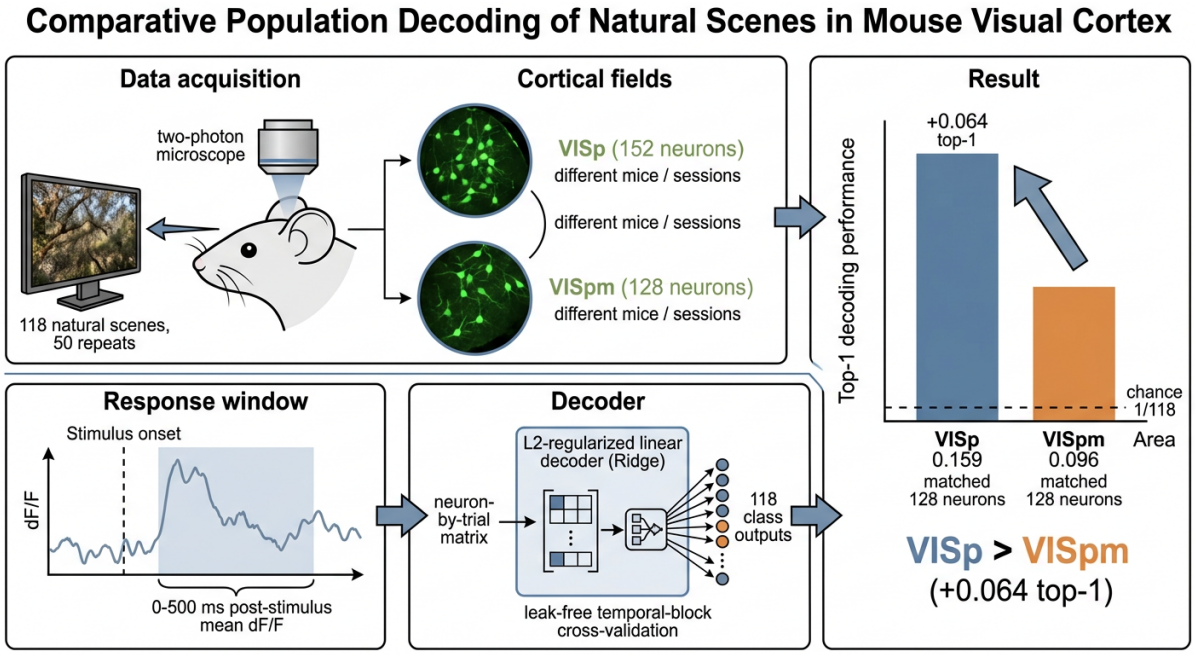
Comparative Population Decoding of Natural Scenes in Mouse Visual Cortex (VISp vs. VISpm)

July 11, 2026



Graphical abstract. Two-photon calcium responses to 118 natural scenes are decoded with a leak-free, L2-regularized linear decoder in primary visual cortex (VISp) and the higher visual area VISpm imaged in separate mice; at a matched neuron count VISp carries a substantially more decodable representation of scene identity.

Abstract

Higher visual areas of the mouse cortex are thought to transform the comparatively general-purpose representation of primary visual cortex (VISp) into more specialized codes, but how much information about the identity of a complex natural scene survives this transformation is not obvious *a priori*. Using the Allen Brain Observatory Visual Coding two-photon survey, we decoded which of 118 natural-scene images was presented on each trial from populations of layer-2/3 excitatory neurons expressing GCaMP6f. Because each two-photon experiment images a single area in a single mouse at a single depth, a within-session cross-area comparison is impossible; we therefore analysed one VISp experiment (`ophys_experiment_id` 501794235, 152 neurons) and one posteromedial (VISpm) experiment (503772253, 128 neurons), drawn from different specimens but matched for Cre line (Cux2-CreERT2), reporter (Ai93/GCaMP6f) and imaging depth (175 µm). Features were the per-neuron mean $\Delta F/F$ in a 0 ms to 500 ms post-stimulus response window; an L2-regularized multiclass linear decoder (`RidgeClassifierCV`, with the regularization strength chosen inside each training fold over a wide interior grid) was

evaluated with a strictly leak-free scheme combining within-fold standardization and contiguous temporal-block cross-validation with a frame embargo. In VISp, the full population decoded scene identity at 0.178 top-1 accuracy and 0.394 top-5 accuracy—roughly 21× and 46× the $1/118$ chance level (exact binomial $p \approx 0$). VISpm decoded markedly less well (0.096 top-1, 0.221 top-5). A neuron-subsampling analysis showed the VISp advantage was not a consequence of its larger population: at a matched count of 128 neurons, VISp still reached 0.159 ± 0.004 top-1 versus VISpm's 0.096, a difference of +0.064 that placed VISpm 15.5 standard deviations below the VISp subsampled distribution. These results, together with per-trial predictions and reproducible code, indicate that VISp maintains a more detailed and robust population representation of natural-scene identity than the higher visual area VISpm under matched recording conditions.

Keywords: two-photon calcium imaging; natural scenes; population decoding; mouse visual cortex; VISp; VISpm; Allen Brain Observatory; cross-validation; ridge classifier.

1 Introduction

The mammalian visual cortex confronts a formidable inference problem: from the spatiotemporal pattern of light falling on the retina it must construct stable, behaviourally useful representations of objects, surfaces and scenes. Decades of work in cats, primates and, increasingly, rodents have established that this computation is distributed across a hierarchy of cortical areas, each emphasizing different features of the visual input [1, 2]. The mouse has become a central model for dissecting this hierarchy because it combines a tractable, retinotopically organized set of visual areas with a genetic toolkit that permits chronic, cell-type-specific recording of large neuronal populations [3, 4]. A recurring and still only partially resolved question is how the representation of naturalistic visual input is reformatted as signals ascend from primary visual cortex (V1, area VISp) into the constellation of higher visual areas (HVAs) that surround it. The present report addresses one concrete facet of this question: how much information about the identity of a complex natural scene can be read out from the population activity of VISp compared with a higher visual area, the posteromedial area (VISpm), when the two are compared on equal footing.

1.1 The Allen Brain Observatory Visual Coding two-photon survey

Our analysis is built on the Allen Brain Observatory Visual Coding two-photon dataset, a large, standardized *in vivo* survey of visually evoked activity in the mouse visual cortex [5, 6]. Motivated by the argument that understanding cortical function at scale requires industrialized, quality-controlled data collection rather than bespoke individual experiments [6], the Allen Institute imaged calcium activity from tens of thousands of neurons across six visual areas, multiple cortical layers and a panel of transgenic Cre lines, all under a fixed battery of visual stimuli. The stimulus battery includes drifting and static gratings, locally sparse noise, natural movies, and—critical for the present work—a set of 118 natural-scene photographs, each presented for 250 ms and repeated 50 times in randomized order. The resulting resource has become a benchmark for testing theories of visual coding and hierarchical processing precisely because its standardization makes cross-area and cross-animal comparisons meaningful [5]. One of the survey's headline observations is that the largest functional class of neurons is not reliably driven by any single stimulus type, and that this weakly responsive fraction grows in higher visual areas, hinting at a progressive change in the nature of the visual code beyond V1 [5].

A structural feature of the two-photon dataset dictates the design of any cross-area study.

Each two-photon "experiment" (identified by an `ophys_experiment_id`) images a single field of view: one mouse, at one cortical depth, in one visual area, during one session type. It is therefore physically impossible to compare two areas *within* a single experiment; any VISp-versus-HVA contrast must draw on two different experiments, and hence two different animals and sessions. We embrace this constraint explicitly. To minimize the confounds that necessarily accompany a between-animal comparison, we selected a VISp experiment and a VISpm experiment that share the same Cre line (Cux2-CreERT2, labelling layer-2/3 and upper layer-4 excitatory neurons), the same GCaMP6f reporter (Ai93), the same nominal imaging depth (175 µm), the same session type (`three_session_B`, which contains the natural-scenes block), and the same stimulus set of 118 scenes each repeated 50 times. What remains uncontrolled—the identity of the individual mouse and the particular imaging session—is stated openly and revisited in the Discussion.

1.2 GCaMP6 calcium imaging as a window on population activity

The signals we decode are relative fluorescence changes ($\Delta F/F$) of the genetically encoded calcium indicator GCaMP6f expressed in cortical excitatory neurons [4]. GCaMP6 indicators were a step change in sensitivity: they reliably report single action potentials in favourable conditions and, in their fast ("f") variant, provide the kinetics needed to follow stimulus-locked responses at the tens-of-Hz frame rates typical of resonant two-photon microscopy [4, 7]. The physics of two-photon excitation—nonlinear absorption confined to a diffraction-limited focal volume—permits optical sectioning and micron-scale resolution hundreds of micrometres deep in scattering tissue, at the cost of a practical depth ceiling of roughly 450 µm that confines most survey imaging to the superficial layers [8, 9]. Calcium fluorescence is, however, an indirect and temporally low-pass proxy for spiking: the indicator integrates spikes over its binding and decay kinetics, so $\Delta F/F$ reflects a smoothed, nonlinearly transformed version of the underlying firing rate [4, 10]. Successive generations of indicators have improved linearity and speed [7, 11], and a substantial methodological literature addresses the inference of spike rates from fluorescence [10]. For a decoding analysis these properties matter chiefly because they bound the information available at the input: any accuracy we report is a lower bound on the information present in the spike trains, filtered through the indicator and the imaging pipeline. We therefore treat trial-averaged $\Delta F/F$ within a fixed post-stimulus window as a conservative, well-defined feature and focus on *relative* comparisons between areas measured with identical indicators and pipelines.

1.3 Functional organization of VISp and the higher visual areas

Mouse visual cortex comprises primary visual cortex (VISp) and roughly nine retinotopically defined higher visual areas that tile the cortex around it [3, 12]. Anatomical tracing shows that these areas are not a flat collection but are organized into candidate dorsal and ventral processing streams, reminiscent—though not homologous—of the primate "where" and "what" pathways [1, 13]. Physiological surveys using both calcium imaging and dense electrophysiology have assigned the areas a functional hierarchy, with VISp near its base and areas such as VISpm, AM and AL occupying higher tiers [2, 14, 15]. The areas differ systematically in their spatiotemporal tuning: relative to VISp, some lateral areas prefer higher temporal frequencies and faster motion, whereas medial areas including the posteromedial area PM (VISpm) tend to prefer high spatial frequencies, low temporal frequencies and small, slowly moving stimuli [2, 14]. Layer- and area-dependent differences in the recruitment of inhibition further shape how signals are transformed along the hierarchy [16].

The posteromedial area VISpm is a useful test case for asking how scene information changes

beyond V1. It sits above VISp in most hierarchy estimates [1, 15] and is generally assigned to the medial group of areas whose tuning emphasizes fine spatial detail at low temporal frequency [14]. Two broad hypotheses bracket the possible outcomes of a scene-decoding comparison. Under a "progressive abstraction" view, higher areas discard image-specific detail in favour of more invariant or categorical features, which would tend to reduce the linear decodability of individual scene identities in VISpm relative to VISp. Under an alternative "specialized re-encoding" view, a higher area might retain or even sharpen information about the particular stimulus classes it is specialized for, which could leave scene-identity decoding comparable across areas. Adjudicating between these requires a decoding measurement made under matched conditions, including matched population size—the central methodological commitment of this report.

1.4 Population decoding of natural scenes

Reading a stimulus identity out of neural population activity—population decoding—has become a standard tool for quantifying how much, and what kind of, information a neural representation makes explicit [17, 18]. In human neuroimaging, linear decoders can identify which natural image or movie a participant viewed from distributed cortical activity [18, 19], and in rodents large-scale calcium imaging has revealed that visual cortical population responses are high-dimensional and carry rich stimulus information [17]. The choice of decoder is consequential. Linear decoders are the workhorse of the field because a linear readout is biologically plausible—a downstream neuron computes a weighted sum of its inputs—and because linear decodability provides a conservative, interpretable measure of "explicit" information that does not credit the experimenter's model with nonlinear feats the brain may not perform [17, 20]. In the high-dimensional, few-trials-per-class regime typical of imaging (here 118 classes with only 50 repeats each), regularization is essential to control variance; ridge (L2) regularization, introduced for exactly this ill-conditioned regression setting [21], is a natural and well-behaved choice. These considerations, together with efficiency at the scale of the subsampling analysis, motivate the L2-regularized multiclass linear decoder used here (Section 2).

A second, equally important consideration is the avoidance of information leakage between training and test data. Because calcium signals are temporally autocorrelated and because any preprocessing step that "sees" the test data can inflate accuracy, naive cross-validation can produce optimistic, irreproducible results—a family of pitfalls extensively documented in systems neuroscience and neuroimaging [22–24]. We therefore adopt a deliberately conservative pipeline in which all standardization is fit inside each training fold and in which cross-validation folds are contiguous blocks of time separated from the test block by a temporal embargo, guaranteeing zero overlap between the response windows used for training and testing.

1.5 Aims and contributions

Within this framing, the report makes three contributions. First, we establish that natural-scene identity is robustly linearly decodable from a single superficial VISp population, quantifying accuracy against the $1/118$ chance level with both top-1 and top-5 metrics and characterizing how accuracy grows as neurons are added. Second, we perform a like-for-like comparison of VISp and VISpm at a matched neuron count, explicitly acknowledging that the two recordings come from different mice and sessions, and quantify the accuracy difference and its statistical reliability. Third, we release per-trial predictions and the complete, leak-free analysis code, so that every reported number is reproducible from the public Allen data. The remainder of the report details the data and methods

(Section 2), presents the decoding results (Section 3), and interprets their biological significance and limitations (Section 4).

2 Methods

2.1 Dataset, experiment selection and *as-of* date

All data were obtained from the Allen Brain Observatory Visual Coding two-photon dataset via the AllenSDK (version 2.16.2) [5, 25]. Data were downloaded and cached from the Allen Institute release, and all experiment-selection queries and downloads were finalized *as of* 2026-07-11 (UTC); the manifest and metadata records generated on that date accompany the code. Two `three_session_B` experiments—the session type that contains the natural-scenes block—were selected (Table 1). The primary visual cortex experiment was `ophys_experiment_id` 501794235 (experiment container 511507650, VISp, 152 neurons, specimen 222424), and the higher-visual-area experiment was `ophys_experiment_id` 503772253 (experiment container 511510822, VISpm, 128 neurons, specimen 225037). The two experiments were deliberately matched on every controllable acquisition variable: both used the Cux2-CreERT2 driver with the Ai93(TITL-GCaMP6f) reporter, were imaged at 175 µm depth with 910 nm excitation over a 400 × 400 µm field of view, were recorded from male mice of closely matched age (107 vs. 103 d), and presented the identical natural-scenes stimulus (118 unique scenes, 50 repeats each, plus 50 blank sweeps; 5900 image trials per experiment). **The two experiments necessarily come from different mice and different sessions**; this between-animal design is an unavoidable consequence of the single-area-per-experiment structure of the dataset and is treated as a limitation (Section 4.4).

Table 1: Selected two-photon experiments. Both are `three_session_B` sessions matched for genetics, reporter, depth and stimulus; they differ in specimen (mouse) and session, as required by the single-area-per-experiment structure of the dataset.

Property	**VISp (primary)**	**VISpm (higher)**
`ophys_experiment_id`	501794235	503772253
Experiment container ID	511507650	511510822
Targeted structure	VISp	VISpm
Specimen (mouse)	222424	225037
Cre line	Cux2-CreERT2	Cux2-CreERT2
Reporter	Ai93 (GCaMP6f)	Ai93 (GCaMP6f)
Imaging depth	175 µm	175 µm
Session type	three_session_B	three_session_B
Neurons (N)	152	128
Natural-scene trials	5900 (118×50)	5900 (118×50)

2.2 Response-window feature extraction

For each experiment we aligned the per-neuron $\Delta F/F$ traces (as provided by the AllenSDK processing pipeline) to the onset of every natural-scene presentation using the stimulus table. The imaging frame rate, computed from the acquisition timestamps, was 30.1 Hz for both experiments. The feature for a given trial and neuron was the mean $\Delta F/F$ over a fixed 0 ms to 500 ms post-stimulus response window (15 imaging frames at 30.1 Hz), yielding one scalar per neuron per trial. This produced a trial-by-neuron feature matrix of 5900 × 152 for VISp and 5900 × 128 for VISpm, with

an integer label in $\{0, \ldots, 117\}$ giving the scene shown on each trial. The 50 blank sweeps were excluded. No cells were dropped by the responsiveness filter in either experiment (0 of 152 and 0 of 128), so the full imaged populations were used. The window length of 500 ms was chosen to span the 250 ms stimulus plus the slow decay of the GCaMP6f transient, capturing the bulk of the evoked response while remaining short enough to avoid contamination from the following trial.

2.3 Leak-free cross-validation and standardization

We used a five-fold cross-validation scheme designed to eliminate temporal leakage (Figure 1). Rather than shuffling trials at random—which, given the autocorrelation of calcium signals and the block structure of the stimulus, can leak information across the train/test boundary [22, 24]—we ordered the image trials in acquisition time and partitioned them into five contiguous temporal blocks. Each fold used one block as the test set and the remaining four as the training set (a group k-fold over temporal blocks). To ensure that no training trial's response window could abut or overlap a test trial's window, we applied a 15-frame embargo: any training trial whose response window fell within 15 frames (one full window length, ≈500 ms) of the test block's temporal span was purged. This guaranteed a minimum train–test gap of 16–17 frames and zero response-window overlap, verified programmatically for every fold (0 overlaps in all 2×5 folds). Each test block contained all 118 scene classes. Folds were constructed independently for each experiment from its own acquisition timeline, because the two sessions have distinct temporal structure; the folds are therefore not shared across areas, and no cross-experiment alignment was performed or implied.

Feature standardization was performed strictly within each fold to prevent leakage of test-set statistics into preprocessing. For every fold we fit a `StandardScaler` (per-neuron z-scoring) on the *raw* mean-$\Delta F/F$ features of the training trials only, then applied that fixed transform to both the training and the test trials. No statistic computed on the test data ever influenced the model or its preprocessing.

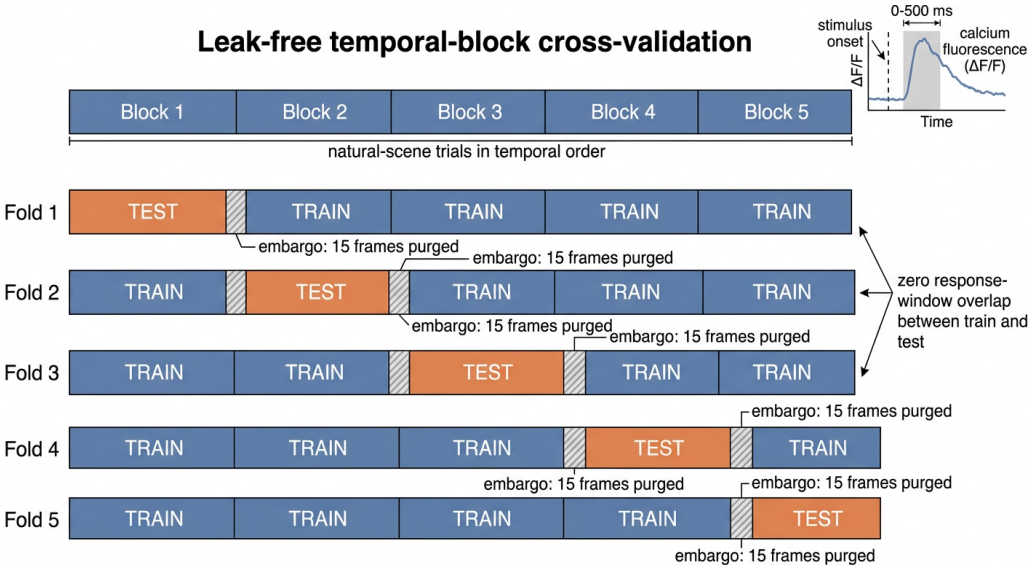


Figure 1: Leak-free temporal-block cross-validation. Natural-scene trials are ordered in acquisition time and split into five contiguous blocks. Each fold tests on one block and trains on the other four; a 15-frame embargo (hatched gaps) purges training trials adjacent to the test block, guaranteeing zero overlap between the 0 ms to 500 ms response windows used for training and testing. Within-fold z-scoring (fit on training trials only) completes the leak-free pipeline.

2.4 L2-regularized multiclass linear decoder

The decoder was an L2-regularized multiclass *linear* classifier implemented as `RidgeClassifierCV` in scikit-learn [21, 26]. This model performs ridge regression of the standardized population features onto a one-hot encoding of the 118 scene classes and assigns each test trial to the class with the highest decision score. It is a canonical population decoder for the high-class-count, modest-trials-per-class regime, and on a benchmark of this dataset it both decoded more accurately and ran roughly two orders of magnitude faster than a multinomial logistic-regression or linear-SVM decoder, which made the many-fit subsampling analysis tractable. The regularization strength α was selected *inside each training fold* by efficient leave-one-out generalized cross-validation over a wide grid, $\alpha \in \{10^{-1}, 10^{0}, \ldots, 10^{6}\}$. The grid was intentionally extended upward until the selected α was interior rather than at a boundary: the optimum was $\alpha = 10^{3}$ for VISp and $\alpha = 10^{4}$ for VISpm in every fold, confirming that the regularization was genuinely cross-validated rather than grid-capped. Because α selection uses only training trials, it introduces no leakage.

Two accuracy metrics were computed. Top-1 accuracy is the fraction of test trials whose highest-scoring class equals the true scene. Top-5 accuracy is the fraction of test trials for which the true scene is among the five highest-scoring classes. Both are reported as the mean $\pm$ standard deviation across the five folds, and pooled across all 5900 test predictions for the exact statistical tests below. For the per-trial export we additionally converted the decision scores to softmax-normalized confidence values; these are documented as confidences rather than calibrated probabilities, because a ridge classifier does not output a proper posterior.

2.5 Neuron-subsampling curve and matched-count comparison

To characterize how decoding scales with population size and to enable a fair cross-area comparison, we subsampled the VISp population. For each target count $n \in \{10, 20, 30, 50, 80, 100, 120, 128, 152\}$ we drew 10 random neuron subsets (without replacement) and, for each subset, ran the full leak-free cross-validation pipeline, recording the mean top-1 and top-5 accuracy across folds. We report the mean and standard deviation across the 10 repetitions; the $n = 152$ point is the full population (a single deterministic run). Random subset selection used a fixed seed for reproducibility.

The cross-area comparison was made at a matched neuron count of $n = 128$, the size of the smaller (VISpm) population. VISp accuracy at 128 neurons was taken from the subsampling analysis (10 random 128-neuron subsets of the 152-neuron VISp population), and VISpm accuracy was its full-population value (128 neurons). We report the difference in mean top-1 accuracy and, to express the size of the gap relative to the sampling variability of the VISp estimate, the z-score of the VISpm point estimate within the distribution of the 10 VISp 128-neuron subsamples.

2.6 Statistical analysis

Decoding accuracy was compared to the $1/118 = 0.008475$ chance level in two complementary ways. First, a one-sample t-test of the five per-fold top-1 accuracies against chance provided a conservative, fold-level test with an accompanying Cohen's d effect size and a 95% confidence interval on the mean fold accuracy. Second, because five folds provide little power, an exact binomial test on the pooled predictions (number of correct trials out of 5900) against $p_0 = 1/118$ provided a high-powered confirmation. The matched-count cross-area difference was assessed with a one-sample t-test of the 10 VISp 128-neuron subsample accuracies against the VISpm point estimate, and summarized by the z-score described above. All analyses used a fixed random seed (42) and are reproducible with

the released code.

3 Results

3.1 Natural-scene identity is robustly decodable from VISp

A single superficial VISp population of 152 neurons decoded which of the 118 natural scenes was presented on a held-out trial far above chance. Averaged over the five leak-free temporal folds, top-1 accuracy was $\mathbf{0.178 \pm 0.016}$ and top-5 accuracy was $\mathbf{0.394 \pm 0.023}$ (mean ± SD; Table 2). Against the $1/118$ chance level of 0.008 475, these correspond to roughly 21× and 46× chance, respectively. The effect was overwhelmingly significant by both tests: the fold-level one-sample t-test gave $t_4 = 23.5$ ($p = 2.0 \times 10^{-5}$, two-sided; Cohen's $d = 10.5$), and the exact binomial test on the pooled 1050/5900 correct predictions gave $p \approx 0$. In practical terms, from half a second of calcium activity in roughly 150 layer-2/3 neurons one can identify a specific photograph out of 118 on nearly one in five trials, and place it in a shortlist of five on nearly two in five trials.

The structure of the errors reinforces this picture. The VISp confusion matrix (Figure 3, left) shows a pronounced diagonal—correct predictions concentrated along the true-equals-predicted line—superimposed on a diffuse, largely unstructured background of errors, indicating that when the decoder is wrong it is not systematically confusing particular scene pairs but rather failing to reach threshold. Decodability was, however, far from uniform across scenes (Figure 4). Per-scene top-1 accuracy ranged from near zero to above 0.6: 81% of the 118 scenes were decoded above chance and 25 scenes exceeded 0.3 top-1 accuracy, while a minority of scenes were essentially indecodable from this population. This heterogeneity is expected if scenes differ in how strongly and reproducibly they drive the imaged neurons.

Table 2: Full-population decoding performance for the 118-way natural-scene task. Chance = $1/118$ = 0.008475. Values are mean ± SD across five leak-free temporal folds; the pooled binomial test uses all 5900 test predictions.

Area	N	**Top-1**	**Top-5**	**Top-1 / chance**	**Binomial** p
VISp	152	0.178 ± 0.016	0.394 ± 0.023	$\sim 21\times$	≈ 0
VISpm	128	0.096 ± 0.004	0.221 ± 0.015	$\sim 11\times$	≈ 0

3.2 Decoding accuracy grows monotonically with population size

The VISp neuron-subsampling analysis (Figure 2A) revealed a smooth, strictly monotonic increase of decoding accuracy with the number of neurons, with no sign of saturation up to the full population. Mean top-1 accuracy rose from 0.033 ± 0.005 at 10 neurons, through 0.091 ± 0.009 at 50, 0.141 ± 0.006 at 100 and 0.159 ± 0.004 at 128, to 0.178 at the full 152 neurons; top-5 accuracy rose in parallel from 0.103 to 0.394 (Table 3). The curve is concave but still climbing at 152 neurons, implying that these superficial VISp populations have not exhausted their scene-identity information at ∼150 cells and that larger populations would decode still better. The tightness of the error bars (SD ≤ 0.009 top-1 across the 10 subsamples at every count) shows that accuracy depends mainly on the number of neurons rather than on which particular neurons are chosen.

Table 3: VISp neuron-subsampling curve. Mean ± SD across 10 random neuron subsets per count (the $n = 152$ point is the full population, a single run).

n **neurons**	10	20	30	50	80	100	120	128	152
Top-1	0.033	0.053	0.063	0.091	0.126	0.141	0.155	0.159	0.178
±SD	0.005	0.007	0.004	0.009	0.009	0.006	0.004	0.004	–
Top-5	0.103	0.150	0.174	0.227	0.297	0.331	0.355	0.364	0.394
±SD	0.016	0.013	0.007	0.006	0.017	0.010	0.008	0.008	–

3.3 VISp outperforms VISpm, and not merely because it has more neurons

At full population, VISpm (128 neurons) decoded scene identity substantially less well than VISp: top-1 accuracy was $\mathbf{0.096 \pm 0.004}$ and top-5 was $\mathbf{0.221 \pm 0.015}$ (Table 2). This is still far above chance (∼11×; pooled 564/5900 correct, binomial $p \approx 0$; fold-level $t_4 = 44.9$, $p = 1.5 \times 10^{-6}$, Cohen's $d = 20.1$), so VISpm clearly carries decodable scene information—but roughly half as much, by top-1 accuracy, as VISp's full population.

Because VISp had more neurons (152 vs. 128), a fair comparison must control for population size. The matched-count analysis at 128 neurons does exactly this (Figure 2B). Subsampled to 128 neurons, VISp still achieved $\mathbf{0.159 \pm 0.004}$ top-1 accuracy versus VISpm's **0.096**—a difference of $+\mathbf{0.064}$ top-1 (and +0.143 top-5) in favour of VISp. The gap is large relative to the sampling variability of the VISp estimate: the VISpm value lies **15.5 standard deviations below** the mean of the VISp 128-neuron subsample distribution, and a one-sample t-test of the 10 VISp subsamples against the VISpm point estimate was unambiguous ($t_9 = 48.9$, $p = 3.1 \times 10^{-12}$). Equivalently, reading off the subsampling curve, VISpm's full 128-neuron accuracy (0.096) is matched by a VISp subpopulation of only about 50–55 neurons—less than half as many cells. The VISp advantage in scene decoding is therefore not an artefact of its larger population; at matched neuron count VISp remains markedly more informative about which natural scene was presented.

The confusion matrices make the difference visible at the level of individual scenes (Figure 3): VISpm's diagonal is visibly fainter and sparser than VISp's, with fewer scenes reaching high per-scene accuracy (13 scenes above 0.3 top-1 versus 25 for VISp; 58% versus 81% of scenes above chance; Figure 4). Both areas show the same qualitative error structure—a diagonal over diffuse noise—but VISp's representation supports correct read-out for roughly twice as many scenes.

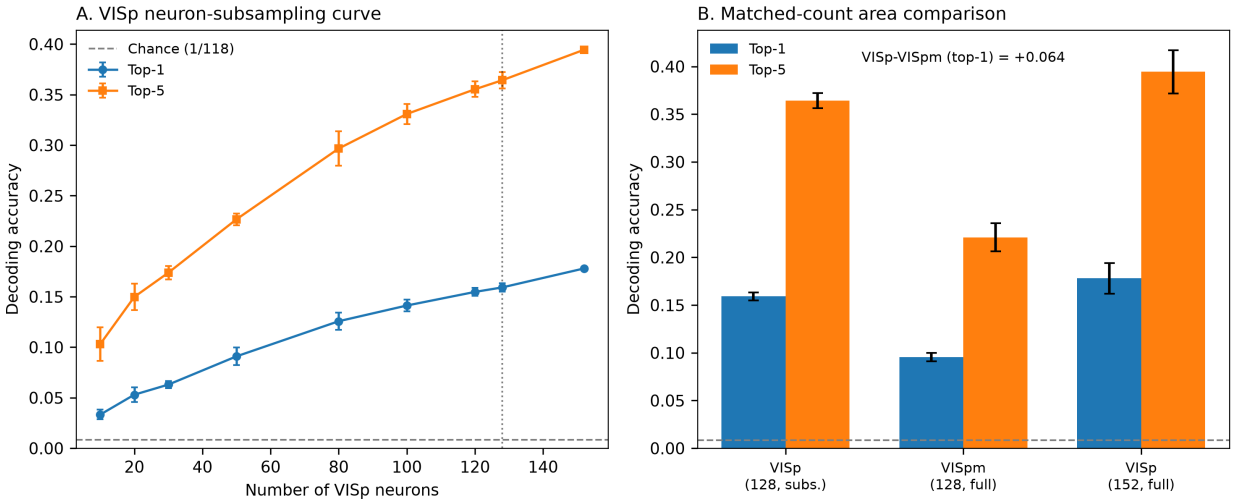


Figure 2: Decoding performance summary. (A) VISp neuron-subsampling curve: mean top-1 (blue) and top-5 (orange) decoding accuracy versus the number of subsampled VISp neurons (mean ± SD over 10 random subsets per count). The dashed line is the $1/118$ chance level; the dotted vertical line marks the 128-neuron matched count. Accuracy rises smoothly and monotonically with no saturation. **(B)** Matched-count area comparison: top-1 and top-5 accuracy for VISp subsampled to 128 neurons, VISpm at its full 128 neurons, and VISp at its full 152 neurons. At the matched count VISp exceeds VISpm by +0.064 top-1.

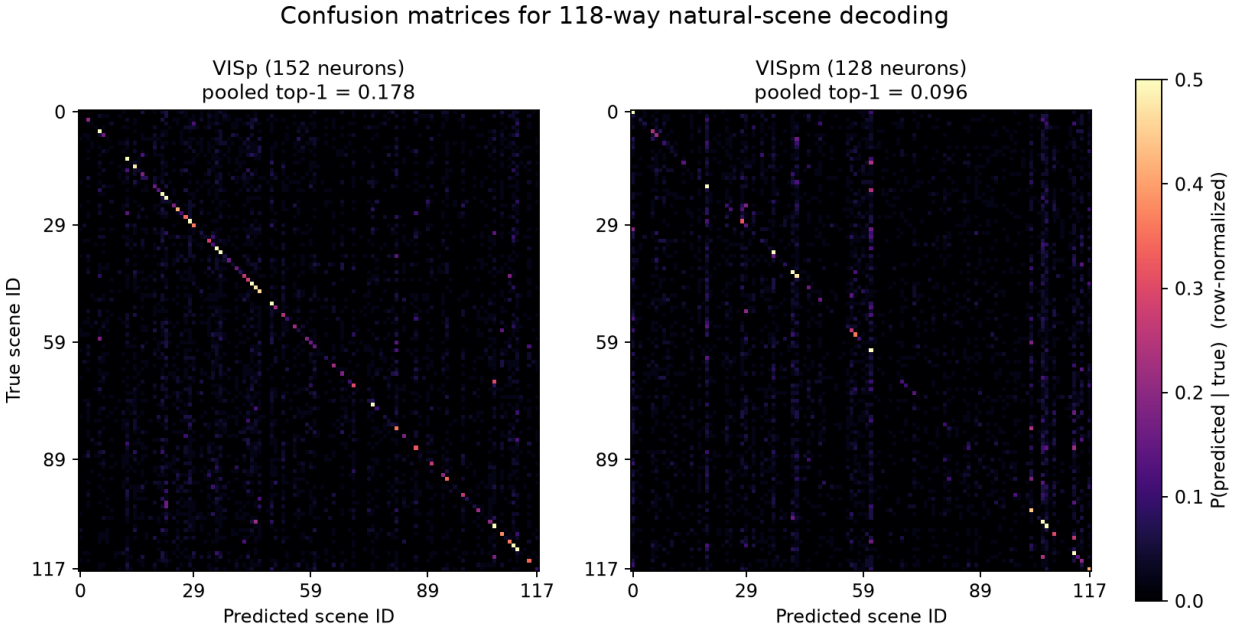


Figure 3: Confusion matrices for 118-way natural-scene decoding (row-normalized, pooled over all five test folds). Rows are true scenes, columns predicted scenes; brightness encodes $P(\text{predicted} \mid \text{true})$. VISp (left) shows a stronger, denser diagonal than VISpm (right), consistent with its higher top-1 accuracy. Off-diagonal errors are diffuse in both areas, indicating no systematic confusion between particular scene pairs.

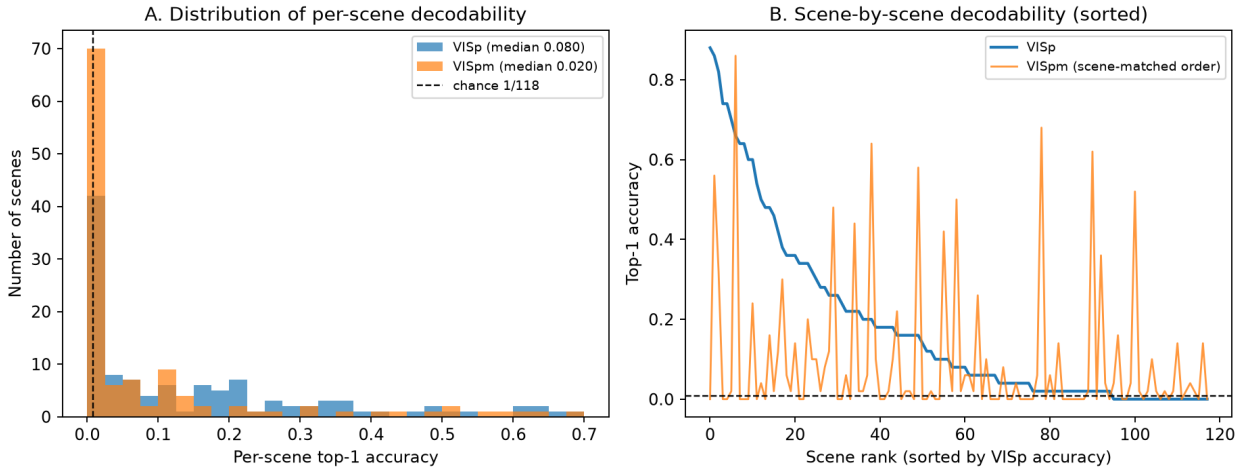


Figure 4: Scene-by-scene decodability. (A) Distribution of per-scene top-1 accuracy for VISp (blue) and VISpm (orange); dashed line marks chance. A large fraction of scenes cluster near chance in both areas, but VISp has a heavier right tail of well-decoded scenes. **(B)** Per-scene accuracy sorted by VISp decodability (blue), with VISpm shown in the same scene order (orange). VISp decodes more scenes, and decodes its best scenes more accurately, than VISpm.

3.4 Per-trial predictions and reproducibility

For full transparency the analysis exports per-trial predictions for both areas (`predictions_visp.csv` and `predictions_vispm.csv`), each with 5900 rows—one per image trial across all five held-out test folds. Each row records the fold, the trial index, the true scene, the predicted (top-1) scene, binary top-1 and top-5 correctness flags, the top-5 predicted scene IDs, and the softmax-normalized decoder confidence for the predicted and true classes. The complete decoding pipeline (feature loading, within-fold standardization, ridge-classifier fitting with interior α selection, subsampling and statistics) and the figure-generation code are released alongside the machine- readable results (`decoding_results.json`), so that every quantity in this report can be regenerated from the public Allen data.

4 Discussion

4.1 Summary of findings

Using a deliberately leak-free decoding pipeline on the Allen Brain Observatory two-photon dataset, we found that the identity of a natural scene, drawn from a set of 118, is robustly linearly decodable from a single superficial VISp population: 17.8% top-1 and 39.4% top-5 accuracy from 152 neurons, more than an order of magnitude above chance. Decoding accuracy grew smoothly and without saturation as neurons were added, indicating that even ∼150 layer-2/3 cells do not exhaust the scene information available in VISp. Critically, when VISp and the higher visual area VISpm were compared at a matched population of 128 neurons—the fair comparison demanded by the fact that the two areas were imaged in different mice—VISp retained a substantial and highly reliable advantage (0.159 vs. 0.096 top-1, +0.064; VISpm 15.5 SD below the VISp subsample distribution). The VISp advantage is therefore a property of the representation, not of the sample size.

4.2 Biological interpretation

The most parsimonious interpretation is that primary visual cortex maintains a more detailed, higher-fidelity population representation of individual natural-scene identities than the posteromedial higher visual area VISpm, at least as read out by a linear decoder from superficial excitatory neurons. This is consistent with the broader picture emerging from the Allen surveys and from anatomy: VISp sits at the base of the visual hierarchy and receives the most direct, retinotopically precise thalamic input, whereas higher areas such as VISpm occupy higher hierarchical tiers and are specialized for particular spatiotemporal feature regimes rather than for veridical image representation [1, 14, 15]. Our result aligns naturally with the "progressive abstraction" hypothesis outlined in the Introduction: as signals ascend beyond V1, the linear decodability of specific image identities falls, plausibly because higher areas re-encode the input in terms of behaviourally relevant features (motion, spatial scale, ethological categories) that discard some of the image-specific detail needed to distinguish one photograph from another. It also resonates with the Allen survey's observation that the weakly stimulus-driven fraction of neurons grows in higher areas [5]: if fewer VISpm neurons respond reliably to any given scene, fewer informative dimensions are available to a decoder, which is exactly the signature we observe in the fainter VISpm confusion diagonal and the smaller fraction of well-decoded scenes.

VISpm's tuning profile offers a complementary account. Medial higher areas, including PM, are biased toward high spatial frequencies and low temporal frequencies and toward small, slow stimuli [2, 14]. A diverse set of natural photographs spans a wide range of spatial-frequency content and does not selectively engage this narrow preferred regime, so a VISpm population may respond informatively to only the subset of scenes that match its tuning—consistent with the sparser set of well-decoded scenes we find in VISpm. Under this view, the cross-area gap reflects not a global loss of information but a narrowing of the stimulus space over which the area carries explicit information. Importantly, our finding that VISpm still decodes at $\sim 11\times$ chance shows that scene information is far from abolished beyond V1; it is attenuated, and re-formatted, rather than erased.

We emphasize what our measure does and does not establish. Linear decodability is a measure of *explicit*, linearly accessible information—the kind a downstream neuron could read with a weighted sum [17, 20]. A lower linear-decoding accuracy in VISpm does not imply that VISpm contains less information in an absolute (e.g. nonlinear or information-theoretic) sense; it could in principle hold comparable information in a format that a linear read-out cannot access. What the result does establish is that, for the biologically plausible and widely used class of linear read-outs, VISp makes scene identity more explicit than VISpm under matched conditions.

4.3 Relation to prior work

Our numbers sit comfortably within the literature on visual population decoding. That linear decoders can identify specific natural images or movies from distributed visual activity is well established in human neuroimaging [18, 19], and large-scale mouse imaging has shown that visual cortical populations carry high-dimensional, richly informative stimulus representations [17]. The monotonic, non-saturating growth of accuracy with neuron count that we observe is the expected behaviour when additional neurons contribute partly independent information, and mirrors the high embedding dimensionality reported for mouse V1 responses to large image sets [17]. Our methodological posture—within-fold standardization and temporally embargoed block cross-validation—follows explicit warnings in the systems-neuroscience and neuroimaging methods literature about circular analysis and leakage-inflated accuracy [22–24], and we regard the resulting numbers as conservative

but trustworthy.

4.4 Limitations

Several limitations temper the interpretation. First and most importantly, the comparison is between two experiments and therefore between two mice and two sessions: the single-area-per-experiment structure of the two-photon dataset makes a within-animal cross-area contrast impossible. We mitigated this by matching Cre line, reporter, imaging depth, session type, age and stimulus, but we cannot exclude that some of the VISp–VISpm difference reflects animal-specific factors (expression level, behavioural state, imaging quality) rather than area identity *per se*. A definitive test would aggregate many experiments per area, or use a modality (e.g. Neuropixels) that records multiple areas simultaneously [15]; the present two-experiment design should be read as a rigorous case study rather than a population-level estimate of the area effect. Second, the signal is GCaMP6f fluorescence, a slow, nonlinear proxy for spiking [4, 10]; our accuracies are lower bounds on the information in the underlying spike trains, and indicator or expression differences between the two mice are a possible (though depth- and line-matched) confound. Third, the analysis is restricted to superficial (layer-2/3) excitatory neurons at a single depth, so it does not speak to laminar differences or to inhibitory populations [16]. Fourth, we used trial-averaged mean $\Delta F/F$ in a single fixed window and a purely linear decoder; a temporally resolved feature, a peak-aligned window, or a nonlinear decoder might recover more information and could in principle narrow or widen the cross-area gap. Finally, the natural-scene set, while diverse, is a particular sample of 118 photographs, and the per-scene heterogeneity we observe cautions against over-generalizing to all natural images.

4.5 Future directions

Several extensions follow directly. Scaling the comparison to the full Allen cohort—many VISp and VISpm experiments, and additional higher areas such as VISl, VISal and VISam—would convert this case study into a statistically powered map of how natural-scene decodability varies across the mouse visual hierarchy, with per-area variability properly estimated. Simultaneous multi-area electrophysiology would remove the between-animal confound entirely [15]. On the analysis side, a peak-aligned or multi-frame response feature, spike-rate inference before decoding [10], and comparison of linear against nonlinear or deep-network decoders [27] would clarify how much of the cross-area gap is intrinsic to the representation versus imposed by our conservative feature and read-out choices. Relating per-scene decodability to image statistics—spatial-frequency content, contrast, semantic category—could test the specific prediction that VISpm preferentially represents the subset of scenes matching its spatiotemporal tuning [2, 28, 29]. Together these directions would sharpen the central conclusion of this report: that under matched conditions, primary visual cortex represents the identity of a natural scene more explicitly and more robustly than the higher visual area VISpm.

Data and code availability

The primary data are openly available from the Allen Brain Observatory Visual Coding two-photon dataset via the AllenSDK [5, 25] (experiments 501794235 and 503772253; accessed as of 2026-07-11). The decoding and figure-generation code, the machine-readable results (`decoding_results.json`),

and the per-trial prediction tables (`predictions_visp.csv`, `predictions_vispm.csv`) accompany this report.

Author contributions and acknowledgements

The analysis used data generated by the Allen Institute for Brain Science; we gratefully acknowledge the Allen Institute founder, Paul G. Allen, and the Allen Brain Observatory team for creating and openly sharing this resource.

References

[1] Lindsey L. Glickfeld and Shawn R. Olsen. Higher-order areas of the mouse visual cortex. *Annual Review of Vision Science*, 3(1):251–273, 2017. ISSN 2374-4650. doi: 10.1146/annurev-vision-102016-061331. URL `http://dx.doi.org/10.1146/annurev-vision-102016-061331`.

[2] James H. Marshel, Marina E. Garrett, Ian Nauhaus, and Edward M. Callaway. Functional specialization of seven mouse visual cortical areas. *Neuron*, 72(6):1040–1054, December 2011. ISSN 0896-6273. doi: 10.1016/j.neuron.2011.12.004. URL `http://dx.doi.org/10.1016/j.neuron.2011.12.004`.

[3] Quanxin Wang and Andreas Burkhalter. Area map of mouse visual cortex. *Journal of Comparative Neurology*, 502(3):339–357, March 2007. ISSN 1096-9861. doi: 10.1002/cne.21286. URL `http://dx.doi.org/10.1002/cne.21286`.

[4] Tsai-Wen Chen, Trevor J. Wardill, Yi Sun, Stefan R. Pulver, Sabine L. Renninger, Amy Baohan, Eric R. Schreiter, Rex A. Kerr, Michael B. Orger, Vivek Jayaraman, Loren L. Looger, Karel Svoboda, and Douglas S. Kim. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300, 2013. ISSN 1476-4687. doi: 10.1038/nature12354. URL `http://dx.doi.org/10.1038/nature12354`.

[5] Saskia E. J. de Vries, Jerome A. Lecoq, Michael A. Buice, Peter A. Groblewski, Gabriel K. Ocker, Michael Oliver, David Feng, Nicholas Cain, Peter Ledochowitsch, Daniel Millman, Kate Roll, Marina Garrett, Tom Keenan, Leonard Kuan, Stefan Mihalas, Shawn Olsen, Carol Thompson, Wayne Wakeman, Jack Waters, Derric Williams, Chris Barber, Nathan Berbesque, Brandon Blanchard, Nicholas Bowles, Shiella D. Caldejon, Linzy Casal, Andrew Cho, Sissy Cross, Chinh Dang, Tim Dolbeare, Melise Edwards, John Galbraith, Nathalie Gaudreault, Terri L. Gilbert, Fiona Griffin, Perry Hargrave, Robert Howard, Lawrence Huang, Sean Jewell, Nika Keller, Ulf Knoblich, Josh D. Larkin, Rachael Larsen, Chris Lau, Eric Lee, Felix Lee, Arielle Leon, Lu Li, Fuhui Long, Jennifer Luviano, Kyla Mace, Thuyanh Nguyen, Jed Perkins, Miranda Robertson, Sam Seid, Eric Shea-Brown, Jianghong Shi, Nathan Sjoquist, Cliff Slaughterbeck, David Sullivan, Ryan Valenza, Casey White, Ali Williford, Daniela M. Witten, Jun Zhuang, Hongkui Zeng, Colin Farrell, Lydia Ng, Amy Bernard, John W. Phillips, R. Clay Reid, and Christof Koch. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, 23(1):138–151, January 2020. ISSN 1546-1726. doi: 10.1038/s41593-019-0550-9. URL `http://dx.doi.org/10.1038/s41593-019-0550-9`.

[6] Michael Hawrylycz, Costas Anastassiou, Anton Arkhipov, Jim Berg, Michael Buice, Nicholas Cain, Nathan W. Gouwens, Sergey Gratiy, Ramakrishnan Iyer, Jung Hoon Lee, Stefan Mihalas, Catalin Mitelut, Shawn Olsen, R. Clay Reid, Corinne Teeter, Saskia de Vries, Jack Waters, Hongkui Zeng, Christof Koch, Costas Anastassiou, Anton Arkhipov, Chris Barber, Lynne Becker, Jim Berg, Barg Berg, Amy Bernard, Darren Bertagnolli, Kris Bickley, Adam Bleckert, Nicole Blesie, Agnes Bodor, Phil Bohn, Nick Bowles, Krissy Brouner, Michael Buice, Dan Bumbarger, Nicholas Cain, Shiella Caldejon, Linzy Casal, Tamara Casper, Ali Cetin, Mike Chapin, Soumya Chatterjee, Adrian Cheng, Nuno da Costa, Sissy Cross, Christine Cuhaciyan, Tanya Daigle, Chinh Dang, Bethanny Danskin, Tsega Desta, Saskia de Vries, Nick Dee, Daniel Denman, Tim Dolbeare, Ashley Donimirski, Nadia Dotson, Severine Durand, Colin Farrell, David Feng, Michael Fisher, Tim Fliss, Aleena Garner, Marina Garrett, Marisa Garwood, Nathalie Gaudreault, Terri Gilbert, Harminder Gill, Olga Gliko, Keith Godfrey, Jeff Goldy, Nathan Gouwens, Sergey Gratiy, Lucas Gray, Fiona Griffin, Peter Groblewski, Hong Gu, Guangyu Gu, Caroline Habel, Kristen Hadley, Zeb Haradon, James Harrington, Julie Harris, Michael Hawrylycz, Alex Henry, Nika Hejazinia, Chris Hill, Dijon Hill, Karla Hirokawa, Anh Ho, Robert Howard, Jaclyn Huffman, Ram Iyer, Tim Jarsky, Justin Johal, Tom Keenan, Sean Kim, Ulf Knoblich, Christof Koch, Ali Kriedberg, Leonard Kuan, Florence Lai, Rachael Larsen, Rylan Larsen, Chris Lau, Peter Ledochowitsch, Brian Lee, Chang-Kyu Lee, Jung-Hoon Lee, Felix Lee, Lu Li, Yang Li, Rui Liu, Xiaoxiao Liu, Brian Long, Fuhui Long, Jennifer Luviano, Linda Madisen, Veronica Maldonado, Rusty Mann, Naveed Mastan, Jose Melchor, Vilas Menon, Stefan Mihalas. Maya Mills, Catalin Mitelut, Kenji Mizuseki, Marty Mortrud, Lydia Ng, Thuc Nguyen, Julie Nyhus, Seung Wook Oh, Aaron Oldre, Doug Ollerenshaw, Shawn Olsen, Natalia Orlova, Ben Ouellette, Sheana Parry, Julie Pendergraft, Hanchuan Peng, Jed Perkins, John Phillips, Lydia Potekhina, Melissa Reading, Clay Reid, Brandon Rogers, Kate Roll, David Rosen, Peter Saggau, David Sandman, Eric Shea-Brown, Adam Shai, Shu Shi, Josh Siegle, Nathan Sjoquist, Kimberly Smith, Andrew Sodt, Gilberto Soler-Llavina, Staci Sorensen, Michelle Stoecklin, Susan Sunkin, Aaron Szafer, Bosiljka Tasic, Naz Taskin, Corinne Teeter, Nivretta Thatra, Carol Thompson, Michael Tieu, Dmitri Tsyboulski, Matt Valley, Wayne Wakeman, Quanxin Wang, Jack Waters, Casey White, Jennifer Whitesell, Derric Williams, Natalie Wong, Von Wright, Jun Zhuang, Zizhen Yao, Rob Young, Brian Youngstrom, Hongkui Zeng, and Zhi Zhou. Inferring cortical function in the mouse visual system through large-scale systems neuroscience. *Proceedings of the National Academy of Sciences*, 113(27):7337–7344, 2016. ISSN 1091-6490. doi: 10.1073/pnas.1512901113. URL `http://dx.doi.org/10.1073/pnas.1512901113`.

[7] Hod Dana, Yi Sun, Boaz Mohar, Brad K. Hulse, Aaron M. Kerlin, Jeremy P. Hasseman, Getahun Tsegaye, Arthur Tsang, Allan Wong, Ronak Patel, John J. Macklin, Yang Chen, Arthur Konnerth, Vivek Jayaraman, Loren L. Looger, Eric R. Schreiter, Karel Svoboda, and Douglas S. Kim. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nature Methods*, 16(7):649–657, 2019. ISSN 1548-7105. doi: 10.1038/s41592-019-0435-6. URL `http://dx.doi.org/10.1038/s41592-019-0435-6`.

[8] Karel Svoboda and Ryohei Yasuda. Principles of two-photon excitation microscopy and its applications to neuroscience. *Neuron*, 50(6):823–839, 2006. ISSN 0896-6273. doi: 10.1016/j.neuron.2006.05.019. URL `http://dx.doi.org/10.1016/j.neuron.2006.05.019`.

[9] Kevin Takasaki, Reza Abbasi-Asl, and Jack Waters. Superficial bound of the depth limit of two-photon imaging in mouse brain. *eneuro*, 7(1):ENEURO.0255–19.2019, January 2020. ISSN 2373-2822. doi: 10.1523/eneuro.0255-19.2019. URL `http://dx.doi.org/10.1523/ENEURO.0255-19.2019`.

[10] Philipp Berens, Jeremy Freeman, Thomas Deneux, Nikolay Chenkov, Thomas McColgan, Artur Speiser, Jakob H. Macke, Srinivas C. Turaga, Patrick Mineault, Peter Rupprecht, Stephan Gerhard, Rainer W. Friedrich, Johannes Friedrich, Liam Paninski, Marius Pachitariu, Kenneth D. Harris, Ben Bolte, Timothy A. Machado, Dario Ringach, Jasmine Stone, Luke E. Rogerson, Nicolas J. Sofroniew, Jacob Reimer, Emmanouil Froudarakis, Thomas Euler, Miroslav Román Rosón, Lucas Theis, Andreas S. Tolias, and Matthias Bethge. Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. *PLOS Computational Biology*, 14(5):e1006157, May 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006157. URL `http://dx.doi.org/10.1371/journal.pcbi.1006157`.

[11] Yan Zhang, Márton Rózsa, Yajie Liang, Daniel Bushey, Ziqiang Wei, Jihong Zheng, Daniel Reep, Gerard Joey Broussard, Arthur Tsang, Getahun Tsegaye, Sujatha Narayan, Christopher J. Obara, Jing-Xuan Lim, Ronak Patel, Rongwei Zhang, Misha B. Ahrens, Glenn C. Turner, Samuel S.-H. Wang, Wyatt L. Korff, Eric R. Schreiter, Karel Svoboda, Jeremy P. Hasseman, Ilya Kolb, and Loren L. Looger. Fast and sensitive gcamp calcium indicators for imaging neural populations. *Nature*, 615(7954):884–891, March 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-05828-9. URL `http://dx.doi.org/10.1038/s41586-023-05828-9`.

[12] Marina E. Garrett, Ian Nauhaus, James H. Marshel, and Edward M. Callaway. Topography and areal organization of mouse visual cortex. *The Journal of Neuroscience*, 34(37):12587–12600, 2014. ISSN 1529-2401. doi: 10.1523/jneurosci.1124-14.2014. URL `http://dx.doi.org/10.1523/JNEUROSCI.1124-14.2014`.

[13] Quanxin Wang, Olaf Sporns, and Andreas Burkhalter. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *The Journal of Neuroscience*, 32(13):4386–4399, March 2012. ISSN 1529-2401. doi: 10.1523/jneurosci.6063-11.2012. URL `http://dx.doi.org/10.1523/JNEUROSCI.6063-11.2012`.

[14] Mark L. Andermann, Aaron M. Kerlin, Demetris K. Roumis, Lindsey L. Glickfeld, and R. Clay Reid. Functional specialization of mouse higher visual cortical areas. *Neuron*, 72 (6):1025–1039, December 2011. ISSN 0896-6273. doi: 10.1016/j.neuron.2011.11.013. URL `http://dx.doi.org/10.1016/j.neuron.2011.11.013`.

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

reveals functional hierarchy. *Nature*, 592(7852):86–92, January 2021. ISSN 1476-4687. doi: 10.1038/s41586-020-03171-x. URL http://dx.doi.org/10.1038/s41586-020-03171-x.

[16] Rinaldo David D'Souza, Andrew Max Meier, Pawan Bista, Quanxin Wang, and Andreas Burkhalter. Recruitment of inhibition and excitation across mouse visual cortex depends on the hierarchy of interconnecting areas. *eLife*, 5, 2016. ISSN 2050-084X. doi: 10.7554/elife.19332. URL http://dx.doi.org/10.7554/eLife.19332.

[17] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D. Harris. High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765): 361–365, 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1346-5. URL http://dx.doi.org/10.1038/s41586-019-1346-5.

[18] Kendrick N. Kay, Thomas Naselaris, Ryan J. Prenger, and Jack L. Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, March 2008. ISSN 1476-4687. doi: 10.1038/nature06713. URL http://dx.doi.org/10.1038/nature06713.

[19] Shinji Nishimoto, An T. Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L. Gallant. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, 21(19):1641–1646, October 2011. ISSN 0960-9822. doi: 10.1016/j.cub.2011.08.031. URL http://dx.doi.org/10.1016/j.cub.2011.08.031.

[20] Sophie Denève and Christian K Machens. Efficient codes and balanced networks. *Nature Neuroscience*, 19(3):375–382, February 2016. ISSN 1546-1726. doi: 10.1038/nn.4243. URL http://dx.doi.org/10.1038/nn.4243.

[21] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, February 1970. ISSN 1537-2723. doi: 10.1080/00401706.1970.10488634. URL http://dx.doi.org/10.1080/00401706.1970.10488634.

[22] Nikolaus Kriegeskorte, W Kyle Simmons, Patrick S F Bellgowan, and Chris I Baker. Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, 12(5): 535–540, April 2009. ISSN 1546-1726. doi: 10.1038/nn.2303. URL http://dx.doi.org/10.1038/nn.2303.

[23] Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8, 2014. ISSN 1662-5196. doi: 10.3389/fninf.2014.00014. URL http://dx.doi.org/10.3389/fninf.2014.00014.

[24] Gael Varoquaux and Bertrand Thirion. How machine learning is shaping cognitive neuroimaging. *GigaScience*, 3(1), November 2014. ISSN 2047-217X. doi: 10.1186/2047-217x-3-28. URL http://dx.doi.org/10.1186/2047-217X-3-28.

[25] Allen Institute for Brain Science. Allen SDK: A Python library for accessing and analyzing data from the Allen Institute. https://github.com/AllenInstitute/AllenSDK, 2024. Software, version 2.16.2; Allen Brain Observatory Visual Coding two-photon dataset.

[26] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[27] Daniel L K Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):356–365, February 2016. ISSN 1546-1726. doi: 10.1038/nn.4244. URL http://dx.doi.org/10.1038/nn.4244.

[28] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1):1193–1216, March 2001. ISSN 1545-4126. doi: 10.1146/annurev.neuro.24.1.1193. URL http://dx.doi.org/10.1146/annurev.neuro.24.1.1193.

[29] Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996. ISSN 1476-4687. doi: 10.1038/381607a0. URL http://dx.doi.org/10.1038/381607a0.