Robust representation of natural images by sparse and variable population of active neurons in visual cortex

Takashi Yoshida[1,2,3,*] and Kenichi Ohki[1,2,3,4,*]

1. Department of Physiology, The University of Tokyo School of Medicine, Tokyo, Japan.

2. Department of Molecular Physiology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

3. CREST, Japan Science and Technology Agency, Tokyo, Japan

4. International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

*Correspondence: Kenichi Ohki, kohki@m.u-tokyo.ac.jp

Takashi Yoshida, takashiy@m.u-tokyo.ac.jp

Abstract

Natural scenes sparsely activate neurons in the primary visual cortex (V1). However, whether and how sparsely active neurons sufficiently and robustly represent natural image contents has not been revealed. We reconstructed the natural images from neuronal activities of mouse V1. Single natural images were linearly decodable from surprisingly small number (~20) of highly responsive neurons. This was achieved by diverse receptive fields (RFs) of the small number of responsive neurons. Furthermore, these neurons robustly represented the image against trial-to-trial response variability. Synchronous neurons with partially overlapping RFs formed functional clusters and were active at the same trials. Importantly, multiple clusters represented similar patterns of local images but were active at different trials. Thus, integration of activities among the clusters led to robust representation against the variability. Our results suggest that the diverse, partially overlapping RFs ensure the sparse and robust representation, and propose a new representation scheme in which information is reliably represented, while representing neuronal patterns change across trials.

Introduction

Sensory information is thought to be represented by relatively small number of active neurons in the sensory cortex. This sparse representation has been observed in several cortical areas[1-9] and is postulated to reflect an efficient coding of the statistical features in sensory inputs[4, 10]. However, it has not been determined whether and how small numbers of active neurons represent sufficient information about sensory inputs.

In the primary visual cortex (V1), a type of neuron termed a simple cell has a receptive field (RF) structure that is spatially localized, oriented, and has a bandpass filter property of specific spatial frequency. This RF structure is modelled by a two-dimensional (2D) Gabor function[11]. According to theoretical studies, single natural images are represented by relatively small numbers of neurons using Gabor-like RFs, whereas information about multiple natural scenes is distributed across the neuronal population[10,12,13]. Indeed, V1 neurons respond sparsely to natural scenes at the single cell level[2, 3, 5-9] and population level[3,5,14]. Population activity with higher sparseness exhibits greater discriminability between natural scenes[5].

What types of information from natural scenes are represented in sparsely active neuronal populations in a brain? The visual contents of natural scenes or movies are reconstructed from populations of single unit activities in the lateral geniculate nucleus (LGN) collected from several experiments[15] and functional magnetic resonance imaging (fMRI) data from the visual cortices[16-19]. However, it has not been addressed experimentally whether the visual contents of natural images are represented by small numbers of sparsely active neurons and whether V1 RFs in the brain are useful to represent the natural image. Furthermore, do the sparsely active neurons reliably represent the natural image contents against trial-to-trial response variability? Although a computational model[20] has suggested that sparse and overcomplete representation is optimal representation for natural images with unreliable neurons, this has not been examined experimentally.

We also addressed how visual information is distributed among neurons in a local population. It has been reported that subsets of neurons are ‘unresponsive’ to visual stimuli (e.g., a responsive rate for visual stimuli in mouse V1 of 26–68%)[21-27], indicating subsets of neurons represent sensory information. However, this may

partly because stimulus properties do not completely cover RF properties of all neurons. Thus, there are two extreme possibilities; sparsely active neurons distributed among all neurons in a local population, or only a specific subset of cells processes the natural images. What proportion of neurons are actually involved in information processing has been debated[28, 29].

Here, we examined whether and how a small number of highly responsive V1 neurons was sufficient for the representation of natural image contents. Using two-photon Ca^{2+} imaging, we recorded visual responses to natural images from local populations of single neurons in V1 of anaesthetized mice. A small number of neurons (<3%) highly responded to each natural image, which was sparser than that predicted by linear encoding model. On the other hand, approximately 90% of neurons were activated by at least one of the natural images, revealing that most neurons in a local population are involved in natural image processing. We reconstructed the natural images from the activities to estimate the information about the visual contents. Visual contents of single natural images were linearly decodable from a small number (~20) of highly responsive neurons. The highly responsive neurons showed diverse RF, which helped small numbers of neurons represent complex natural images. Furthermore, the highly responsive neurons robustly represent the image against trial-to-trial response variability. We found that subsets of the neurons whose RFs partially overlapped formed functional clusters based on correlated activities. Importantly, between the clusters, represented local images were similar to each other, while across-trial response variabilities were almost independent. Thus, integration of activities among the clusters led to a robust representation. We also found that the responsive neurons were only slightly shared between images, and many natural images were represented by the combinations of responsive neurons in a population. Finally, visual features represented by a local population were sufficient to represent the features in all the natural images we used. These results revealed new robust representation of natural images by a small number of neurons in which information is reliably represented, while representing neuronal patterns change across trials. Preliminary results of this study have been published in an abstract form[30] and on a preprint server[31].

Results

The main purpose of this study is to examine whether and how the natural images are represented in the sparse representation scheme. We first confirm the sparse response to natural images in our dataset. Next, we demonstrate that the natural images were reconstructed from a relatively small number of responsive neurons. Finally, we address how the small number of neurons robustly represent natural images against trial-to-trial response variability.

Sparse visual responses to natural images in mouse V1

We presented flashes of natural images as visual stimuli (Fig. 1a, see Methods) and simultaneously recorded the activities of several hundreds of single neurons from layers 2/3 and 4 of mouse V1 using two-photon calcium (Ca^{2+}) imaging (560 [284–712] cells/plane, median [25–75th percentiles], n = 24 planes from 14 mice, 260–450 microns in depth, Fig. 1b for representative response traces). Fig. 1c presents plots of significant visual response events for all images (x-axis) across all neurons (y-axis) in a plane (n = 726 cells, depth: 360 microns from the brain surface). Significant response for each image was defined as an evoked response which was significantly different from 0 ($p < 0.01$ using the signed-rank test) and whose normalized response amplitude (z-score) was greater than 1 (see Methods). Hereafter, we call these significant visual responses highly responsive. A few to 10% of neurons were highly responsive to a single image (5.1% [3.9–6.7%] cells/image, Fig. 1c bottom panel), indicating sparse visual responses to natural images. In contrast, nearly all neurons (98%, 711/726 cells) responded to at least one image (each cell responded 4.5% [2.5–7.5%] images, Fig. 1c right panel). Across planes, 2.7% cells were activated by a single image ([1.8–3.2%], Fig. 1f), whereas almost all cells responded to at least one image (90% [86–93%], Fig. 1g). This low responsive rate to each image was not due to poor recording conditions. The same neurons responded well to moving gratings (27% [22–34%] for one direction, and 75% [66–79%] for at least one of 8 directions, Fig. 1h and i).

The highly responsive neurons only slightly overlapped between images. Fig. 1d presents representative activity patterns for three natural images (Fig. 1d, left column). Each image activated different subsets of

neurons that exhibited small overlaps between images (Fig. 1d, right column). Of the responsive cells, 4.8% exhibited overlap between two images (25–75th percentiles for 24 planes: 4.0–5.5%, Fig. 1j). We further computed the distributions of the response amplitudes to single images (Fig. 1e). Only a small number of neurons exhibited visual responses with greater amplitudes, which is a characteristic property of a sparse representation (Fig. 1e). Population sparseness[2, 3], a measure of sparse representation, was comparable to a previous report for mouse V1[5] (0.36 [0.30–0.42], Fig. 1k, see Methods). Thus, each natural image activated a relatively small number of neurons, whereas most neurons in a local population were activated by at least one of the images, suggesting the sparsely distributed representation of natural images in V1 that was originally proposed in a previous study[10]. The latter result also provides the first report that most neurons in mouse V1 are visually responsive to natural image stimuli[28, 29].

Partially overlapping representations of visual features among local V1 populations

We created encoding models for the visual responses of individual neurons to examine the visual features represented by each neuron. We used a set of Gabor wavelet filters (1248 filters, Supplementary Fig. 1a and b, see Methods) to extract the visual features from the natural images. Natural images were applied to each Gabor filter and transformed into sets of feature values (Gabor feature values). For each neuron, we first selected the Gabor features that exhibited strong correlations with the visual response. The correlation threshold for the selected feature was adjusted to maximize the visual response prediction (Supplementary Fig. 1c–e, see Methods). Then, the visual response was represented by a linear regression of the selected feature values followed by a non-linear scaling (Fig. 2a, see Methods). The visual response prediction of the model was estimated with a different dataset from the dataset used in the regression (10-fold cross validation, see Methods).

Visual response of an individual neuron was represented by a small number of Gabor features. In the example cells (Fig. 2b and c), the correlation coefficients between the observed responses and the responses predicted by the model were 0.76 and 0.89. These neurons were represented by 19 and 13 Gabor features,

respectively (Fig. 2b and c, right panels), and their encoding filters (weighted sums of the Gabor filters) were spatially localized (Fig. 2b and c, insets in the right panels). In the representative plane presented in Fig. 1, the median of the prediction performance of the encoding model (i.e., the correlation coefficient between the observed and predicted responses) was 0.34 (25–75th percentiles: 0.16–0.52, n = 726 cells, Supplementary Fig. 1f), and the median performance of all cells across planes was 0.24 (25–75th percentiles: 0.07–0.45, n = 12755 cells across 24 planes, Supplementary Fig. 1i). An examination of the non-linear scaling function revealed that this step suppressed weak predicted responses and enhanced strong predicted responses (see Fig. 2d and e for a representative cell and average across planes, respectively), suggesting that this non-linear step enhanced the sparseness of the predicted response obtained from the linear step (i.e., linear regression by feature values). On average, 2.0% of the features (25/1248 features, 25–75th percentiles: 2.0–2.1%) were represented in each cell of the example plane (upper panels in Fig. 2f and Supplementary Fig. 1g), and 2.1% were represented in each cell of all recorded cells across all planes (~26/1248 features, 25–75th percentiles: 0.9–4.9%, n = 12755 cells, Fig. 2h and Supplementary Fig. 1k). These features were related to the RF structure of each cell (Supplementary Fig. 2). The RF structure of each cell was estimated using the regularized inverse method[32-34] (see Methods). The regression weights of the Gabor features in the encoding model were positively correlated with the similarity between the corresponding Gabor filter and the RF structure (Supplementary Fig.2a–d).

The Gabor features encoded in one cell partially overlapped with those of other cells in a local population (Fig. 2i). Among 19 and 13 Gabor features represented by the two example cells (Fig. 2b and c), only two features overlapped. For all cell pairs across all planes, the median overlap was 3.4% (25–75th percentile: 0.0–9.6% relative to features represented by each cell, Fig. 2i and Supplementary Fig. 1h and 1l). The feature overlap between neurons was positively correlated with the similarity of RF structure (Supplementary Fig. 2e–j). Based on these findings, the Gabor features encoded by individual neurons in a local population were highly diverse and partially overlapped.

The analysis of the encoding model also revealed how the individual Gabor features were encoded across neurons (upper left and bottom panels in Fig. 2f and g). As the spatial frequency (SF) of the Gabor filter

increased (i.e. the scale decreased), the corresponding feature contributed to the visual responses of fewer neurons (Fig. 2g). This pattern likely reflected the fact that Gabor filters with a low SF (i.e., a large scale) covered more of the neuron's RF, whereas Gabor filters with a high SF (i.e. a small scale) affected the responses of fewer neurons. Furthermore, almost all features contributed to the responses of at least one cell (100% in the plane presented in Fig. 2f and 100% [99.4–100%] across all planes, median [25–75th percentiles], Fig. 2j).

Image reconstruction from the activities of the neuronal population

The encoding model revealed the Gabor features represented by each neuron. We next examined whether the features encoded in a local population of neurons were sufficient to represent the visual contents of the natural images. We reconstructed stimulus images from the neuronal activities to evaluate information about visual contents in the population activity[15-19]. Using the same Gabor features as in the encoding model, each Gabor feature value was subjected to a linear regression of the neuronal activities of multiple neurons (Fig. 3a and Supplementary Fig. 3a). Each Gabor feature value was independently reconstructed. Then, the sets of reconstructed feature values were transformed into images (Fig. 3a, see Methods). The reconstruction performance was estimated with a different dataset from the dataset used in the regression (10-fold cross validation, see Methods).

We first used all simultaneously recorded neurons to reconstruct the image. In the examples of the plane (n = 726 neurons, presented in Figs. 1 and 2), the rough structures of the stimulus images were reconstructed from the population activities ("All-cells" in Fig. 3b). The reconstruction performances (pixel-to-pixel correlations between stimulus images and reconstructed images) were 0.45 [0.36–0.56] (median [25–75th percentiles] of 200 images) in the representative plane (n = 726 cells, Fig. 3c upper panel) and 0.36 [0.31–0.38] across all planes (n = 24 planes, "All cells" in Fig. 3d). Thus, the visual contents of natural images were extracted linearly from the neuronal activities of the local population in V1

The encoding model used in the previous section revealed how each neuron encodes the Gabor features

(Fig. 2f). We next examined whether these encoded features were sufficient for the representation of visual contents. In this analysis, each Gabor feature value was reconstructed with a subset of neurons selected using the encoding model (cell-selection model, Supplementary Fig. 3a, and see Methods). In this model, different subsets of neurons were used to reconstruct different features (Fig. 2f). Across all features, almost all neurons were used to reconstruct at least one feature (Fig. 2j). The examples of the reconstructed images from the cell-selection model are presented in Fig. 3b (Cell-selection). The reconstruction performance of the cell-selection model was comparable to or even slightly higher than the model using all cells (R = 0.49 [0.37–0.59] for the representative plane, Fig 3c lower panel, and 0.36 [0.32–0.39] for all planes, median [25–75th percentiles], $p = 4.0\times10^{-4}$ using the signed-rank test, Fig. 3d). Thus, the Gabor features encoded in individual cells in a population captured sufficient information about the visual contents of the natural image. When the neurons were selected to maximize the reconstruction of each feature, the image reconstruction performance was only slightly improved (Supplementary Fig. 3b–h). Thus, main information about the visual contents was captured by the cell-selection model.

Visual contents of natural images are linearly decodable from small numbers of responsive neurons

Single natural images activated small numbers of neurons in a local population (Fig. 1). We next examined whether these small number of highly responsive neurons were sufficient to reconstruct a single image. For this purpose, we changed the number of neurons used in the reconstruction of each image and examined how many responsive neurons were sufficient for each image reconstruction. Parameters (weights and biases) of the cell-selection model were used in the reconstruction, and only the number of neurons used in the reconstruction was changed in this analysis.

Representative results are presented in Fig. 4a–c. In each image, neurons were sorted by visual response amplitude (descending order) first among the highly responsive neurons (red dots in Fig. 4a–c) and then among the remaining neurons (black dots in Fig. 4a–c). The image was reconstructed by top N neurons (N = 1–726 cells), and the reconstruction performances were plotted against the numbers of neurons used (Fig.

4a–d). All highly responsive neurons or even fewer neurons were sufficient to reconstruct the image to a level that was fairly comparable to the image created with all neurons (Fig. 4a–d). In summary, the performance of the highly responsive neurons was slightly better than the performance of all neurons (Representative plane: R = 0.52 [0.40–0.64] for the responsive neurons and 0.49 [0.37–0.59] for all neurons, Fig. 4f. Across planes: R = 0.38 [0.34–0.44] for the responsive neurons and 0.35 [0.31–0.40] for all neurons, median [25-75th percentiles], $p = 3.2\times10^{-4}$ using the signed-rank test, n = 24 planes, Fig. 4g). On average, only approximately 20 neurons were sufficient to achieve 95% of the peak performance (vertical line in Fig. 4d). Thus, the visual contents of the single natural images were linearly decodable from small numbers of highly responsive neurons.

The features represented by individual neurons should be diverse to represent features in a natural image using a small number of neurons. Fig. 4e illustrates how individual responsive neurons contributed to the image reconstruction in the case presented in Fig 4a. Each neuron had a specific pattern of contributions (reverse filter: sum of Gabor filters × weights, see Methods), and the patterns varied across neurons (Fig. 4e top panels), while partially overlapping in the visual field. In neuron pairs that were highly responsive to the same image, the number of overlapping Gabor features were slightly increased compared to all pairs, but the percentage was still less than 10% (7.1% [1.0–16%] of features for the all pairs and 8.1% [6.3–10%] of features for 24 planes, Fig. 4h–j, cf. Fig. 2g). These small overlaps and diversity in the represented features among neurons should be useful for the representation of natural images by the relatively small number of highly responsive neurons.

Robust image representation by neurons with spatially overlapping representation

We next examined whether a single image was robustly represented by the small number of responsive neurons. We computed reconstruction performance after dropping single cells (Fig. 5a and b. Cell # on the x-axis is the same as in Fig. 4d). Single cell-drop had only a small effect on the reconstructed image (middle panels in Fig. 5a). On average, at most 5% reduction of reconstruction performance was observed for the best-responding neurons, and there were almost no effects in most neurons (Fig. 5b). This indicates that an

image was robustly represented by highly responsive neurons against cell drop.

We found that this robustness was due to spatial overlap of representation patterns (i.e., reverse filters) among responsive neurons (Fig. 5c). We selected nine neurons which represented the upper part of the image and whose representation patterns spatially overlapped but variable in structure (overlapping cells, top panels in Fig. 5c and Supplementary Fig. 4). Although single-cell drop had almost no effect on the reconstructed local image (bottom panels in Fig. 5c), sequential drop of these cells gradually degraded the upper part of the reconstructed image (Fig. 5d). Pixel values in the overlapping area of the reconstructed image gradually decreased as the number of dropped cells increased (Fig. 5e and f). These results indicate that the robust image representation was due to neurons with spatially overlapping representation.

Independent activities among subsets of neurons provide robust image representation against trial-to-trial variability

We further analyzed whether this overlapping representation is useful to reduce trial-to-trial variability of image representation. Cortical neurons often show trial-to-trial variability in response to repetitions of the same stimulus. If neurons with spatially overlapping representations showed independent or negatively correlated activities, integration of activities among these neurons should reduce the variability of image representations[35, 36].

Across-trial variability of the reconstructed images of the example case (shown in Fig. 5) is exhibited in Figure 6. Single-trial reconstructed images from all responsive neurons (57 cells) were mostly stable across trials and were distorted only in a few trials (e.g., trial 10, Fig. 6a). By contrast, reconstructed images from individual neurons were variable across trials (Fig. 6c). Importantly, some neuron pairs showed positively correlated representation across trials, other pairs showed almost independent representation across trials. Thus, integration of activities among the neurons with overlapping representation resulted in reliable representation across trials, even though the activity patterns of individual neurons were variable across trials (Fig. 6d).

Based on this observation, we hypothesized that some neurons which show positively correlated activities

form a functional cluster and work together, while neurons between different clusters show independent or negatively correlated activities to reduce variability of image representations. In the case shown in Fig. 6, the nine neurons formed three clusters based on their noise correlations (Fig. 7a and Supplementary Fig. 5a, see Methods). Neurons with overlapping representations usually formed two clusters (Fig. 7b). Importantly, the neuron pairs between different clusters exhibited almost zero or slightly negative correlations (between-cluster pair: -0.05 [-0.22–0.12], and within-cluster pair: 0.26 [0.09–0.42], median and 25-75th percentile, Fig. 7c, blue). This tendency was independent of the number of clusters (Supplementary Fig. 5i). Similarity of reverse filters for the within-cluster pair was almost comparable with that for the between-cluster pair (Supplementary Fig. 5b), indicating that reverse filter structures did not simply explain the structure of noise correlation. Further, cortical positions of neurons did not explain the structure of noise correlation, because neurons in different clusters were spatially intermingled in FOVs (Supplementary Fig. 6).

We next compared reconstructed images obtained from different clusters (Supplementary Fig. 5c, d, h). Importantly, the images were similar between clusters (pixel-to-pixel correlation of reconstructed image: 0.33 [0.11–0.52], median [25-75th percentile], Supplementary Fig. 5d), indicating that the clusters represented similar information. At single-trial, the reconstructed images from individual clusters were still variable across trials (Fig. 7d), due to the relatively high noise correlation within clusters. We further compared trial-to-trial variability of reconstructed images between clusters. As predicted from almost zero noise correlation between clusters, the trial-to-trial variability of reconstructed image was almost independent between the clusters (Fig. 7d for the representative case and Supplementary Fig. 5e, i for summary data. Across-trial correlation coefficients of reconstructed images between clusters were: -0.08 [-0.25–0.09], median [25–75th percentile]). Integration of the multiple clusters resulted in more reliable image representation compared to individual clusters (Supplementary Fig. 5f). These results indicate that integration of activities among the clusters provides robust representation against trial-to-trial response variability.

Representation of multiple natural images in a local population

Finally, we examined how multiple natural images were represented in a population of responsive neurons (Fig. 8a–c). Figs. 8a and b provide an example of the representative plane shown in the previous figures (n = 726 cells). Natural images were sorted by reconstruction performance (y-axis in Fig. 8a), and the cells responding to each image are plotted in each row. First, as the number of images increased, new responsive cells are added, and the total number of responsive cells used for the reconstructions quickly increased (right end of the plot on each row, Fig. 8a). At approximately 50 images, the number of newly added responsive cells quickly decreased, and the increase in the total number of responsive cells slowed, indicating that the newly added image was represented by a combination of the already plotted responsive cells (i.e., the neurons that responded to other images), which was due to the small overlap in responsive cells between images (Fig. 1j). These findings are summarized in Figures 8b and c in which the number of newly added cells quickly decreased to zero as the number of images increased (red lines in Fig. 8b and c for the representative case and for all planes, respectively). Therefore, although only 4.8% responsive neurons overlapped between images (Fig. 1j), this small overlap is useful for the representation of many natural images by a limited number of responsive neurons.

We also analyzed whether the features represented by the local populations of the responsive neurons were sufficient to represent all features of the natural images. If the features in a local population are sufficient to represent all natural images, all features of the natural images should be accurately represented by the combination of features in the individual cells in a population. We represented a set of features in each image by linear regression of weights (i.e., features) of all responsive cells from the reconstruction model (cell-selection model) and computed the fitting errors (see Methods, Fig. 8d). The median error was less than 10% for all images and all planes (8.2% [4.5–15.2%] for all image cases and 5.7% [4.9–16%], n = 24 for the planes, Fig. 8e and f). Based on this result, features that sufficiently represent the visual contents of natural images are encoded in neurons in a local population.

Discussion

In the mouse V1, single natural images activated a small number of neurons (2.7%) which was sparser than that predicted by the linear model. The Gabor features represented in the individual neurons only slightly overlapped between neurons, indicating diverse representations. Visual contents of natural images were linearly decodable from the small number of active neurons (about 20 neurons), which was achieved by the diverse representations. A local part of the image was robustly represented by neurons whose representation patterns partially overlapped. These neurons with overlapping representation formed a small number of functional clusters which represented similar local image but were active independently across trials. Thus, integration of activities across the clusters led to robust representation against across-trial response variability. Further, small share of responsive neurons between the images helped a limited number of the responsive neurons to represent multiple natural images. Finally, the visual features represented by all the responsive neurons provided a good representation of the original visual features in the natural images.

Visual responses to natural images or movies in V1 are sparse at the single cell level (high lifetime sparseness)[2, 3, 5-9] and at the level of populations (population sparseness)[3, 5, 6, 14]. Recently, recordings of local population activity using two-photon Ca^{2+} imaging have enabled us to precisely evaluate the population sparseness[5, 14, 37]. We confirmed that a single natural image activated only a small number of neurons. Encoding model analysis indicated that visual responses in individual neurons were sparser than that predicted from a linear model (Fig. 2d, e). Here, this sparse activity was shown to contain sufficient and even robust information to represent the natural image contents. Image reconstruction is useful for evaluating the information contents represented by the neuronal activity and is widely used to analyze populations of single unit activities in response to natural scenes or movies in LGN[15] and fMRI data from several visual cortical areas[16-19]. The former[15] study used “pseudo-population” data collected from several experiments, and the latter studies[16-19] used blood oxygen level-dependent (BOLD) signals that indirectly reflect the average of local neuronal activity. Thus, it has not been examined whether and how the visual contents of natural images are represented in simultaneously recorded populations of single neurons in the cortex. We revealed that visual contents of single natural images were linearly decodable from relatively small number of responsive neurons

in a local population. It has been proposed that information is easy to be read out from the sparse representation[4]. Indeed, the sparse population activity increases the discriminability of two natural scenes by rendering the representations of the two scenes separable[5]. Our results extend this in that information about visual contents encoded in sparsely active neurons is linearly accessible, suggesting that downstream areas are easy to read out images from the sparse representation in V1.

The visual features encoded by individual neurons should be diverse so that a small number of active neurons represent the complex visual features of the image. Although RF structures in the local population of mouse V1 have already been reported[21, 22, 33, 34], their diversity has not been analyzed with respect to natural image representation. In the present study, the visual features represented by sparsely active neurons were sufficiently diverse to represent visual contents of natural images. Computational models for natural image representation have suggested that sparse activity and number of available neurons affect diversity of RF structures[20, 38-40].

We also demonstrated that sparsely active neurons robustly represented an image against trial-to-trial response variability. Although a computational model proposed sparse and overcomplete representation as optimal representation for natural images with unreliable neurons[20], this has never been addressed experimentally. We demonstrated that the robust representation was mainly achieved by the diverse, partially overlapping representations, consistent with the overcomplete representation. It has been reported that subregions of receptive fields of some V1 neurons partially overlap[21]. Our results suggest that such overlap may be useful for the robust image representation. We further revealed that neurons with overlapping reverse filters formed functional clusters and integration across the clusters reduced the trial-to-trial variability, suggesting a new representation scheme in which information is reliably represented, while representing neuronal patterns change across trials. This seems to be similar to “drop-out” in deep learning[41] and may be useful to avoid overfitting and local minimum problems in learning.

Our analysis also revealed how multiple natural images were represented in a local population of responsive neurons. A single natural image activated specific subsets of neurons, whereas most neurons in a

local population responded to at least one of the images, supporting sparse, distributed code proposed in a previous study. The overlap of responsive neurons between images involved only 4.8% of the responsive cells (Fig. 1i). However, due to this small overlap, many natural images were represented by a limited number of responsive neurons (Fig. 5a–c). Furthermore, the features of all responsive neurons in a local population were sufficient to represent all the natural images used in the present study (Fig. 5d–f). Based on these findings, any natural image could be represented by a combination of responsive neurons in a local population.

In summary, this work highlighted how the visual contents of natural images are sufficiently and even robustly represented in sparsely active V1 neurons. The diverse, but partially overlapping representation helps the small number of neurons to represent a complex image robustly against across-trial variability. We propose a new representation scheme in which information is reliably represented with variable neuronal patterns across trials and which may be effective to avoid over-fitting in learning.

Acknowledgements

We thank Ms. Y. Sono, A. Hayashi, T. Inoue, A. Ohmori, A. Honda, M. Nakamichi for animal care, and all members of Ohki laboratory for support and discussions. This work was supported by grants from Core Research for Evolutionary Science and Technology (CREST)—Japan Agency for Medical Research and Development (AMED) (to K.O.), Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant number 25221001 and 25117004 to K.O. and 15K16573, 17K13276 to T.Y.), International Research Center for Neurointelligence (WPI-IRCN), JSPS (to K.O.), Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS)—AMED (to K.O.), Strategic International Research Cooperative Program (SICP)—AMED (to K.O.), grants from the Ichiro Kanehara Foundation for the Promotion of Medical Sciences and Medical Care, and the Uehara Memorial Foundation (to T.Y.).

Author contributions

T.Y. and K.O. designed the research. T.Y. performed experiments. T.Y. and K.O. analyzed data and wrote the

manuscript. K.O. supervised the research.

Competing financial interests.

We declare no competing financial interests.

References

1. Rolls, E.T. & Tovee, M.J. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol* **73**, 713-726 (1995).
2. Vinje, W.E. & Gallant, J.L. Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision. *Science* **287**, 1273-1276 (2000).
3. Weliky, M., Fiser, J., Hunt, R.H. & Wagner, D.N. Coding of natural scenes in primary visual cortex. *Neuron* **37**, 703-718 (2003).
4. Olshausen, B.A. & Field, D.J. Sparse coding of sensory inputs. *Curr Opin Neurobiol* **14**, 481-487 (2004).
5. Froudarakis, E.*, et al.* Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat Neurosci* **17**, 851-857 (2014).
6. Yen, S.C., Baker, J. & Gray, C.M. Heterogeneity in the responses of adjacent neurons to natural stimuli in cat striate cortex. *J Neurophysiol* **97**, 1326-1341 (2007).
7. Yao, H., Shi, L., Han, F., Gao, H. & Dan, Y. Rapid learning in cortical coding of visual scenes. *Nat Neurosci* **10**, 772-778 (2007).
8. Tolhurst, D.J., Smyth, D. & Thompson, I.D. The sparseness of neuronal responses in ferret primary visual cortex. *J Neurosci* **29**, 2355-2370 (2009).
9. Willmore, B.D., Mazer, J.A. & Gallant, J.L. Sparse coding in striate and extrastriate visual cortex. *J Neurophysiol* **105**, 2907-2919 (2011).
10. Field, D.J. What Is the Goal of Sensory Coding. *Neural Comput* **6**, 559-601 (1994).
11. Jones, J.P. & Palmer, L.A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* **58**, 1233-1258 (1987).
12. Olshausen, B.A. & Field, D.J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607-609 (1996).
13. Bell, A.J. & Sejnowski, T.J. The "independent components" of natural scenes are edge filters. *Vision Research* **37**, 3327-3338 (1997).
14. Tang, S.*, et al.* Large-scale two-photon imaging revealed super-sparse population codes in the V1 superficial layer of awake monkeys. *Elife* **7** (2018).

15. Stanley, G.B., Li, F.F. & Dan, Y. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J Neurosci* **19**, 8036-8042 (1999).

16. Miyawaki, Y., *et al.* Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders. *Neuron* **60**, 915-929 (2008).

17. Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M. & Gallant, J.L. Bayesian reconstruction of natural images from human brain activity. *Neuron* **63**, 902-915 (2009).

18. Nishimoto, S., *et al.* Reconstructing visual experiences from brain activity evoked by natural movies. *Curr Biol* **21**, 1641-1646 (2011).

19. Horikawa, T., Tamaki, M., Miyawaki, Y. & Kamitani, Y. Neural decoding of visual imagery during sleep. *Science* **340**, 639-642 (2013).

20. Doi, E. & Lewicki, M.S. Sparse Coding of Natural Images Using an Overcomplete Set of Limited Capacity Units. *In: Advances in Neural Information Processing Systems (NIPS 2004)* **17**, 377-384 (2005).

21. Smith, S.L. & Hausser, M. Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nat Neurosci* **13**, 1144-1149 (2010).

22. Bonin, V., Histed, M.H., Yurgenson, S. & Reid, R.C. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J Neurosci* **31**, 18506-18521 (2011).

23. Kampa, B.M., Roth, M.M., Gobel, W. & Helmchen, F. Representation of visual scenes by local neuronal populations in layer 2/3 of mouse visual cortex. *Front Neural Circuits* **5**, 18 (2011).

24. Ko, H., *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87-91 (2011).

25. Marshel, J.H., Garrett, M.E., Nauhaus, I. & Callaway, E.M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72**, 1040-1054 (2011).

26. Miller, J.e.K., Ayzenshtat, I., Carrillo-Reid, L. & Yuste, R. Visual stimuli recruit intrinsically generated cortical ensembles. *Proceedings of the National Academy of Sciences* **111**, E4053-E4061 (2014).

27. Rikhye, R.V. & Sur, M. Spatial Correlations in Natural Scenes Modulate Response Reliability in Mouse Visual Cortex. *J Neurosci* **35**, 14661-14680 (2015).

28. Olshausen, B.A. & Field, D.J. How Close Are We to Understanding V1? *Neural Comput* **17**, 1665-1699 (2005).

29. Shoham, S., O'Connor, D.H. & Segev, R. How silent is the brain: is there a "dark matter" problem in neuroscience? *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **192**, 777-784 (2006).

30. Yoshida, T. & Ohki, K. Visual image reconstruction from neuronal activities in the mouse primary visual cortex. *Program No. 415.17. 2015 Neuroscience Meeting Planner. Chicago, IL:*

447 *Society for Neuroscience, 2015. Online.* (2015).
448 31. Yoshida, T. & Ohki, K. Representation of natural image contents by sparsely active neurons in
449 visual cortex. *bioRxiv* (2018).
450 32. Smyth, D., Willmore, B., Baker, G.E., Thompson, I.D. & Tolhurst, D.J. The Receptive-Field
451 Organization of Simple Cells in Primary Visual Cortex of Ferrets under Natural Scene Stimulation.
452 *The Journal of Neuroscience* **23**, 4746-4759 (2003).
453 33. Ko, H., *et al.* The emergence of functional microcircuits in visual cortex. *Nature* **496**, 96-100
454 (2013).
455 34. Cossell, L., *et al.* Functional organization of excitatory synaptic strength in primary visual
456 cortex. *Nature* **518**, 399-403 (2015).
457 35. Shadlen, M.N. & Newsome, W.T. Noise, neural codes and cortical organization. *Current Opinion*
458 *in Neurobiology* **4**, 569-579 (1994).
459 36. Zohary, E., Shadlen, M.N. & Newsome, W.T. Correlated neuronal discharge rate and its
460 implications for psychophysical performance. *Nature* **370**, 140 (1994).
461 37. Greenberg, D.S., Houweling, A.R. & Kerr, J.N. Population imaging of ongoing neuronal activity
462 in the visual cortex of awake rats. *Nat Neurosci* **11**, 749-751 (2008).
463 38. Rehn, M. & Sommer, F.T. A network that uses few active neurones to code visual input predicts
464 the diverse shapes of cortical receptive fields. *J Comput Neurosci* **22**, 135-146 (2007).
465 39. Olshausen, B.A., Cadieu, C.F. & Warland, D.K. Learning real and complex overcomplete
466 representations from the statistics of natural images. *SPIE Optical Engineering + Applications* **7446**,
467 11 (2009).
468 40. Olshausen, B.A. Highly overcomplete sparse coding. **8651**, 86510S (2013).
469 41. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple
470 way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929-1958 (2014).
471 42. Nimmerjahn, A., Kirchhoff, F., Kerr, J.N. & Helmchen, F. Sulforhodamine 101 as a specific
472 marker of astroglia in the neocortex in vivo. *Nat Methods* **1**, 31-37 (2004).
473 43. Ohki, K., Chung, S., Ch'ng, Y.H., Kara, P. & Reid, R.C. Functional imaging with cellular
474 resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597-603 (2005).
475 44. Hagihara, K.M., Murakami, T., Yoshida, T., Tagawa, Y. & Ohki, K. Neuronal activity is not
476 required for the initial formation and maturation of visual selectivity. *Nat Neurosci* **18**, 1780-1788
477 (2015).
478 45. Mank, M., *et al.* A genetically encoded calcium indicator for chronic in vivo two-photon imaging.
479 *Nat Methods* **5**, 805-811 (2008).
480 46. van Hateren, J.H. & van der Schaaf, A. Independent component filters of natural images
481 compared with simple cells in primary visual cortex. *Proc Biol Sci* **265**, 359-366 (1998).

47. Olmos, A. & Kingdom, F.A. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* **33**, 1463-1473 (2004).
48. Peirce, J.W. Generating Stimuli for Neuroscience Using PsychoPy. *Front Neuroinform* **2**, 10 (2008).
49. Kerlin, A.M., Andermann, M.L., Berezovskii, V.K. & Reid, R.C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67**, 858-871 (2010).
50. Treves, A. & Rolls, E.T. What determines the capacity of autoassociative memories in the brain? *Network: Computation in Neural Systems* **2**, 371-397 (1991).
51. Lee, T.S. Image Representation Using 2D Gabor Wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 959-971 (1996).
52. Kay, K.N., Naselaris, T., Prenger, R.J. & Gallant, J.L. Identifying natural images from human brain activity. *Nature* **452**, 352-355 (2008).
53. Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag New York, Inc., 2006).

Methods

All experimental procedures were approved by the local Animal Use and Care Committee of Kyushu University.

Animal preparation for two-photon imaging

C57BL/6 mice (male and female) were used (Japan SLC Inc., Shizuoka, Japan). Mice were anaesthetized with isoflurane (5 % for induction, 1.5 % for maintenance during surgery, ~0.5% during imaging with a sedation of < 0.5mg/kg chlorprothixene, Sigma-Aldrich, St. Louis, MO, USA). The head skin was removed from the head, and the skull over the cortex was exposed. A custom-made metal plate for head fixation was attached with dental cement (Super Bond, Sun Medical, Shiga, Japan), and a craniotomy (~3mm in diameters) was performed over the primary visual cortex (center position: 0–1 mm anterior to lambda, +2.5–3mm lateral to the midline). A mixture of 0.8 mM Oregon Green BAPTA1-AM (OGB1, Life Technologies, Grand Island, NY, USA) dissolved with 10% Pluronic (Life Technologies) and 0.025 mM sulforhodamine 101[42] (SR101, Sigma-Aldrich) was pressure-injected with a Picospritzer III (Parker Hannifin, Cleveland, OH, USA) at the

depth of 300–500 μm from the brain surface. The cranial window was sealed with a coverslip and dental cement. The imaging experiment began at least one hour after the OGB1 injection.

Two-photon Ca^{2+} imaging.

Imaging was performed with a two-photon microscope (A1R MP, Nikon, Tokyo, Japan) equipped with a 25× objective (NA 1.10, PlanApo, Nikon) and Ti:sapphire mode-locked laser (MaiTai Deep See, Spectra Physics, Santa Clara, CA, USA)[43, 44]. OGB1 and SR101 were excited at a wavelength of 920 nm. Emission filters of 525/50nm and 629/56nm were used for the OGB1 and SR101 signals, respectively. The fields of view (FOVs) were 338 × 338 μm (10 planes from 7 mice) and 507 × 507 μm (14 planes from 7 mice) at 512 × 512 pixels. The sampling frame rate was at 30Hz using a resonant scanner.

Visual stimulation.

Before beginning the recording session, the retinotopic position of the recorded FOV was determined using moving grating patches (lateral or upper directions, 99.9% contrast, 0.04 cycle/degrees, 2 Hz temporal frequency, 20 and 50 degrees in diameter) while monitoring the changes in signals over the entire FOV. The lateral or upper motion directions of the grating were used to activate many cells because the preferred directions of mouse V1 neurons are slightly biased towards the cardinal directions[44, 45]. First, the grating patch of 50 degrees in diameter was presented in one of 15 (5 × 3) positions that covered the entire monitor to roughly determine the retinotopic position. Then, the patch of 20 degrees in diameter was presented on the 16 (4 × 4) positions covering an 80 × 80-degree space to finely identify the retinotopic position. The stimulus position that induced the maximum visual response of the entire FOV was set as the centre of the retinotopic position of the FOV.

A set of circular patches of grey-scaled, contrast-enhanced natural images (200 image types) was used as the visual stimuli for response prediction and natural image reconstruction (60 degrees in diameter, 512 × 512 pixels, with a circular edge (5 degrees) that was gradually mixed to grey background). Each natural image was

adjusted to almost full contrast (99.9%). The mean intensity across pixels in each image was adjusted to an approximately 50% intensity. Original natural images were obtained from the van Hateren's Natural Image Dataset (http://pirsquared.org/research/#van-hateren-database)[46] and the McGill Calibrated Color Image Database (http://tabby.vision.mcgill.ca/html/welcome.html)[47]. During image presentation, one image type was consecutively flashed three times (three 200-ms presentations interleaved with 200 ms of grey screen), and the presentation of the next image was initiated after the presentation of the grey screen for 200 ms. Images were presented in a pseudo-random sequence in which each image was presented once every 200 image types. Each image was presented at least 12 times (i.e., 12 trials) in a total recording session. We did not set a long interval between image flashes to reduce the total recording time and increase the number of repetitions. In this design, the tail of the Ca^{2+} response to one image invaded the time window of the next image presentation (Fig. 1b). Although this overlap may have affected the visual responses between two adjacent images, many trial repetitions (> 11 times for each image) in the pseudo-random order and the sparse responses to natural images (Fig. 1) minimized the effects of response contamination between two consecutive images.

Moving, square gratings (8 directions, 0.04 cycle/degrees, 2 Hz temporal frequency, 60-degree patch diameter) were presented at the same position as the natural image on the screen. Each direction was presented for 4 secs interleaved by 4 secs of the grey screen. The sequence of directions was pseudo-randomized, and each direction was presented 10 times in a recording session.

All stimuli were presented with PsychoPy[48] on a 32-inch LCD monitor (Samsung, Hwaseong, South Korea) with a 60-Hz refresh rate, and the timing of the stimulus presentation was synchronized with the timing of image acquisition using a TTL pulse counter (USB-6501, National Instruments, Austin, TX, USA).

The entire recording session for one plane was divided into several recording sessions (4–6 trials/sub-session and 15–25 min for each session). Each recording session was interleaved by approximately 5–10 minutes of rest time during which the slight drift of the FOV was manually corrected. Every two or three sessions, the retinotopic position of the FOV was checked with the grating patch stimuli during the resting period. The recording was terminated, and then data were discarded if the retinotopic position was shifted

(probably due to eye movement). The recordings were performed in one to three planes of different depths and/or positions in each animal (1.7 ± 0.8 planes, mean ± standard deviation).

Data analysis.

All data analysis procedures were performed using MATLAB (Mathworks, Natick, MA, USA). Recorded images were phase-corrected and aligned between frames. The averaged image across frames was used to determine the region of interests (ROIs) of individual cells. After removing slow SF component (obtained with a Gaussian filter with a sigma of approximately five times the soma diameter), the frame-averaged image was subjected to a template matching method in which two-dimensional difference of Gaussian (sigma1: 0.26 × soma diameter that was adjusted for zero-crossing at the soma radius, sigma2: soma diameter) was used as a template for the cell body. Highly correlated areas between the frame-averaged image and the template were detected as ROIs for individual cells. ROIs were manually corrected via visual inspection. SR101-positive cells (putative astrocytes[42]) were removed from the ROI. Time course of calcium signal in each cell was computed as an average of all pixel's intensities within an ROI. Signal contamination from out of focus plane was removed by a previously reported method[44, 49]. Briefly, a signal from ring-shaped area surrounding each ROI was multiplied by a factor (contamination ratio) and subtracted from the signal of each cell. The contamination ratio was determined to minimize the difference between the signal from a blood vessel and the surrounding ring shape region multiplied by the contamination ratio. The contamination ratios were computed for several blood vessels in the FOV, and the mean value for several blood vessels was used for all cells in the FOV.

The average response of 200-ms grey screen period just before each image was subtracted from the average response of the time course during the last 200 ms of the stimulus period (during 3rd flash of each image at approximately time of the peak Ca^{2+} transient) to compute visually evoked responses. The evoked response was normalized for each cell by dividing by the standard deviation across all visual responses (200 images × trials, z-scored response). If the z-scored response to one image was significantly different from 0 (p

< 0.01 using signed-rank test across trials) and the across-trial average of the z-scored response was greater than 1, the response was considered significant for the image. The population sparseness (s) was computed using the equation described in previous studies[2, 3, 50] as follows: $s = [1-(\sum Ri)^2/(N\sum Ri^2)]/(1-1/N)$, where Ri is the evoked response of ith cell, and N is the number of cells (i = 1–N).

Natural images were scaled so that maximum intensity and minimum intensities were 1 and -1, respectively, and gray intensity was 0. A square position (50 × 50 degrees) of the centre of natural image patch was extracted and down-sampled to 32 × 32-pixel image. The down-sampled image was used to analyze the Gabor features, response prediction and image reconstruction.

Gabor features.

A set of spatially overlapping Gabor filter wavelets was prepared to extract the visual features of the natural images[10, 51, 52]. The down-sampled images were first subjected to the set of Gabor filters to obtain Gabor feature values. Single feature value corresponds to a single wavelet filter.

Gabor filters have four orientations (0, 45, 90, and 135 degrees), two phases, and 4 sizes (8 × 8, 16 × 16, 32 × 32, and 64 × 64pixels) located on 11 × 11, 5 × 5, 3 × 3, and 1 × 1 grids (Supplementary Fig. 1a and b). Therefore, the three smaller scale filters were spatially overlapped with each other. The spatial frequencies of the four scale sizes of the Gabor wavelets were 0.13, 0.067, 0.033, and 0.016 cycle/degrees (cpd). This filter set was almost self-inverting, i.e., the feature values obtained by applying an image to the wavelet set were transformed back to the image by summing the filters after multiplying by the feature values[51]. The Gabor filters and the transformations were based on an open source program (originally written by Drs. Daisuke Kato and Izumi Ohzawa, Osaka University, Japan, https://visiome.neuroinf.jp/modules/xoonips/detail.php?item_id=6894).

Encoding model.

In the encoding model for response prediction, single-cell responses were predicted using a linear regression

analysis of selected Gabor feature values (Fig. 2a and Supplementary Fig. 1a–e). The encoding model was created independently for each cell. First, Pearson's correlation coefficients between the response and each feature value were computed. Then, using one of the preset values for the correlation coefficient as a threshold (12 points ranging from 0.05 to 0.35, Supplementary Fig.1c–e), only the more strongly correlated features were selected (feature selection) and used in the regression analysis. The weight and bias parameters of the regression were estimated by Bayesian linear regression with an expectation-maximization algorithm which is almost equivalent to linear regression with L2 regularization[53]. After the regression analysis, the non-linearity of predicted response was adjusted via a rectification step using the following equation[34], predicted response = $A/[1 + \exp(\alpha x + \beta)]$, where x is the output of the regression and A, α, and β are parameters to be estimated. This step only scaled the regression output without changing the regression parameters (i.e., weights and biases). The response prediction of the model was estimated by 10-fold cross validations (CVs) in which the response data for 180 images were used to estimate the parameters, and the remaining data for 20 images were used to evaluate the prediction. In the 10-fold CVs, all images were used once as test data. The prediction performances were estimated using Pearson's correlation coefficients between the observed (trial-average) and predicted responses. Encoding models were created for all preset threshold values for feature selection, and the model that exhibited the best prediction performance was selected as the final model. In the analysis of weights (i.e., feature) overlap between the two cells, the percentage of overlapping weights relative to the number of non-zero weights was computed for each cell and averaged between the two cells in the pair.

Using the same dataset as used in the encoding model, The RF structure was estimated for each cell using a regularized inverse method[32-34]. The regularized inverse method uses one hyper-parameter (regularized parameter). In the 10-fold CVs, the RF structure was estimated with the training dataset using one of preset regularized parameters (13 logarithmically spaced points between 10^{-3} and 10^{3}). The visual response was predicted using the estimated RF and test dataset. The Prediction performance of visual response was estimated by determining Pearson's correlation coefficients between the observed and the predicted responses. RFs were estimated for all values of the pre-set regularized parameters, and the value that resulted in the best

response prediction was selected for the final RF model.

Image reconstruction.

In the image reconstruction, each Gabor feature value was linearly regressed by the single-trial activities of multiple cells. In the 10 CVs, the weights and a bias parameter were estimated using the same algorithm as in the encoding model with the training dataset (see above), and each Gabor feature value was reconstructed from the visual response in the test dataset. After each Gabor feature was independently reconstructed, sets of reconstructed feature values were transformed into images as described above (**Gabor features** section, also see Fig. 3a). Reconstruction performance was evaluated by determining pixel-to-pixel Pearson's correlation coefficient between the stimulus and reconstructed images. In the cell-selection model (Fig. 3), each feature value was reconstructed with the subset of cells that were selected using the encoding model (Fig. 2f and Supplementary fig. 3a), and almost all cells were used across features (Fig. 2j). In the encoding model, each cell was represented by a subset of features that affected the cell's response. Thus, in the cell-selection model, each feature was only reconstructed by the cells that encoded information about the reconstructed feature (Supplementary Fig. 3a).

In the analysis of the weights (i.e., feature) overlap between cells, the percentage of overlapping weights relative to the number of non-zero weights was computed for each cell and averaged between the two cells in the pair.

In the analyses shown in Fig. 4a–d and Fig. 5a–b, cells were separated into responsive and non-responsive cells in each image and sorted by their response amplitude in descending order (i.e., from highest to lowest response amplitude). Then the cells were added (in Fig. 4) or dropped (in Fig. 5) one by one first from the responsive cells and then from non-responsive cells.

In the analysis of robustness (Fig. 5c–f), first, z-scored reverse filter was computed in each neuron. A cluster of pixels whose absolute z-scores were more than 1.5 were defined as a representation area after smoothing their contours (e.g., red contours in Fig. 5c and Supplementary Fig. 4). If multiple areas were

obtained, the largest one was used. In each stimulus image, one responsive neuron was selected as a reference cell, and correlation coefficients of binarized representation areas were computed between the reference cell and other responsive cells to an image. Cells whose correlation coefficients were more than 0.4 were selected. A set of neurons including both the reference and the selected cells were called “overlapping cells”. To evaluate the effects of cell drop, cells were randomly removed from the overlapping cells, and reconstructed image was computed after each cell-drop. The reference cell was removed at first, and then other remaining overlapping cells were removed in each cell-drop sequence. The changes of reconstructed images were estimated by quantifying pixel values of a local part of the image. The local part of the image was determined as the reference cell’s representation area overlapped by at least one remaining overlapping cell (overlapping area in Fig. 5d and supplementary fig. 4). Absolute pixel values were averaged inside the local part of the image (Note that stimulus images were scaled from -1 to 1. See the section of **Data analysis**) and used for the evaluation of the local part of reconstructed image. Random drops of overlapping cells were repeated for 120 times, and the results were averaged across the random orders in each reference cell. All responsive cells were used once as the reference cell in each stimulus image. Data including at least 10 responsive cells and 5 overlapping cells were only used in this analysis.

In the cluster analysis (Fig. 7), the overlapping cells selected as described above were clustered by k means analysis with noise correlation of responses to an image for distance measure (predetermined number of clusters, k = 2, 3, 4, and 5 were used). We used the number of cluster (k) which showed the minimal between-cluster noise correlation in each overlapping neuron set (Supplementary Fig. 5a). In the analysis of correlation of trial-to-trial variability of reconstructed image between clusters (Fig. 7d and Supplementary Fig. 5e, i), trial-to-trial variability of the reconstructed image was evaluated by pixel values of the local part of the image as the analysis in Fig. 5e and f, and correlation coefficient of the pixel value change was computed between clusters. The local part of the reconstructed image was determined as described above. In the analysis of reliability of reconstructed image across trials (Supplementary Fig. 5f), correlations between single-trial reconstructed image and trial-averaged reconstructed image were computed and averaged across trials. The

main results were independent of the choice of cluster number (Supplementary Fig. 5g–i). Data including at least 10 responsive neurons and 5 overlapping neurons were only used in this analysis.

The feature values of each image were linearly regressed with the weights of image reconstruction model (cell-selection model) in all responsive cells in a local population to examine whether all the features of natural images were represented by the features of the responsive cells (Fig. 8d–f). The fitting error rate (% error) was computed in each image using a following equation, % error = $\sum(F_{fitted}-F_{image})^2/\sum(F_{image}-F_{mean})^2\times100$, where F_{fitted} is the set of fitted features, F_{image} is the set of features of the natural image, and F_{mean} is the mean of F_{image}.

Statistical analyses.

All data are presented as the median and 25–75Th percentiles unless indicated otherwise. The significant level was set to 0.05, with the exception of the criteria of significant visual response (0.01). When more than two groups were compared, the significant level was adjusted with the Bonferroni correction. Two-sided test was used in all analyses. The experiments were not performed in a blind manner. The sample sizes were not predetermined by any statistical methods, but are comparable to the sample size of other reports in the field.

Data availability.

The datasets of the current study and the code used to analyze them are available from the corresponding authors on reasonable request.

Legends

Figure 1. Sparse visual response to a natural image in mouse V1

a. Experimental schema. Natural image flashes were presented as visual stimuli, and the neuronal activities of single cells in the mouse V1 were recorded using two-photon Ca^{2+} imaging.

b. Trial-averaged time courses of visual responses to 10 natural images (image# 21–30 in a row) in 10 representative cells (cells# 1–10 in a column). Three lines for each response indicate the mean and the mean ± the standard error across trials. Black: significant responses, grey: non-significant response, red line: stimulus periods during which each image type was flashed for three times.

c. Plots of significant responses of all cells in a representative plane (n = 726 cells, upper left panel). Responses shaded by the red line in the upper left panel correspond to responses presented in (b). Percentage of responsive cells for each image (bottom) and percentage of images to which each cell responded (right) are shown as line graphs. Red lines (bottom and right) indicate median values. Cell numbers (cell #) were sorted by the percentage of images to which they responded, and image numbers (image #) were sorted by the percentages of cells that responded to each image in descending order. Single images activated relatively fewer neurons (bottom).

d. Examples of population response patterns to three images. (Left) Natural image stimuli and the spatial distributions of responsive cells in an imaging area (side length: 507 microns). The red filled and grey open circles indicate the highly responsive and remaining cells, respectively. (Right) Histograms of the visual responses of neurons in a local population. In the top panel, cells are divided into responsive (red bars) and the remaining groups (black bars) and are sorted by the response amplitude of each group to the natural image presented in the upper left panel (descending order). Visual responses to other images are plotted in the middle and bottom panels. The cell # order was fixed among the three histograms. Only small numbers of responsive neurons are duplicated among the three images.

e. Distribution of the amplitude of responses to single images. The cell # is sorted by the amplitude of the response to each image and averaged across images in a plane. After normalizing cell# (x-axis), data were collected across planes (n = 24 planes). The median (thick line) and 25–75th percentiles (thin lines) are shown. Small percentages of neurons exhibited higher response amplitudes.

f and **g**. Response rate to natural images. The percentages of cells responding to a single natural image (f) and to at least one image (g). Small percentage of cells responded to single natural image, whereas almost all cells

responded to at least one of images.

h and **i**. Response rate to moving gratings. Percentages of cells that responded to one direction (h) and to at least one direction (i) of moving grating.

j. Percentages of overlapping responsive cells between the natural images. Only a small percentage of cells exhibited overlapping significant responses between images, indicating that the cells responding to each image were distributed in a population.

k. Population sparseness.

f–k. Each dot indicates data from one plane, and the medians of 24 planes are shown as bars

Figure 2. Small overlap in the encoded visual features among cells in a local population

a. Scheme of encoding model for a single cell's visual response. The visual response is represented by weighted sum of the selected Gabor feature values obtained from a set of Gabor filters. The predicted visual response to ith image (R_i) is represented by the following equation, $R_i = f(\sum W_j \times F_{ji})$, where f is non-linear scaling function, W_j is the weight for jth Gabor feature, and F_{ji} is the feature value for the jth Gabor filter (Gj) obtained from ith image (Si). Gabor feature was selected based on the correlation between its feature values and visual response (Feature selection, see Methods).

b. and **c**. Examples of response predictions for two neurons. (Left panels) Blue and red lines indicate the observed and predicted responses, respectively. (Right panels) Weight parameters of the example neurons presented in the left panels. The weights of one of 10 models (each model corresponds to one of the 10-fold cross validations) are shown. The number of non-zero weights (i.e., number of used feature) is shown above the panels. Encoding filters (weighted sums of Gabor filters) are shown in the insets (red and blue indicate positive and negative values, respectively).

d. Comparison of response predicted by only the linear step (regression of Gabor feature values without non-linear (NL) scaling) and the observed response in the example neuron shown in Fig. 2c. Each dot indicates a response to one image. The red curve indicates NL scaling function curve (see Methods). The NL step

resulted in an enhancement of the sparse visual responses. The black line indicates y = x line.

e. NL scaling function curves across planes. Each grey curve was obtained by averaging the NL scaling curves across cells in each plane. Red curve indicates the averaged curve across planes (n = 24 planes). The black line indicates y = x line.

f. Upper left panel: raster plot of the weights in the plane illustrated in Fig. 1c (red: positive weight, blue: negative weight). The median values for the models of the 10-fold CVs are shown. Right panel: Percentage of features used for each cell. Bottom panel: Percentage of cells in which each feature was involved in the response prediction. The coloured bar under the x-axis indicates spatial frequency of the Gabor filter corresponding to each feature. Red lines in the bottom and right panels indicate median values. Only half of the Gabor features (624/1248 which have one of two phase sets) are shown for visibility, but the remaining features were included in the data shown in the right panel.

g. Participation rate of each feature in the response predictions for a population. Features were divided in terms of spatial frequency (SF) and averaged in each SF. Mean ± standard errors across all planes (n = 24 planes) are shown.

h. Distribution of percentages of features used in each cell (n = 12755 cells across 24 planes).

i. Distribution of percentages of features that overlapped between cells (n = 3993653 cell pairs across 24 planes).

j. Percentage of features used in at least one cell's response prediction.

Figure 3. Image reconstruction based on population activity

a. Scheme of image reconstruction model. Each Gabor feature value ($F_{ji,}$ i: image #, j: Gabor feature #) was independently linearly regressed (weights: H_{jn}, n: cell #) by multiple cell responses (R_{ni}) to ith image (Si) ($F_{ji} = \sum(H_{jn} \times R_{ni})$). Then, a set of reconstructed features (F_{1i}, F_{2i}, …., F_{ji}) are transformed into an image (S'_i). In the all-cells model, each feature value was reconstructed with all cells. In the cell-selection model, only cells that were selected by the encoding model were used to reconstruct each feature value (Cell selection, see Methods).

Thus, each feature value was reconstructed from different subsets of cells. The flow of the reconstruction model is represented by black arrows from the bottom to top.

b. Representative reconstructed images. Stimulus images (Stim. Image), images that were reconstructed using the all-cells (All-cells) and using the cell-selection model (Cell-selection) are shown. Each reconstructed image is trial-averaged. The reconstruction performance (pixel-to-pixel Pearson's correlation coefficients between the stimulus and reconstructed images, R) was computed in each trial and then averaged. The trial-averaged R is presented above each reconstructed image.

c. Distributions of Rs of the all-cell model (upper panel) and cell-selection model (bottom panel) in the representative plane shown in Figs. 1 and 2 (n = 200 images reconstructed using 726 cells from a plane).

d. R across planes. *: $p = 4.0\times10^{-4}$ using the signed-rank test (n = 24 planes). Reconstruction performance of the cell-selection model was comparable with that of the all-cells model.

Figure 4. Sparse responsive neurons sufficiently represent visual contents of single natural images.

a–c. Examples of images reconstructed from only the highly responsive cells (top panels), and reconstruction performances (R) plotted against the number of cells used for the reconstructions (bottom panels). In this analysis, using the parameters of cell-selection model, the number of cells used to reconstruct each image increased one by one. The cells were first collected from the responsive cells and then from the remaining cells. (Top panels), Stimulus image (1st panel) and reconstructed images (trial-average) from a subset of cells (2nd–4th panels). R and the number of cells used for the reconstruction are shown under the panel. (Bottom panels), Responsive cells (red dots) and the remaining cells (black dots) were separately sorted by response amplitude (descending order), and cells were added one by one from responsive cells with higher response amplitude to the remaining cells with lower response amplitude. The horizontal lines indicate 95% peak performances, and the numbers of cells for which the performance curves crossed the 95% level are indicated by the vertical lines. In each case, the responsive cells alone were sufficient to reconstruct the image nearly as well as all cells.

d. Averaged performance curve (n = 24 planes) plotted against cell number. The thick black line and grey lines indicate the means and the means ± standard errors, respectively. The horizontal line indicates the 95 % peak performance, and the vertical line indicates the first crossing point of the performance curve on the 95% line.

e. Contributions of the top 18 responsive cells to the image reconstruction shown in (a). (Top panels) Reverse filters (weighted sum of Gabor filters) multiplied by the visual responses reveal the spatial patterns of individual cell's contributions to the reconstructed image. The patterns vary among cells. (Bottom panels) Reconstructed image gradually changes by consecutively adding single cells.

f. Distributions of R for all cells (top, same as bottom of Fig. 3c) and the highly responsive cells (bottom) in the representative plane (726 cells). N = 200 images. Red vertical lines indicate medians.

g, R for all cells and only the responsive cells. *: $P = 4.0\times10^{-4}$ by signed- rank test (n = 24 planes). Each dot indicates data for each plane, and bars indicate median. The Rs were comparable between the two models.

h–j. Overlapping weights (i.e., features) between the cells highly responding to the same image.

h, Schema of the analysis.

i, Distribution of percentage of overlapping features in all cell pairs responding to the same image collected across planes.

j. The median of the percentages of overlapping features in the cell pairs responding to the same image. Each dot indicates the median in each plane (n = 24 planes) and the bar indicates median across planes. The percentages of overlapping features were still low even in the cell pairs responding to the same image.

Figure 5. Robust image representation by small numbers of responsive neurons

a. Examples of reconstructed images after single cell drop. Top: Stimulus and reconstructed image from all responsive neurons (57 cells). Middle: Reconstructed images after the single cell drop. Bottom: Reconstructed images from the dropped cells. Cell # is the same as in Fig. 4e.

b. Reduction of reconstruction performance after removing a single cell. Cell # on X axis were ordered from largest to smallest response amplitudes. Cell # was the same order as Fig. 4d. The median (horizontal line

inside a box), 25–75th percentiles (box), and 5–95th percentiles (whiskers) were obtained across planes (n = 24 planes).

c. Top: Reverse filters of overlapping cells (top panels). Representation area of each neuron was contoured by red line. Middle: Reconstructed images by the nine cells. Bottoms: Reconstructed images by single cells (upper panels) and reconstructed images after single cell drop (lower panels). Single cell drop had only small effect on the reconstructed images.

d. Examples of reconstructed images during sequential drop of the nine overlapping neurons. Red contours indicate overlapping area of the nine cells. Image around the overlapping areas gradually degraded after each cell drop.

e and **f**. Pixel value in a local part of reconstructed image (overlapping area) against number (or percentage) of dropped cells for the representative case shown in Fig. 5c (e) and for summary of all data (f). The intensity was obtained by averaging absolute pixel values within a local part of the image after each cell drop and compared to that from all overlapping cells. In each case, the reference cell was dropped at first, then cells were randomly selected and dropped from the remaining overlapping cells sequentially. In each stimulus, all responsive cells were used once as the reference cell. Data were collected and averaged across cells and across stimulus in each plane, and then collected across planes. Middle, lower and upper lines of a box, and whiskers indicate median, 25–75th and 5–95th percentiles across repetitions of random drop (n = 120 repetitions, e) or across planes (n = 24 planes, f), respectively.

Figure 6. Representation by multiple neurons is more reliable than that by single neuron

a. Trial-to-trial variability of the representative reconstructed image. Images were reconstructed by single-trial responses of all responsive neurons (57 cells) to the image (Stim. upper left).

b. Trial-to-trial variability of visual responses to the representative stimulus image. The nine neurons are the same as in Fig. 5c.

c. Trial-to-trial variability of reconstructed images by single neurons. Cell 1 is a reference cell in these

overlapping cells. Each image was reconstructed by a single trial response of a single cell. The images were variable across trials in many cells.

d. Trial-to-trial variability of reconstructed images by the nine cells (top) and their activity patterns in FOV (bottom). The reconstructed images were relatively reliable while activity patterns were variable across trials. (Bottom), One side length of FOV: 507 µm. Color code is same as in **b**.

Figure 7. Independent correlated activities between subsets of responsive neurons provide robust representation against trial-to-trial variability

a. Noise correlations between the overlapping neurons. These neurons formed three clusters (black squares). Cell order is same as in Fig. 6. The number of clusters (i.e., three in this case) which showed minimal between-cluster correlation was used (see Supplementary Fig. 5a).

b. Distribution of the number of clusters which showed minimal noise correlation between clusters.

c. Response correlation of within- (red) and between-cluster (blue) neuron pairs for all data (across reference cells, stimulus and planes). Between-cluster neuron pairs show almost-zero correlations on average, indicating independent activities between clusters.

d. Trial-variability of reconstructed images by the three clusters shown in (**a**). The three clusters represent the local part of the image in different trials, and integration of clusters provides robustness against trial-to-trial variability (Fig. 6d, also see Supplementary Fig. 5e, f). Correlation coefficients of trial-variability of the images were shown on the right side of the panels.

Figure 8. Visual features of natural images are distributed among most neurons in a population

a. Raster plots of highly responsive cells for each image in the representative plane shown in the previous figures (n = 711/726 responsive cells). The image # is sorted by the image reconstruction performance (descending order, right panel). In each line, cells that did not respond to the previously plotted images are added on the right side. As image # increased, the number of newly added cells decreased, and then, cell #

quickly reached a plateau level, indicating that many images are represented by the combination of cells that responded to other images. Thus, most images could be represented with some degree of accuracy by the combination of subsets of responsive cells of the population.

b. The numbers of responsive cells (black line) and numbers of newly added responsive cells (red line) are plotted against image # for the case shown in (a). Again, the number of newly added cells quickly decreased as the image # increased.

c. The numbers of responsive cells (black line) and numbers of newly added responsive cells (red line) are plotted against image #. N = 24 planes. Three lines in each colour indicate the mean and the mean ± the standard errors.

d, Schema of the analysis. The feature set of each natural image was linearly regressed by the weights from the cell-selection model of all the responsive cells in each plane, and the fitting error (% error, see Methods) was computed in each image. If the features encoded in all the responsive cells were sufficient to represent natural images, the weights of the responsive cells should work as basis functions to represent visual features of the natural images.

e. Distributions of the errors of all images collected across planes.

f. The median of % error across planes (bar, n = 24 planes). Each dot indicates the median in each plane.

Supplementary figure legends

Supplementary Figure 1. Gabor filter set and prediction performance of encoding model

a. Scheme of the process used to transform natural image features with Gabor filters. Each natural image was subjected to Gabor filters to obtain the corresponding Gabor feature values. Conversely, a set of Gabor feature values were transformed into an image by summing the Gabor filters after multiplying by the corresponding Gabor feature values.

b. Scheme of the Gabor filter set. Four orientations, 2 phases, 4 scales (or spatial frequencies) of Gabor filters were used. The 4 scales of Gabor filters (spatial frequency: 0.016, 0.033, 0.067, and 0.13 cpd) were positioned

on a 1 × 1, 3 × 3, 5 × 5, and 11 × 11 grids. A total of 1248 filters were used.

c. Effect of feature selection on the response prediction performance. The response prediction performances (correlation coefficients between the observed and predicted responses) are plotted against the threshold values of the feature selection for 30 representative cells. Pearson's correlation coefficient (corr. R) between each feature values and response was computed, and the features with corr. R greater than the pre-set threshold values were used for the regression analysis of the encoding model. The threshold value for the final model was selected from the pre-set values to maximize the response prediction.

d and **e**. Averaged prediction performance plotted against the feature selection threshold among all cells in the representative plane shown in the main figures (n = 726 cells from one plane, d) and among all planes (n = 24 planes, e). As the threshold increased, the prediction performance increased slightly and reached a plateau level. The threshold was independently adjusted for each cell such that the response prediction was maximized. Means ± standard errors are shown.

f–h. Results of the response predictions for the representative plane shown in the main figures (n = 726 cells from one plane). Distributions of the response prediction performances (n = 200 images, f), percentages of features used for each cell (n = 726 cells. g), and percentages of overlapping features between cells (n = 263175 pairs, h). In each cell pair, the number of overlapping features was divided by the number of features used in the encoding model for each cell and averaged between the two cells in a pair.

i. Distribution of the response prediction performance of all cells across all planes (n = 12755 cells from 24 planes).

j–l. Results of the response prediction performance across planes (n = 24 planes). Distributions of the response prediction performances (j), percentages of features used for each cell (k), and percentages of overlapping features between cells (l) are shown.

Supplementary Figure 2. Relationships between the weights of the encoding model and the RF structure

a–d. Relationship between the weights of the Gabor features and the RF structure.

a and **b**. (a) Schema of the analysis. In each cell, the RF structure was determined using a pseudo-inverse method (see Methods), and pixel-to-pixel Pearson's correlation coefficients between the RFs and Gabor filters were computed (R1 and R2 in (a); R (RF vs. Gabor)). (b) Then, Pearson's correlation coefficient (R) between the R (RF vs. Gabor) and weight values (W1 and W2 in (a)) was computed for each cell. In the example cell shown in (b), the R (RF vs. Gabor) and weight values were positively correlated (R = 0.83), indicating that Gabor filters similar to the RF were assigned high weight value.

c and **d**. Distributions of R between R(RF vs. Gabor) and weights for all cells in the representative plane (c) and for all planes (d).

e–i. Relationship between the spatial overlap of the encoding filter and percentage of overlapping features between cells.

e. Examples of the response prediction performances of the two representative cells. The black and coloured lines indicate the observed and predicted responses, respectively.

f. Weights and encoding filters of the two cells shown in (e). In these cells, the percentage of overlapping non-zero feature (% overlapping feature) was 12%, and the spatial similarity of the encoding filters (pixel-to-pixel correlation, R (filter)) was 0.26. The encoding filter was obtained by computing the sum of Gabor filters multiplied by the weights of the encoding model. The red and blue colours of the encoding filter indicate positive and negative values, respectively.

g. Relationship between the % overlapping feature and absolute value of the R (filter). These two values were positively correlated, indicating that the % overlapping feature reflects the spatial similarity of the encoding filters. All pairs in the representative plane are illustrated by color code.

h. Diverse structure of the encoding filters. The absolute R (filter) was relatively small for all pairs (median: 0.05), which indicated the diverse structure of the encoding filters between cells.

i. Distribution of the correlation coefficient between the % overlapping feature and the absolute value of R (filter) (|R(filter)|, N = 24 planes).

Supplementary Figure 3. Image reconstructions based on three models: all-cells, cell-selection (encoding-optimized) and image-optimized models

a. Relationship between the encoding model (cyan arrows, Fig. 2 in the main text) and the image reconstruction model (red arrows, Fig.3 in the main text). In the image reconstruction model, each Gabor feature value was independently linearly regressed by the visual responses of multiple neurons. In the "all-cells" model of image reconstruction, each feature value was reconstructed using all cells (not shown in the figure). In the "cell-selection" model of image reconstruction, each feature was reconstructed using a subset of cells (thin red arrows) selected by the encoding model (thin cyan arrows).

b–c. The image reconstruction performances (R) were compared across the three models, i.e., all-cells, cell-selection (encoding-optimized) and image-optimized model. The performances were almost comparable between the all-cells and cell-selection models (see the main text). The performance was only slightly improved when the cell selection for each Gabor feature was performed to maximize the reconstruction of the feature value ("image optimized" model in Supplementary Fig. 3b and c). In the image-optimized model, cells were first selected based on the correlation between the cell's response and the feature value, similar to the feature selection in the encoding model (see Methods). Then, each feature was linearly regressed only by the selected cells. Threshold of the correlation coefficient for cell selection was determined such that the reconstruction performance for each feature value was maximized.

b. Representative reconstructed images. Top panels: stimulus images; second row panels: reconstructed images from the all-cells model; third row panels: reconstructed images from the cell-selection model (also called "encoding-optimized" model hereafter in the legends); fourth row panels: reconstructed images from the image-optimized models; bottom panels: reconstructed images using overlap of the cells for each feature (overlap) between the encoding-optimized and image-optimized models. In the overlap model, each feature value was reconstructed using the cells employed in both the encoding-optimized and image-optimized models.

c. Comparison of R. Mean ± standard errors (n = 24 planes) are shown. Each dot indicates the average of all

images in each plane. The image-optimized model exhibited an only slightly better R than the other models. P = 2.4×10^{-3} (all-cells vs. encoding-optimized), 1.1×10^{-4} (all-cells vs. image-optimized), 2.1×10^{-4} (all-cells vs. overlap), 1.1×10^{-4} (encoding-optimized vs. image-optimized), 1.1×10^{-4} (encoding-optimized vs. overlap), and 1.1×10^{-4} (image-optimized vs. overlap), using signed rank test with the Bonferroni correction.

d. Percentages of cells used for each feature are plotted against the spatial frequencies of Gabor filters. The image-optimized model (blue) used more cells to reconstruct each feature value than the encoding-optimized model (i.e., cell-selection model, red). The plot of the encoding-optimized model is the same as the plot shown in Fig.2g in the main text.

e. Overlap of the cells between the image-optimized and encoding-optimized models. In the reconstruction of each feature value, the cells used in the encoding-optimized model were almost included in the cells used in the image-optimized model. Image-only: percentage of cells used only in the image-optimized model. Encoding-only: percentage of cells used only in the encoding-optimized model. Overlap: percentage of cells used in both models. The results in (c–e) indicate that the overlapping cells mainly contributed to the image reconstruction and the image-only cells contributed to the slight increase in R of the image-optimized model.

f–h. Comparison of reverse filters among the overlap, image-only and encoding-only cells.

f. Examples of reverse filters (reconstruction filters, i.e., the sums of Gabor filters multiplied by the weights in each cell) in the overlap cells (upper panels), image-only cells (middle), and encoding-only cells (lower) used to reconstruct the feature whose Gabor filter is shown in the left panel (reconstruction target). The centre panels present reverse filters of single cells, and the right-most column presents averaged reverse filters among the cells. The reverse filters of the overlap cells appear to be similar to the Gabor filter of the reconstructed feature. In the image-only cells, the reverse filters of the individual cells do not appear to be similar to the Gabor filter of reconstructed feature, whereas the average of reverse filters resembles the Gabor filters. This suggests that the overlapping cells represent the reconstructed feature at the single-cell level, whereas the image-only cells represent the feature as a population.

g. Correlation between the reverse filters of single cells and the Gabor filters of the reconstructed feature. At

the single-cell level, the reverse filters of the overlapping cells (black line) are more similar to the Gabor filters of reconstructed features than the reverse filters of the image-only and encoding-only cells. The similarity of the reverse filter to the Gabor filter of reconstructed feature was indistinguishable between the image-only and encoding-only cells (red and blue lines). Mean ± s.e. are shown. P = 1, 1, 0.087 and 0.080 (0.016–0.13 cpd) for encoding-only vs. image-only; 2.4×10^{-3}, 2.2×10^{-4}, 2.2×10^{-4}, and 2.2×10^{-4} for encoding-only vs. overlap; 2.2×10^{-4}, 2.2×10^{-4}, 2.2×10^{-4}, and 2.2×10^{-4} for image-only vs. overlap; using signed rank test with the Bonferroni correction.

h. Correlation between the averaged reverse filters and Gabor filters of the reconstructed feature. As a population, the similarities of the average reverse filters of the image-only cells (blue) to the Gabor filters are slightly less than the similarity of the overlapping cells (black) but greater than the encoding-only cells (red). Based on these results, the image-only cells were responsible for the slight increase in R by representing the reconstructed feature as a population. Mean ± s.e. are shown. P = 2.4×10^{-3}, 2.2×10^{-4}, 2.5×10^{-4}, and 2.5×10^{-4} (0.016–0.13 cpd) for encoding-only vs. image-only; 2.4×10^{-3}, 2.2×10^{-4}, 2.2×10^{-4}, and 2.2×10^{-4} for encoding-only vs. overlap; 4.1×10^{-4}, 2.2×10^{-4}, 2.2×10^{-4}, and 2.5×10^{-4} for image-only vs. overlap; using signed rank test with the Bonferroni correction.

In summary, the cells used in both the encoding- and image-optimized models (i.e., overlapping cells) likely represented the reconstructed feature at the single-cell level, whereas the additional cells that were used only in the image-optimized model (i.e., image-only cells) probably represented the reconstructed feature as a population. Thus, the main information about visual contents was obtained from the encoding-optimized model (i.e., cell-selection model), and additional information was extracted when more cells that were not captured by the cell-selection using encoding model were integrated into the reconstruction model.

Supplementary Figure 4. Spatial overlap of reverse filters among responsive neurons

a. Spatial overlap of reverse filters. Reverse filter was transformed to z-score, and area in which absolute z-score was greater than 1.5 was defined as a significant area (red contours in left panels). The significant area

was transformed to a binary image, and a pixel-to-pixel correlation coefficient of the binary image was computed between two cells. Cell pair whose correlation coefficient was greater than 0.4 was defined as spatially overlapping cell pair.

b. Distribution of the number of spatially overlapping cells.

Supplementary Figure 5. Independent activities between subsets of responsive neurons provide robust representation against trial-to-trial response variability.

a. Noise correlations in between-cluster cell pairs in the representative neuron set (reference and overlapping cells) shown in Fig. 5c¸ 6b, 6c and 7a. Cluster number which showed minimal noise correlation between clusters was used in the analyses of main figures.

b. Correlation of reverse filters between cells in within- (red) and between-cluster pairs (blue). Distributions highly overlapped between the two pairs, although median correlation coefficient of the within-cluster pair was higher than those of the between-cluster pair (0.23[0.05–0.43], n=620990 in within-cluster pairs, and 0.19[0.02–0.39], n=809833, in between-cluster pairs, $p < 1.0\times10^{-9}$ by signed rank test).

c. Comparisons of reconstructed images by the clusters shown in Fig 7a. Similarity of reconstructed images between clusters was estimated by pixel-to-pixel Pearson's correlation coefficient. Trial-averaged images are shown.

d. Distribution of reconstructed image correlation between clusters. Positive correlation of the distribution indicates that clusters represent similar image patterns. Median [25–75th]: 0.33 [0.11–0.52], n=57420 cluster pairs.

e. Correlation of across-trial local image intensity change between clusters. The local image intensity was obtained by averaging absolute pixel values inside the overlapping area (red contours in Fig. 5d in the main figure). As predicted by response correlations (Fig. 7c), the distribution shows almost-zero correlation on average (-0.08[-0.24–0.09], n=57420 cluster pairs), indicating independent representations between clusters.

f. Across-trial reliability of reconstructed by individual clusters (blue) and all cells in a set of overlapping cells

(red). The reliability was estimated by computing correlation coefficient between single-trial and trial-averaged reconstructed images and averaging across trials. P = 1.8×10^{-5} by signed rank test (n = 24 planes). Middle, lower and upper lines of a box, and whiskers indicate median, 25–75th and 5–95th percentiles across planes (n = 24 planes).

g-i. The main results (Fig. 7 and Supplementary Fig. 5) are independent of the number of clusters. (g) Noise correlation of within-cluster pairs and between-cluster pairs (blue), corresponding to Fig. 7c. Almost zero-correlation in between-cluster neuron pairs were observed in all the number of clusters. (h) Correlation of reconstructed image between clusters corresponding to supplementary fig. 5c and d. Positive correlations were observed in all the number of clusters. (i) Correlation of across-trial image intensity change between clusters, corresponding to supplementary fig. 5e. Image intensity was computed by averaging absolute pixel values inside the overlapping area. Almost zero correlation between clusters were observed in all the number of clusters. (g–i) Middle, lower and upper lines of a box, and whiskers indicate median, 25–75th and 5–95th percentiles across all samples. (g) n = 739579, 1001691, 1134950, 1218522 in between-cluster (blue), and n = 691244, 429132, 295873, 212301 in within cluster pairs (red), for cluster number 2, 3, 4, 5 respectively. (h, i) n = 30573, 91719, 183438, 305730 cluster pairs for cluster number 2, 3, 4, 5, respectively.

Supplementary Figure 6. Neurons were spatially intermingled between clusters.

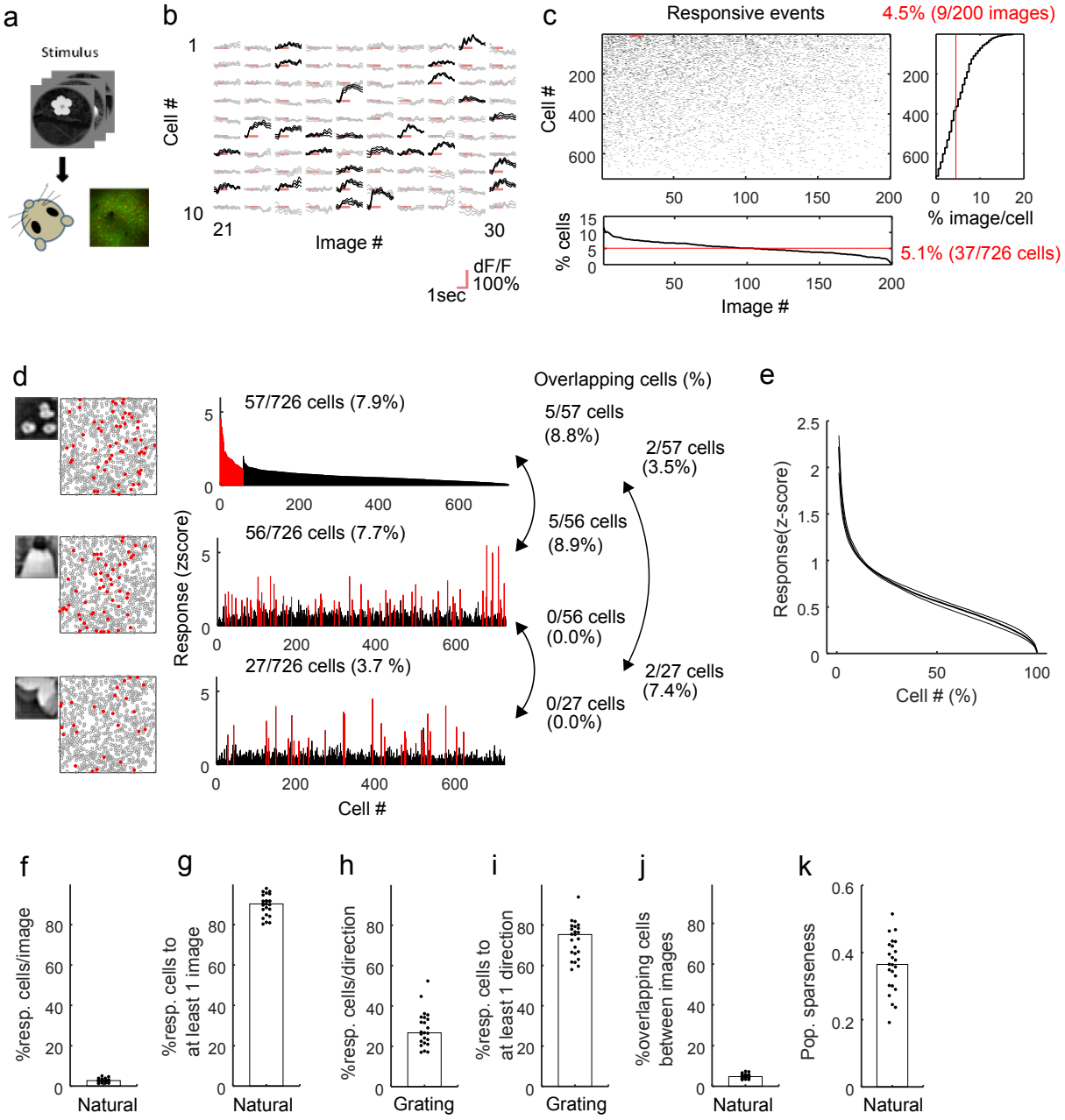
a–c. Representation areas of a set of overlapping cells (top) and cell's location in the imaging field (bottom). Top. Contours of representation areas are drawn on a stimulus image by colour lines (each colour indicates each cluster). Bottom. Grey dots indicate locations of recorded neurons, and coloured lines connect cells in one cluster. Different colors indicate different clusters. Three sets of overlapping cells are presented in a–c.

d–f. Same as (a–c) except for the stimulus image and reference cells. Scales of figures are same in a–f. Scale bar indicates 100 microns.

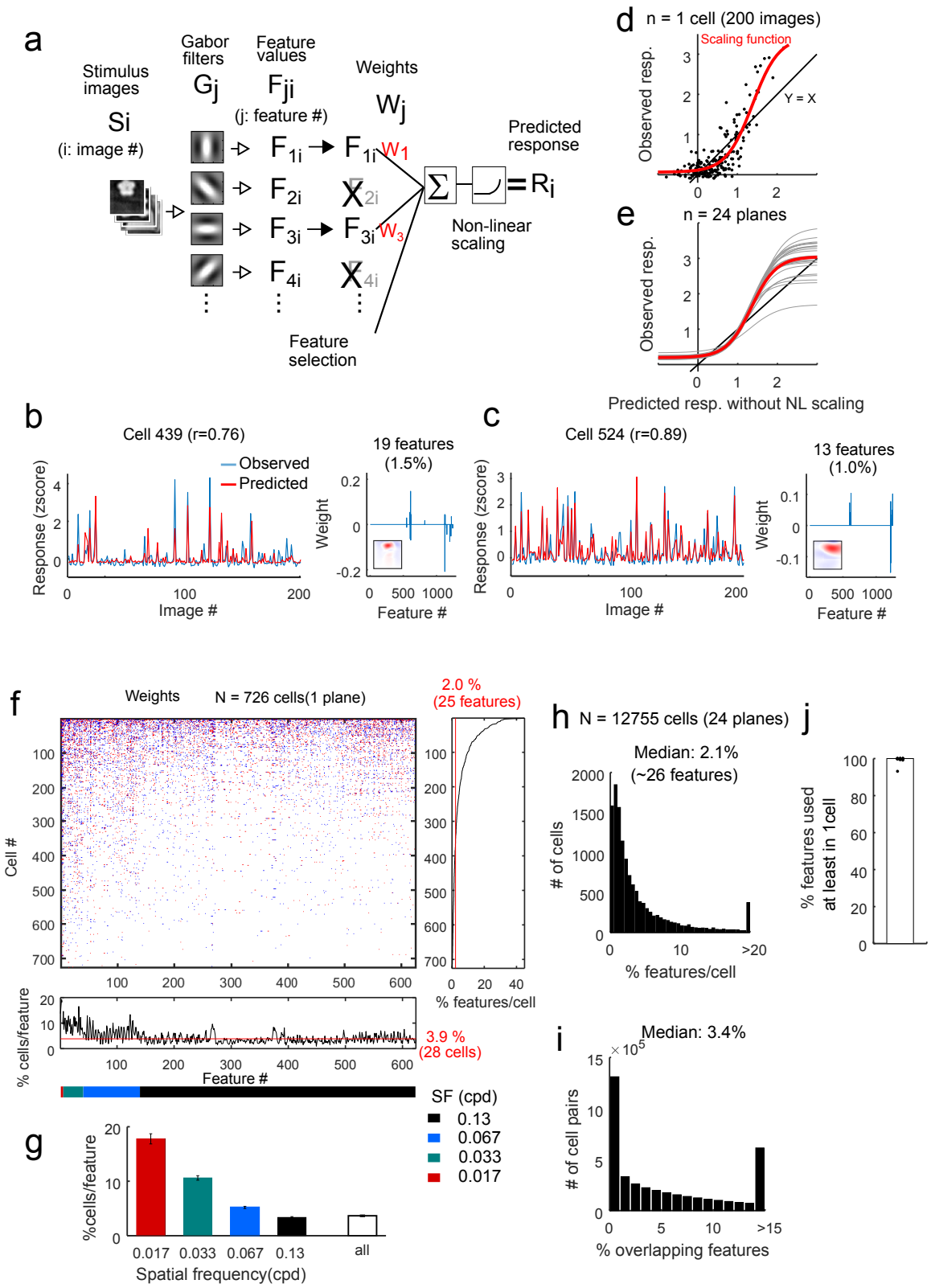
g. Distribution of cell-cell distance of within-cluster pair (red) and between-cluster pair (blue). Distributions almost overlap, indicating that cells are spatially intermingled between clusters. Thus, cell potions do not

simply explain the structure of noise correlations. Median [25–75th percentiles]: 165 [103–242] microns, n = 809833 for between-cluster cell pairs and 162 [100–238] microns, n = 620990 for within-cluster cell pairs.

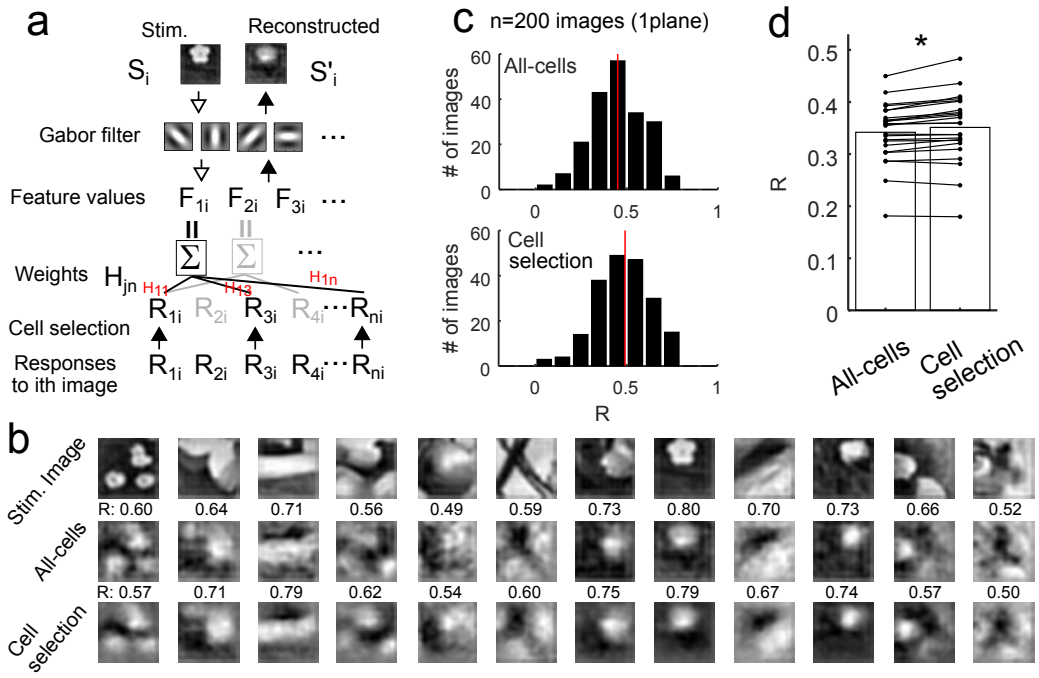
Yoshida and Ohki, Figure 1



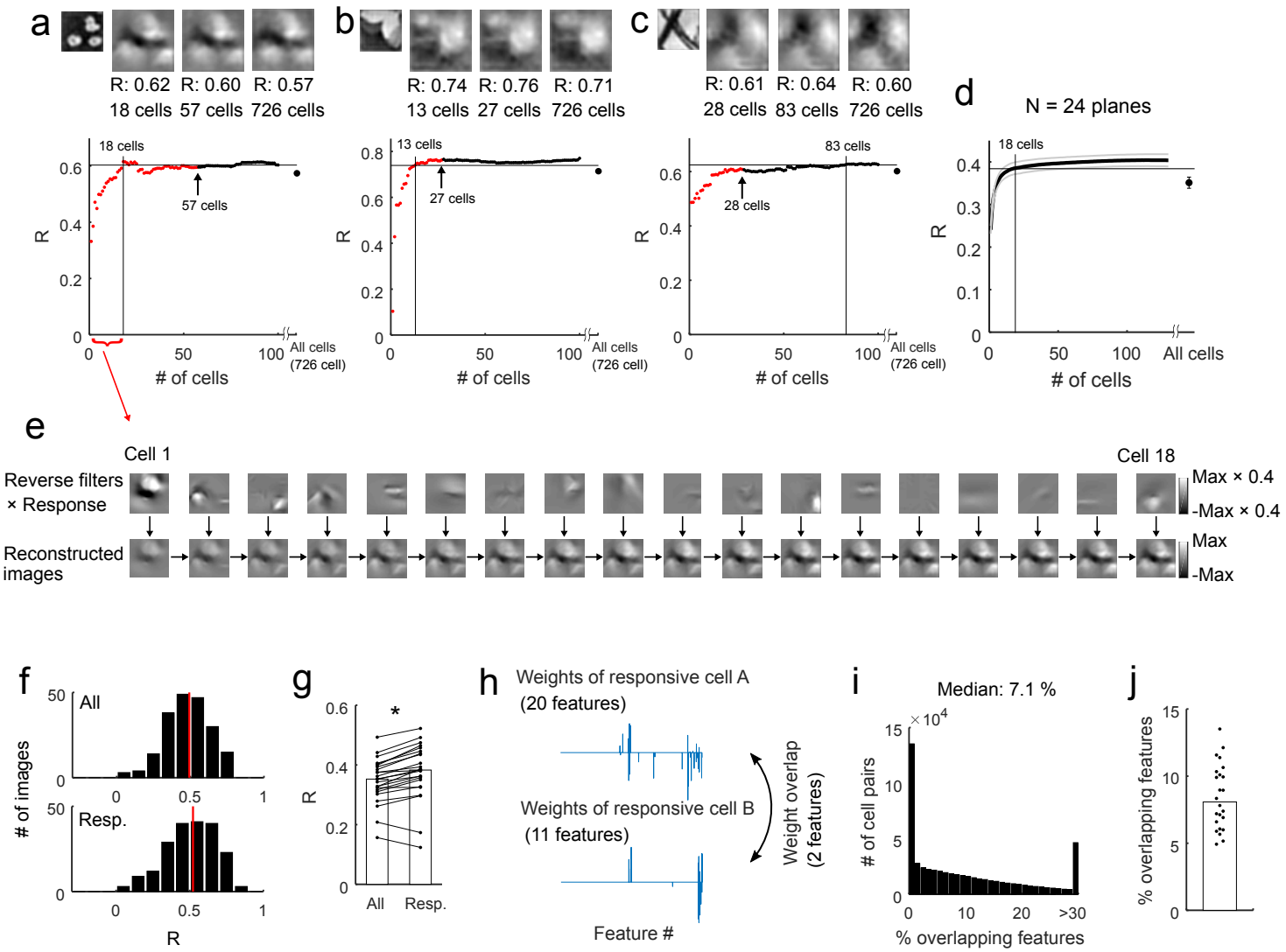
Yoshida and Ohki, Figure 2



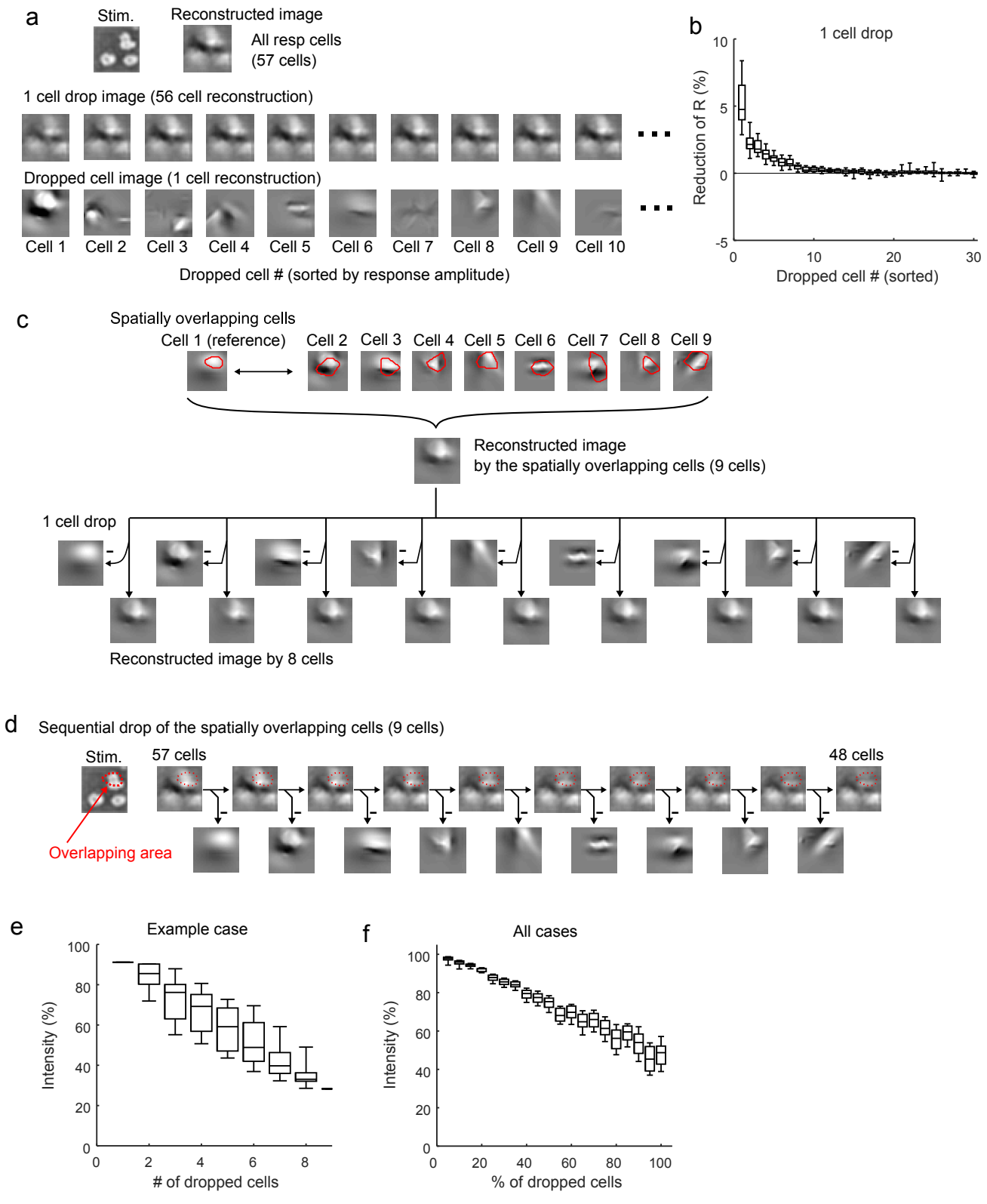
Yoshida and Ohki, Figure 3



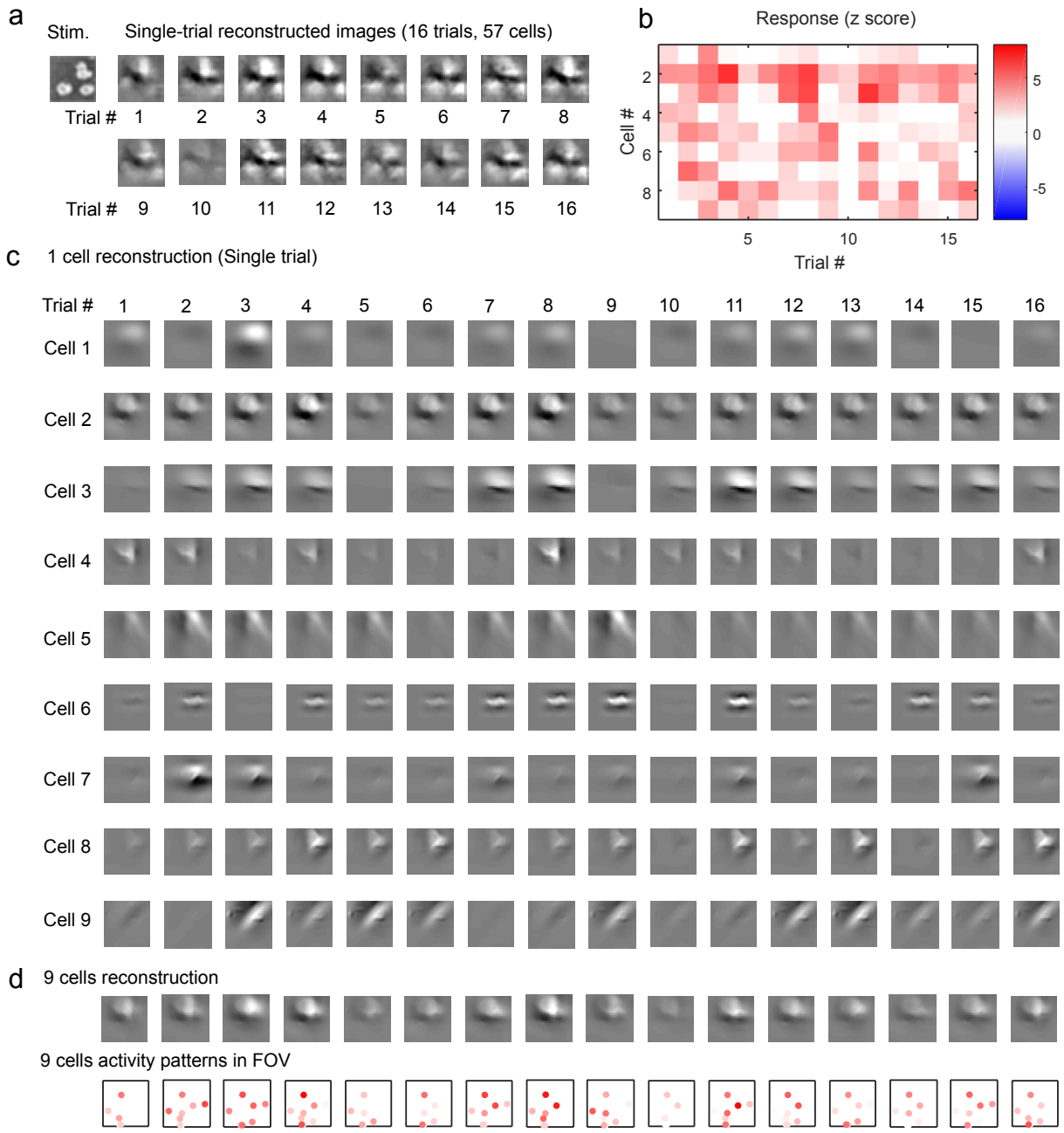
Yoshida and Ohki, Figure 4



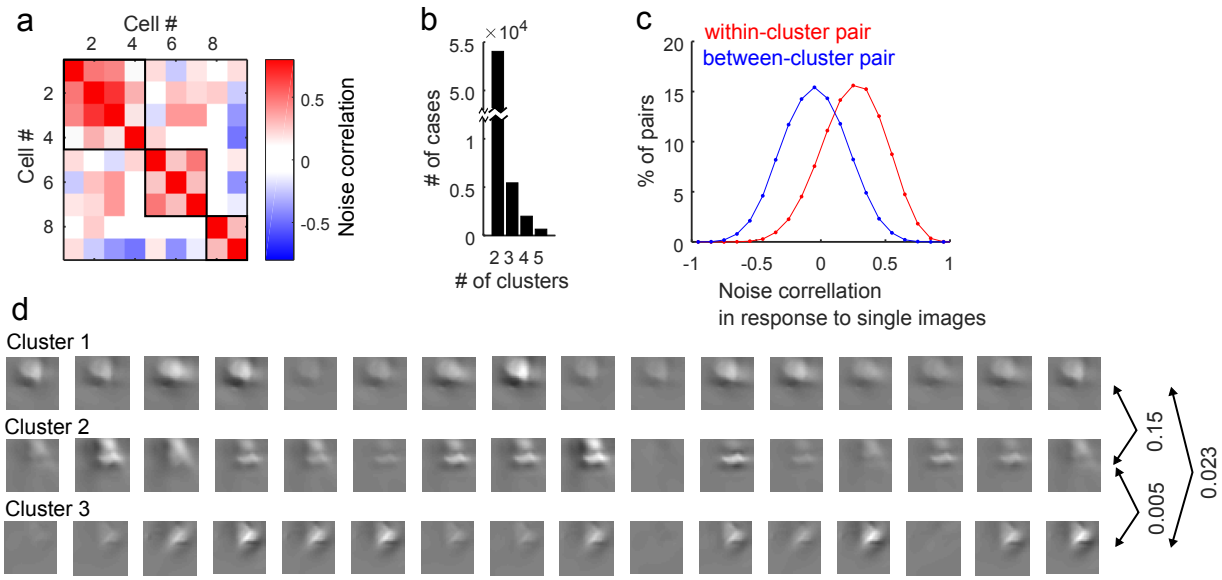
Yoshida and Ohki, Figure 5



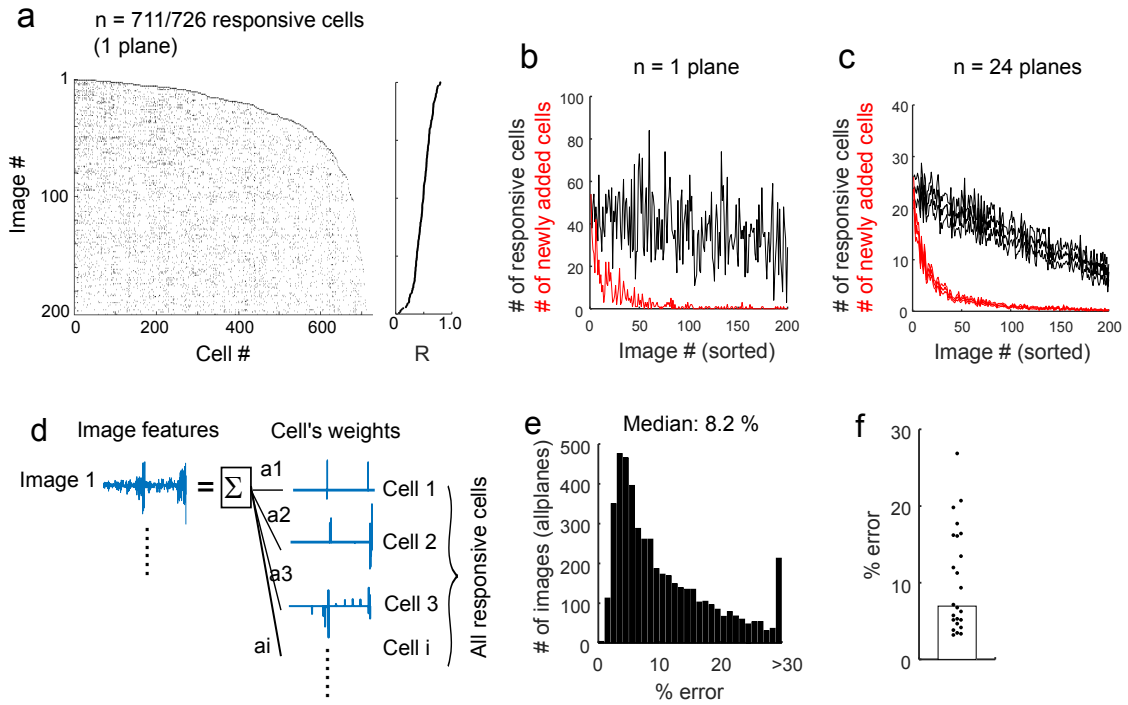
Yoshida and Ohki, Figure 6



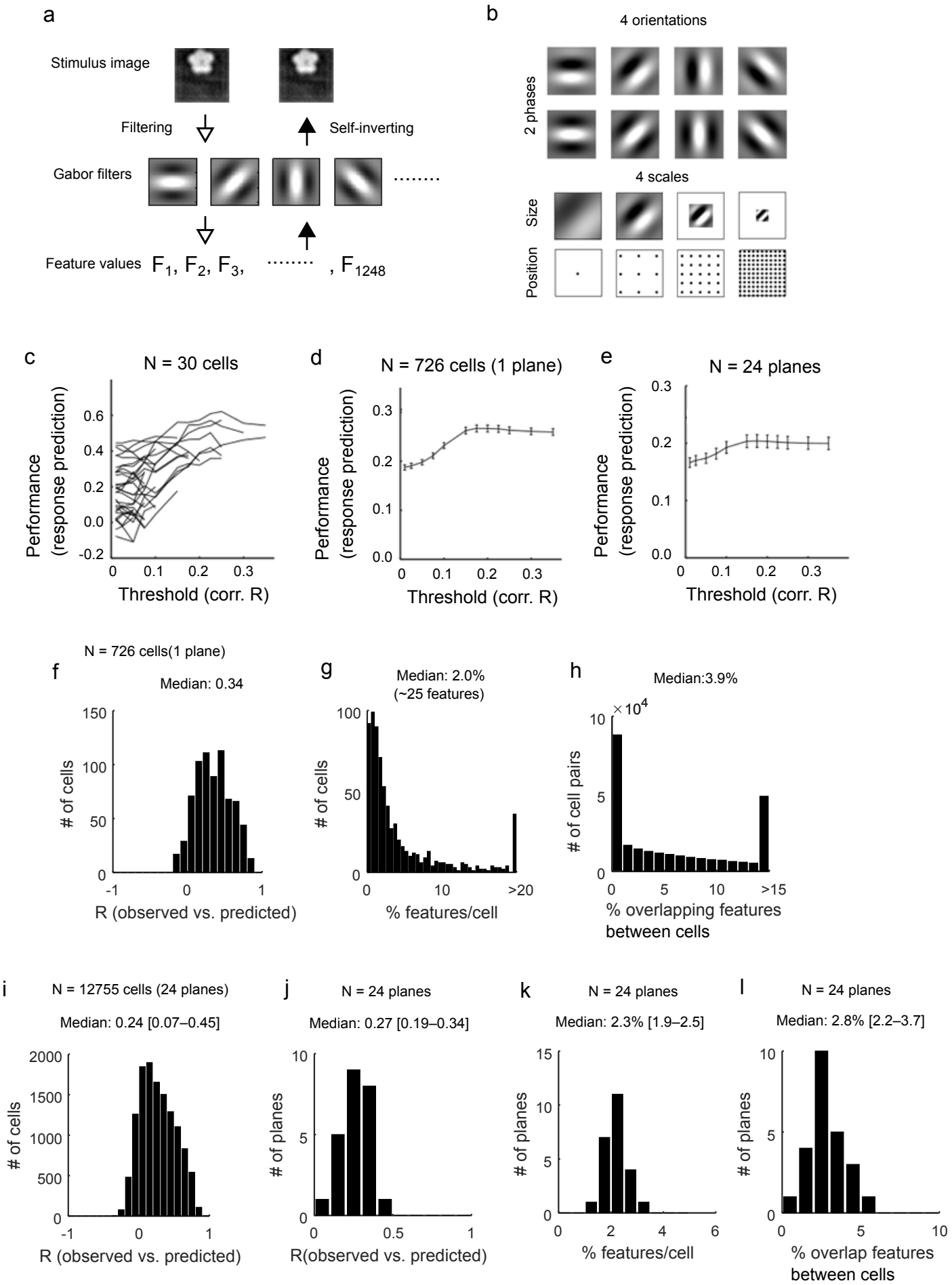
Yoshida and Ohki, Fgiure 7



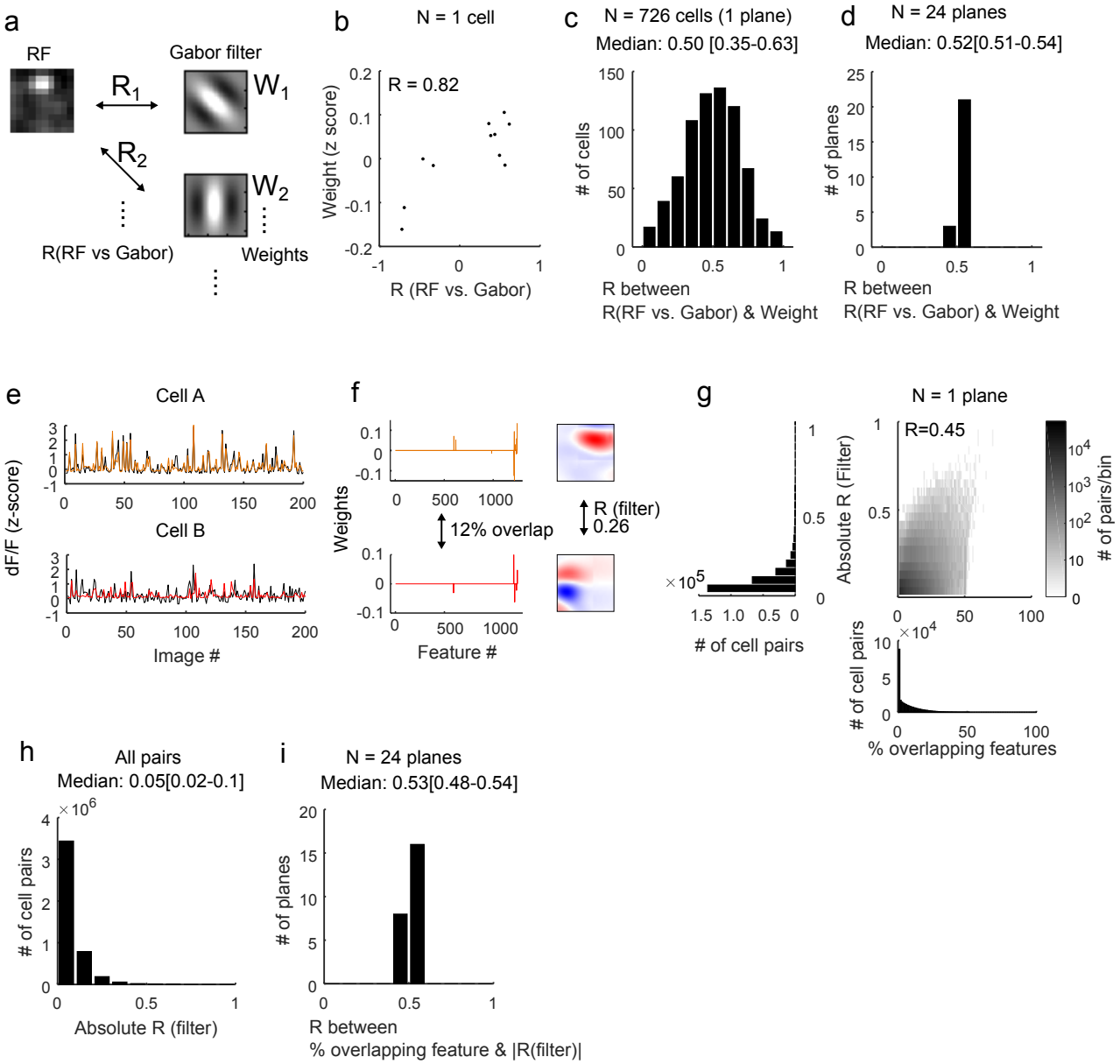
Yoshida and Ohki, Figure 8



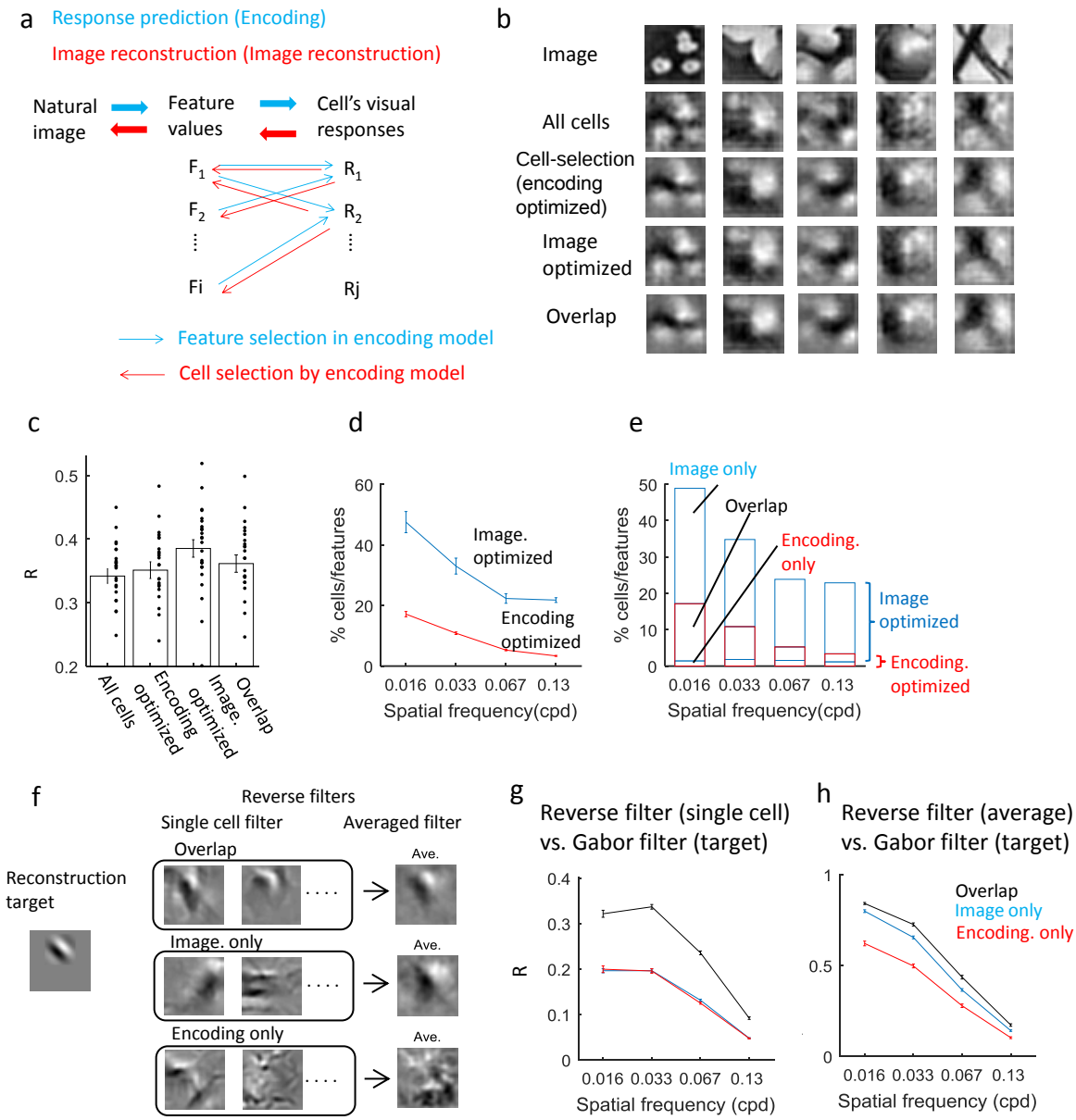
Yoshida and Ohki, Supplementary Figure 1



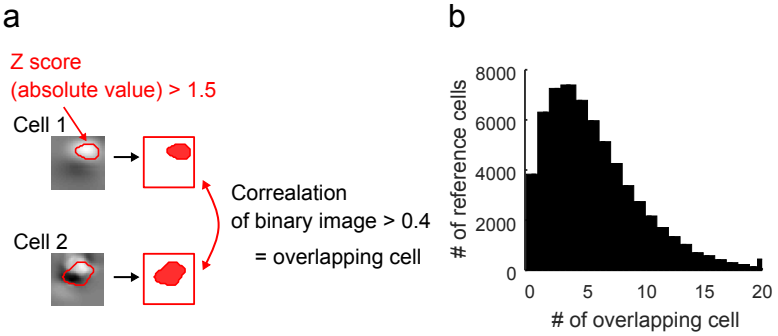
Yoshida and Ohki, Supplementary Figure 2



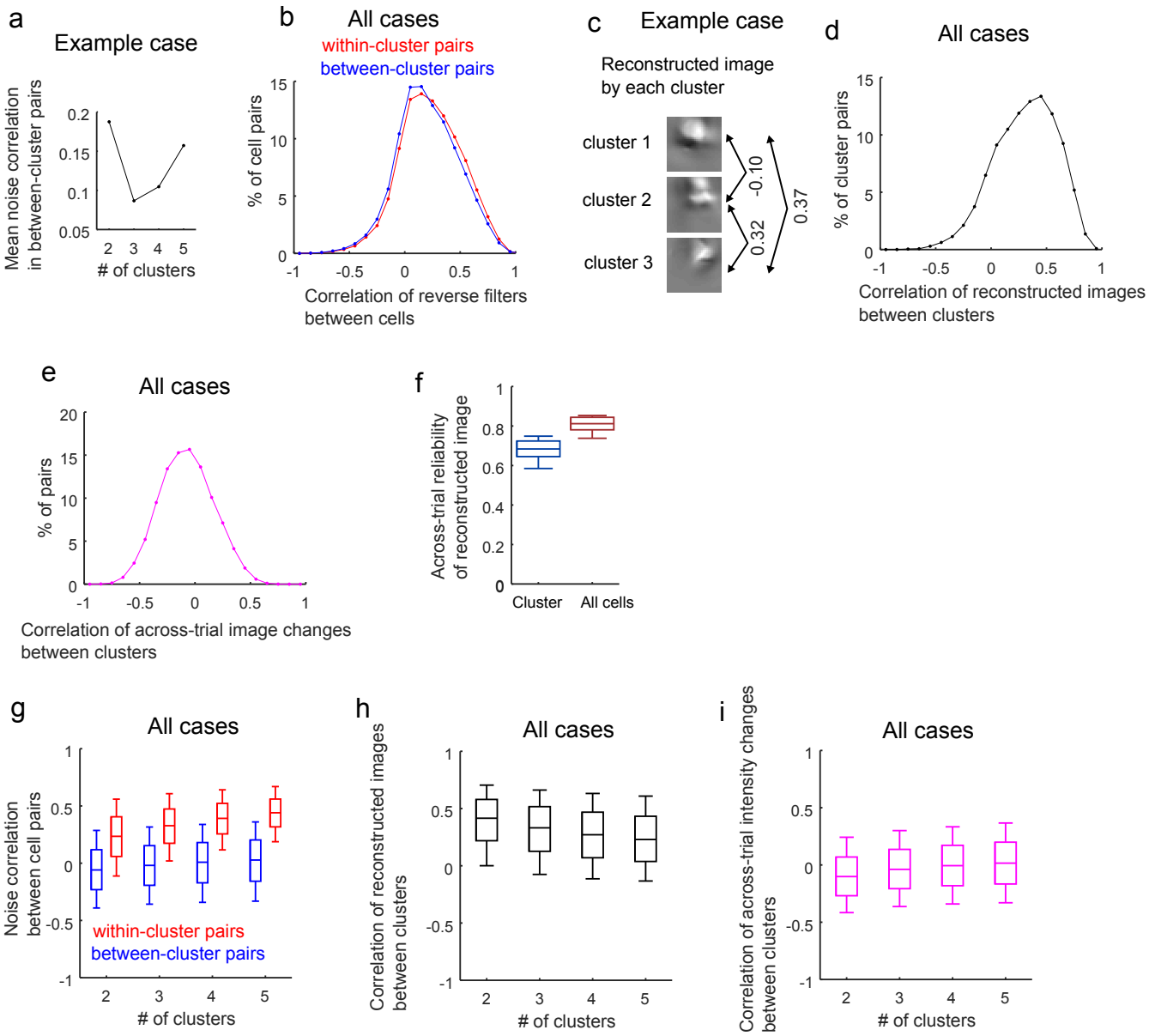
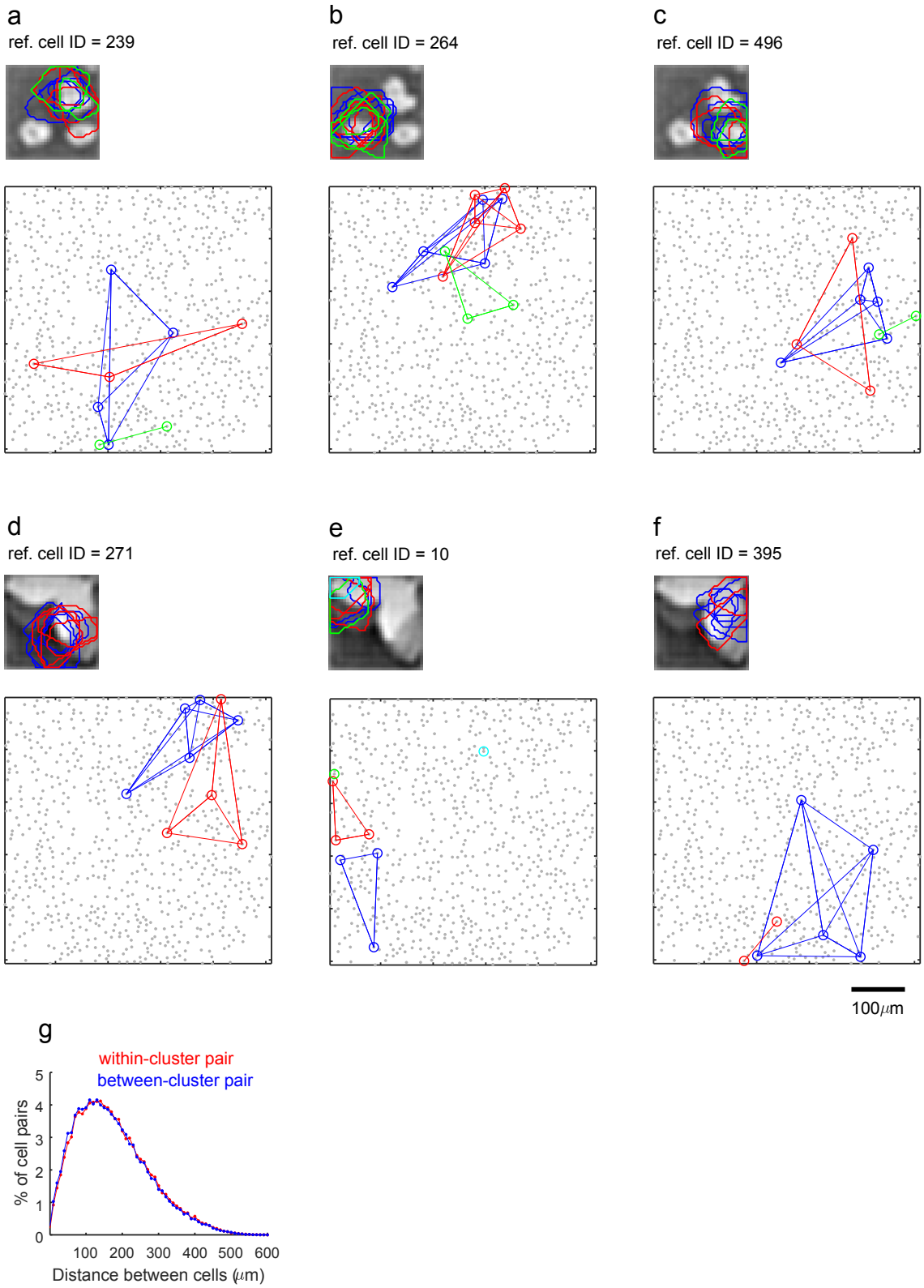
Yoshida and Ohki, Supplementary figure 3



Yoshida and Ohki, Supplementary figure 4



Yoshida and Ohki, Supplementary figure 5



Yoshida and Ohki, Supplementary figure 6