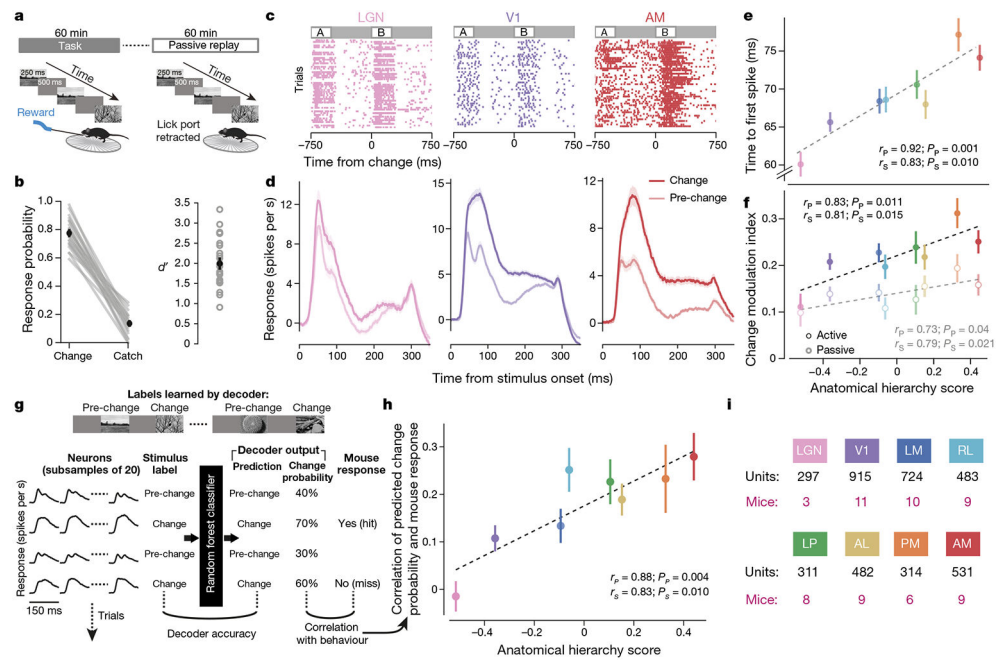
Fig. 4 |. Higher-order areas signal behaviourally relevant changes in image identity more strongly than lower-order areas.

a, Experimental setup for the active (left) and passive (right) change detection tasks. **b**, After training, mice had high hit and low false alarm rates, with an average d' of 2.0 ± 0.1 (n = 12 mice, 21 sessions). **c**, Raster plots of exemplar units from the LGN, V1 and AM before and after change (n = 50 trials). **d**, Population response averaged over all units in the LGN, V1 and AM. For each area, the response to the change and pre-change image is shown as a darker and lighter line, respectively. The line represents the mean and the shaded areas represent s.e.m. **e**, Correlation between mean time to first spike after image change and anatomical hierarchy score across all eight areas; data are mean ± 95% bootstrap confidence intervals. **f**, Correlation between mean change modulation index and anatomical hierarchy score across all eight areas. Closed circles indicate responses during active behaviour, and open circles indicate responses during passive stimulus replay; data are mean ± 95% bootstrap confidence intervals. **g**, Schematic of random forest decoding analysis to identify change versus non-change trials, and comparison with mouse behaviour. **h**, Pearson correlation of decoder prediction (change probability) and mouse behavioural response (hit/miss) across trials. Data are mean ± s.e.m. across sessions; see Methods for details of included units. **i**, Key indicating the colour code used in the graphs in **e, f** and **h**, the number of units per area and the total number of mice per area. The natural scene images in **a** and **g** are shown for schematic purposes. The images shown to the mice are from refs. 51,52.