Panoramic visual statistics shape retina-wide organization of receptive fields

Divyansh Gupta*, Wiktor Młynarski*, Olga Symonova, Jan Svatoň and Maximilian Joesch#

Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria

* These authors contributed equally # Correspondence: maxjosch@ist.ac.at

Abstract

Visual systems have adapted to the structure of natural stimuli. In the retina, center-surround receptive fields (RFs) of retinal ganglion cells (RGCs) appear to efficiently encode natural sensory signals [1–4]. Conventionally, it has been assumed that natural scenes are isotropic and homogeneous; thus, the RF properties are expected to be uniform across the visual field. However, natural scene statistics such as luminance and contrast are not uniform and vary significantly across elevation [5,6]. Here, by combining theory and novel experimental approaches, we demonstrate that this inhomogeneity is exploited by RGC RFs across the entire retina to increase the coding efficiency. We formulated three predictions derived from the efficient coding theory: (i) optimal RFs should strengthen their surround from the dimmer ground to the brighter sky, (ii) RFs should simultaneously decrease their center size and (iii) RFs centered at the horizon should have a marked surround asymmetry due to a stark contrast drop-off. To test these predictions, we developed a new method to image high-resolution RFs of thousands of RGCs in individual retinas. We found that the RF properties match theoretical predictions, and consistently change their shape from dorsal to the ventral retina, with a distinct shift in the RF surround at the horizon. These effects are observed across RGC subtypes, which were thought to represent visual space homogeneously, indicating that functional retinal streams share common adaptations to visual scenes. Our work shows that RFs of mouse RGCs exploit the non-uniform, panoramic structure of natural scenes at a previously unappreciated scale, to increase coding efficiency.

Introduction

The idea that sensory neurons exploit the statistical structure of natural stimuli to minimize the metabolic cost of information transmission has been a guiding principle in neuroscience for over half a century [7,8,9]. This conceptual framework, known as the efficient coding hypothesis [10], has provided successful theoretical accounts of sensory coding across species and sensory systems [11–14] with the retina being the paramount example [15]. Most of the work in the retina has been focused on retinal ganglion cells (RGCs), the neurons that relay visual information from the eye to the brain. It has been demonstrated that multiple properties of RGCs: the shape of receptive fields [1–4], organization of RF mosaics [4,16,17] and the ratio of ON to OFF RGC cell-types [18], can be explained as adaptations to the natural sensory environment. In all of the mentioned cases, *ab initio* theoretical predictions about efficient encoding of natural scenes have led to a better understanding of the physiological and anatomical properties of the retina.

One way a sensory neuron could implement an efficient code is by removing predictable (or redundant) components from sensory stimuli, in a transformation known as predictive coding. This prominent hypothesis suggests that the center-surround structure of RGC RFs is a manifestation of such design principle [1]. According to this hypothesis, the surround computes a prediction of the stimulus value in the center of the RF. The predicted value is then "subtracted" from the center through inhibition, which dramatically reduces the amount of neural resources used to convey the stimulus downstream. Predictive coding and related information-theoretic principles [2–4,19] typically assume that the structure of natural scenes is uniform across the visual field. However, as demonstrated recently, local contrast and luminance vary prominently across the elevation within the natural visual field of a mouse [5,6]. Such systematic variation affects the signal-to-noise ratio (SNR) of the input to RGCs. To understand how this inhomogeneous noise structure should shape RGC RFs, we developed a simple, predictive-coding model. When adapted to natural statistics of mouse vision, our model generates three key predictions linking the shape of optimal RFs and their position within the visual field. First, the relative surround strength should increase with increasing elevation, due to a consistent decrease in brightness between the bright sky and the dim ground. Second, the center size should decrease along the same axis. Third, due to a rapid change of signal intensity between lower and upper fields of view, RFs centered on the horizon should have asymmetric surrounds, with the upper half being stronger than the bottom one.

To test these predictions experimentally, we established a novel system that enables recording and characterization of the RF structure at high resolution, at the scale of thousands of RGCs in a single retina. Such technological development enabled us to collect a dataset of 26558 RGC RFs covering the entire central retina, which was crucial to test our theory. We found a close agreement between theoretically optimal RF architecture and the variation of RF shapes across the retina, suggesting that RGCs exploit global asymmetries of natural scenes for maximizing coding efficiency. Finally, we explored these adaptations across the diversity of functional RGCs types [20]; each thought to share the same physiology, morphology, intra-retinal connectivity [21–23]. We identified a systematic dorsoventral variation of RF shape, regardless of the functional type. Our results thus indicate that even RGC-types, which are frequently interpreted as independent and parallel streams [20,24], share similar adaptations to the structure of the panoramic visual field.

Efficient coding predicts organization of RF shapes across the visual field

To understand how the statistical structure of natural scenes shapes receptive fields (RFs) across the visual field, we developed a model of sensory coding in retinal ganglion cells (RGCs; Fig. 1a). Our approach is closely related to the predictive coding (PC) theory, which postulates that RGCs recode outputs of photoreceptor cells to minimize the metabolic cost of sensory information transmission [1]. Following this theory, we modeled neural responses as a linear combination of the receptive field and natural stimuli, distorted by different sources of constant noise, e.g., biochemical or synaptic [15,25,26] (Fig. 1a, right panel). The computation performed by such RF can be understood as a difference of the weighted center of the stimulus and its surrounding

neighborhood. Our model generates predictions consistent with PC (Suppl. Fig.1) as well as related theories of efficient sensory coding [2,4].

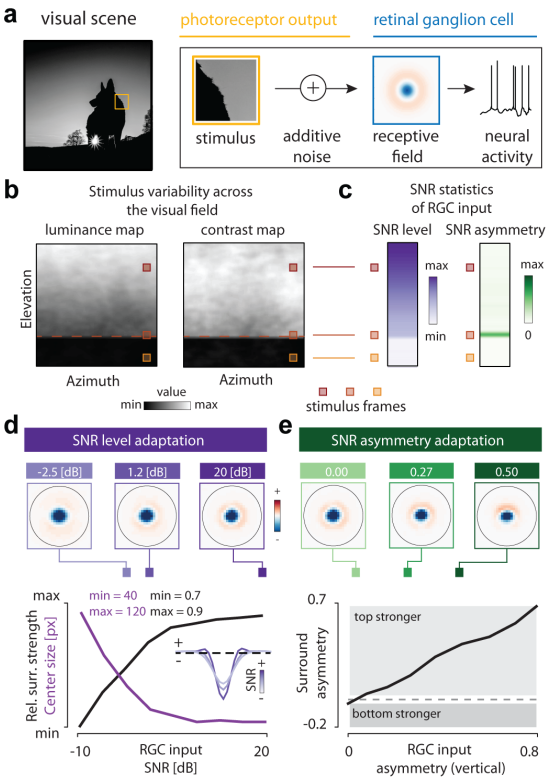


Figure 1. Predictive coding and natural scene statistics. a. *Linear model of a receptive ganglion cell (natural scene photograph - courtesy of Anton Sumser).* ***b.*** *Average luminance (left panel) and local contrasts (right panel) in the mouse field of view in the ultraviolet range. Orange dashed line denotes the horizon. (natural image data - courtesy of Hiroki Asari [5].* ***c.*** *Vertical profiles of photoreceptor SNR level (left panel, purple colors) and vertical photoreceptor SNR asymmetry index (right panel, green colors). Colorful squares correspond to positions along the vertical axis of the visual field depicted in (**b**).* ***d.*** *Relative surround strength (black line) and center size (purple line) of optimal predictive coding RFs either increase or decrease, respectively, with increasing signal-to-noise ratio of the photoreceptor output (SNR). Receptive fields in **d** & **e** were smoothed with a 2 x 2-pixel window for display purposes.* ***e.*** *Surround asymmetry of optimal predictive coding RFs increases with increasing vertical SNR asymmetry of photoreceptor output (solid black line)*

Natural scenes are spatially inhomogeneous. To understand this inhomogeneity, we examined a set of natural images collected specifically to study mouse vision [5]. In agreement with previous studies [5,6], we found that local luminance and contrast decrease gradually with the elevation, and drop-off suddenly close to the horizon line (Fig. 1b). Under the assumption of constant noise level, the SNR of RGC inputs should therefore follow the analogous pattern (Fig. 1c, left panel). Moreover, due to an abrupt change of luminance and contrast, stimuli centered at the horizon yield a highly non-uniform SNR pattern (Fig. 1c, right panel). Taking this into consideration, we numerically optimized RFs to minimize the strength of neural responses averaged across a set of

natural image stimuli (see Methods). The shape of the optimal RF depends on the relative strength and structure of noise.

When the SNR decreases, the center of the optimal RF broadens, and the surround becomes more diffused (Fig. 1d). This qualitative change is manifested in decreasing ratios of relative surround-to-center strength (Fig. 1d, black line) and center sizes (Fig. 1d, purple line). The optimal RF shape is additionally modulated by the spatial pattern of SNR (Fig. 1e). When the SNR is spatially non-uniform (e.g., when the signal is stronger in the upper half of the stimulus), the optimal RF becomes asymmetric (Fig. 1e). This effect is particularly visible in the increasing asymmetry of the surround as a function of SNR uniformity (Fig. 1e, bottom panel). Because of such systematic variation of contrast and luminance, the PC model predicts three links between the position of a neuron in the retina and the shape of its RF. First, the strength of the RF surround relative to the center should be increasing with elevation across the visual field. Second, the size of the center should increase in the opposite direction. Third, RFs located at the horizon should have surrounds that are stronger in the upper, than in the lower half. Such distribution of RF shapes would indicate that RGCs exploit global statistics of the visual field to maximize the efficiency of sensory coding. Our predictions about the dependence of the RF shape on the overall SNR level and on the asymmetry of the SNR pattern are considered independently throughout the remainder of the paper.

Large-scale characterization of receptive fields across the retina

Testing these theoretical predictions requires a high-resolution characterization of RGC RFs from extended regions of the retinal surface. Currently, however, it is not practical to perform such large-scale characterizations with any of the existing methods. Multiphoton imaging approaches can measure large numbers of RGCs [20,27], but only at a moderate throughput (~150 RGCs at a time [20,27]). Multi-electrode array recording approaches have improved this number but are limited by the recording area that is placed on top of the electrode array [17]. Moreover, RF estimates generated by current approaches lack a clear surround structure [20]. To circumvent these limitations, we designed a high-throughput and low-cost epi-fluorescent approach that enables imaging larger field-of-view (FOV) (1.7 mm^2 sampling at ~1 μm/pixel) and permits >1 hr long recordings of the same FOV while avoiding artifacts caused by small retina wrinkles and laser scanning. Our method takes advantage of red Calcium sensors (e.g., RCamp1.07 [28]) that separate the Ca^{2+} indicator's red-shifted excitation light from the opsin absorption spectrum (Fig. 2a, b; see methods), and allows robust responses to UV visual stimulation. We used the VGluT2-cre driver line to specifically target RGCs (Fig. 2c), leading to a uniform expression across the entire retina. All RCamp1.07 positive somata correspond to RGCs, as seen by the RGC specific marker RBPMS [29] (Fig. 2d). Double-positive cells accounted for ~40 % of all RGCs. This expression pattern appears to be RGC-type-specific, as seen by the SMI-32 alpha-RGCs co-labeling. Alpha-RGCs were consistently excluded from the expression profile, apart from a single, sparse and spatially distributed type (Fig. 2e). Using this line, we were able to reproduce and expand previous large-scale imaging results in single retinas, as seen in direction-selective (DS) and non-DS responses (Fig. 2f), the cardinal DS response distributions (Fig. 2g), and

clustering and reproducibility of responses to changes in frequency, contrast and luminance, known as the "chirp" stimulus [20] (Fig. 2h & Suppl. Fig. 2). Moreover, by sequentially recording 4-7 FOVs (Fig. 2c, orange box depicts one FOV), each for approximately 25 min, we were able to record neural activity from up to ~6 mm^2 of retinal surface (~40 % of the total retinal area). By experimental design, the position of each FOV was random. Moreover, the strength of the functional responses in each consecutive session was unaltered (Fig. 2i). Importantly, using a novel "shifting" white-noise approach (Suppl. Fig. Fig. 3 & see methods), we were able to estimate high-resolution and high-SNR spatio-temporal RFs for ~85 % of recorded cells (Fig. 2j, top row). The quality of these RF estimates allowed for automatic parameterization of the spatial RF into center and surround using a difference of Gaussians model (Fig. 2j, bottom row & k, Suppl. Fig. 4, see methods). In total, we recorded 9 retinas, reconstructing and parameterizing 26558 spatio-temporal receptive fields, enabling an unprecedented opportunity to index RGC responses across single retinas. Taken together, our methodology is a simple, efficient, and affordable extension to the current retinal research toolbox.

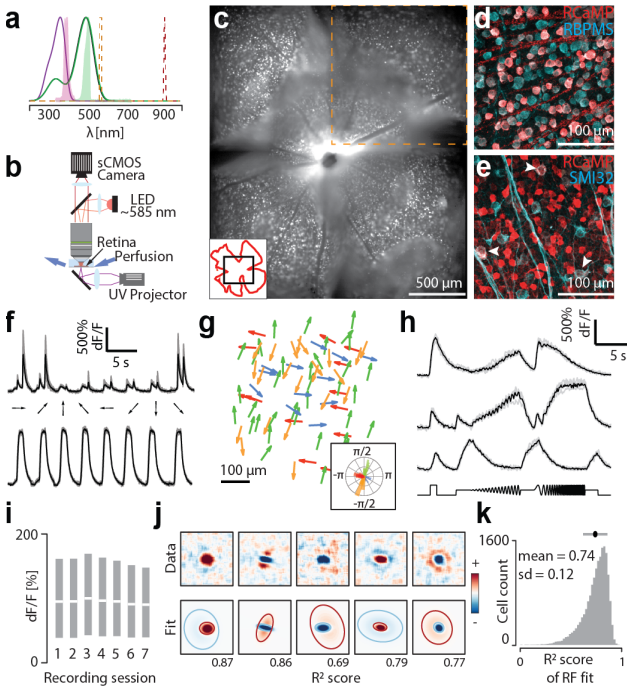


Figure 2. Large-Scale retinal receptive field mapping. a*. Normalized absorption spectra of mouse photoreceptors (purple: S opsin; green: M opsin). Overlaid, the normalized emission spectra of the UV and green light emitted by the DLP projector (filled purple: UV, filled green: green light), epi-fluorescent excitation (orange), and two-photon (red) excitation.* ***b****. Schematic of our epi-fluorescent imaging setup.* ***c****. Montage of five consecutively recorded fields (orange inset) of a whole-mounted mouse retina from a Vglut2-ires-cre; TITL-R-CaMP1.07-D; ROSA26-ZtTA triple transgenic mouse. Inset - black: imaged montage, red: retinal outline.* ***d****. Double-label immunostaining of RCamp1.07 expressing RGCs (red) and RBPMS (cyan).* ***e****. As in (****d****) but labeling SMI32 (cyan). Arrowheads depict double-labeled cells.* ***f****. Example Ca^{2+} signals (grey: 5 repetitions; black: mean) from direction (top panel) and non-direction selective responsive RGCs (bottom).* ***g****. Example distribution of preferred directions in one FOV. Inset: polar plot of*

DS preference. ***h****. Example Ca^{2+} signals to chirp stimulus from 3 different RGCs (grey: 5 repetitions; black: mean).* ***i****. Recording Ca^{2+} signal stability across sequentially imaged FOV for all nine retinas (each session lasts ~25 min, 3-7 sessions per retina) White lines and gray bars denote medians and 25th and 75th percentile range of the dF/F distribution respectively.* ***j****. Example receptive fields recorded using "shifting" white noise (top) and their respective parametrizations (bottom). Blue and red ellipses correspond to two standard-deviation contours of the ON and OFF Gaussians, respectively.* ***k****. Histogram of goodness of fit for all recorded receptive fields.*

Receptive fields are adapted to anisotropic scene statistics across the visual field

Taking advantage of the high-resolution RFs, we examined variations in RF shapes and strengths across the retina. Given that the PC model does not determine the polarity of the optimal RF and to globally compare all retinal RFs, ON-center RFs were flipped in sign, such that all centers were negative and all surrounds were positive. This allowed us to pool across all cells within small bins on the retinal surface and visualize the average spatial RF at different locations of the retina (Fig. 3a, Suppl. Fig. 5). We observed a general and reproducible trend across retinas: a streak-shaped area where all RF surrounds were oriented towards the optic nerve, and below which, hardly any RF surrounds were observed (Fig. 3a, Suppl. Fig. 5). To compare these spatial variations of RFs with our theoretical predictions, we oriented all recorded retinas to a common coordinate system (see methods). Average receptive fields in the ventral, central-dorsal and peripheral-dorsal retina (Fig. 3b, bottom row) qualitatively matched model RFs predicted for the upper, medial and lower visual fields, respectively (Fig. 3b, top row). To confirm the change of relative surround strengths independently from surround asymmetry, we computed the radial profiles of RFs and these also strongly resembled the radial profiles for model RFs (Fig. 3c). Overall, model RFs qualitatively reproduced all aspects of average spatial RFs across different elevations with remarkable detail.

To measure these phenomena quantitatively, we made use of the RF parametrizations and pooled RF parameters for all cells in different 2-d (Fig. 3 d,f,h) or 1-d (Fig. 3 e,g,i) bins across the retina. In line with our theoretical predictions (Fig. 1d, black), our analysis shows that the relative surround strength increases gradually along the dorsoventral axis (Fig. 3d), a trend visible in every single retina (Fig. 3e). No correlation was observed in the temporal-nasal axis (Suppl. Fig. 6). Next, we explored if we could observe any global change in RF center size. As predicted (Fig. 1d, purple), center sizes decrease only across the dorsoventral axis (Fig. 3 f,g, Suppl. Fig. 6). While examining the spatial distribution of differences in upper and lower halves of the RF surrounds, we identified a consistent and prominent asymmetric streak in the dorsal retina, between 700-900 µm dorsally from the optic nerve (Fig. 3h, i), as one would expect from asymmetric visual inputs (Fig. 1 e). To test whether this streak corresponds to the horizon line within the animal's visual field, we transformed the retinal coordinates to visual coordinates [30] and used the S-opsin gradient [31] to define the dorsoventral axis (Fig. 3j, k, see methods). We computed the relative offset of the center-of-mass for the center and surround fits, as a measure of magnitude and orientation of surround asymmetry (Suppl. Fig. 4). Overlaying the measured RF asymmetries with the opsin gradient indicated that the asymmetry is pronounced in the opsin

transition zone (Fig. 3j, left panels). In visual coordinates, the center of this asymmetric streak is located at 0 deg elevation, spanning the entire azimuth of our imaged FOVs (Fig. 3k), in line with our theoretical predictions (Fig. 1c).

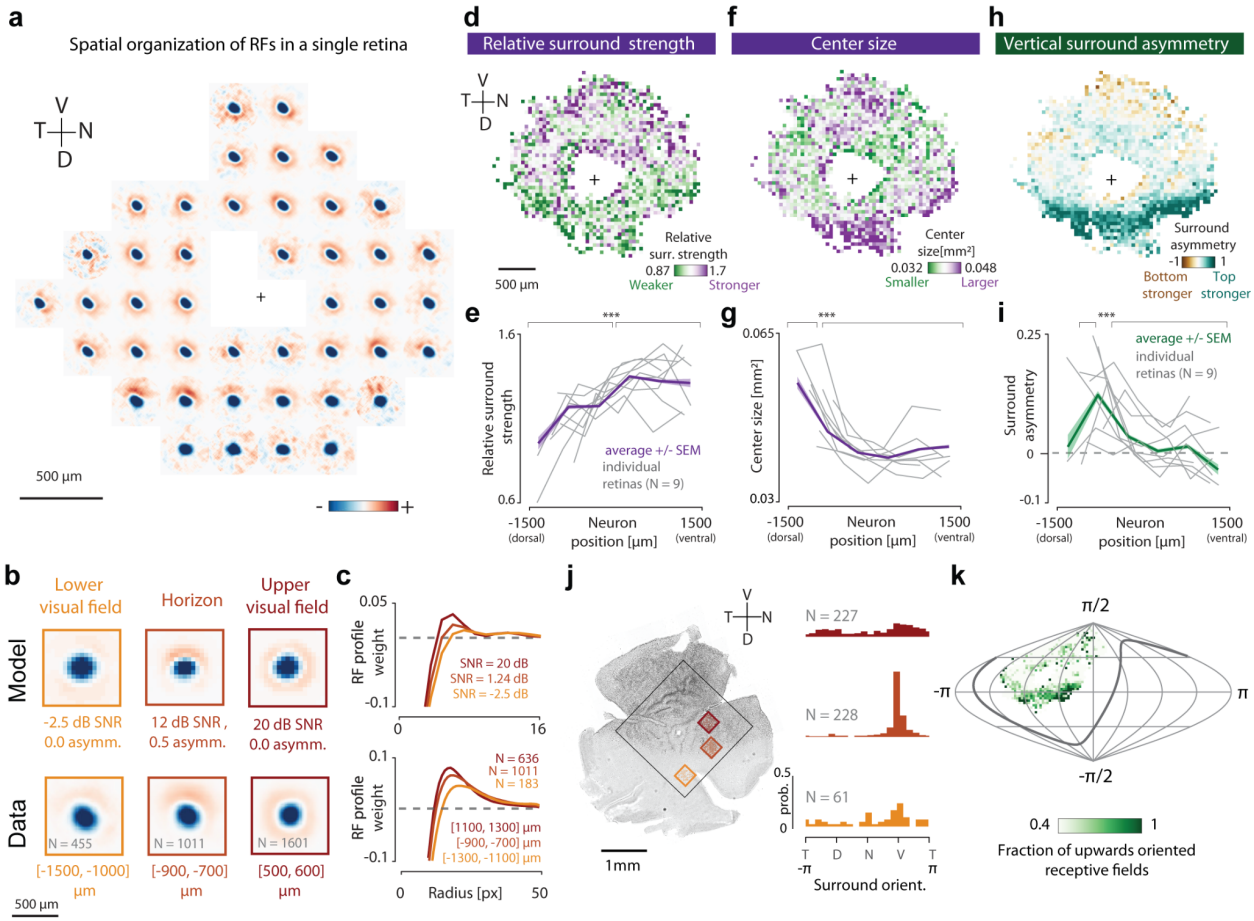


Figure 3. Retina-wide receptive field architecture. a. *Average spatial RFs of all RGCs pooled from square bins of size 300 µm at different positions of one retina (n = 64 ± 52 cells per bin, Black cross: optic nerve head).* ***b.*** *Top row: optimal RFs predicted by the model at different elevations of the visual scene. Bottom row: Average spatial RFs of neurons along different dorsoventral locations on the retina.* ***c****. Top: Radial profiles of model RFs at different SNR levels. Bottom: Mean radial profiles of RGC RFs in bins along the dorsoventral axis.* ***d****. Mean relative surround strengths of RGCs within 50 µm bins, pooled from n = 9 retinas.* ***e****. Relative surround strengths for RGCs within 6 equally spaced bins along the dorsoventral axis (color: mean and SEM pooled from all n = 26558 RFs, grey lines: individual retinas). Note: that the bins for each retina may not align due to slight differences in sampling the retina's surface.* ***f*** *&* ***h****. Same as* ***d****, but for center size and vertical surround asymmetry, respectively.* ***g*** *&* ***i*** *same as* ***e,*** *but for center size and vertical surround asymmetry, respectively.* ***j****. Left: One of the retinas, immunostained for S-Opsin. Black box shows the region imaged for RF mapping. Right: Normalized histograms of surround orientations of RGCs within corresponding bins marked on the right.* ***k****. Proportion of imaged neurons, within local neighborhoods, that have ventrally oriented surrounds, overlaid on a sinusoidal projection of visual space. The animal is centered at (0 deg latitude; 0 deg longitude), facing towards the viewer and the black line shows the area of the visual field viewed by one eye. (p-values for Kolmogorov-Smirnov test:* ***e****: 6.92x10^{-5},* ***g****: 7.20x10^{-5},* ***i****: 5.46x10^{-4}; see Suppl. Fig. 6 for extensive statistical comparisons). V - Ventral, N - Nasal, D - Dorsal, T - Temporal.*

Adaptations to natural scene statistics across retinal pathways

It has long been assumed that specific RGC pathways have stereotyped response properties, shaped by the interactions between direct excitation in their center and indirect inhibition in their surround [32]. Thus, one would expect that these center-surround interactions are uniform across visual space for most RGCs pathways. To assess the specificity of the observed RF adaptations across functional RGC pathways, we clustered cells based on the temporal RF profile of 26558 RGCs into functional groups using a Gaussian Mixture Model (GMM), as done previously [20]. Consistent with the proportion of RGCs labeled in our line (Fig. 2d), we defined 10 functional clusters (Fig. 4a). Each cluster group had distinctly shaped temporal filters, corresponding to different functional properties such as ON or OFF selectivity, transient or sustained responses, and monophasic or biphasic selectivity, as seen in their average profiles (Fig. 4a, right panels). Cluster membership statistics were conserved across different retinas (Suppl. Fig. 7). Moreover, the relative positions of RGCs belonging to individual clusters tile the retina in a mosaic-like arrangement in many cases (Fig. 4b, left panels), confirming that the clusters are indeed functionally distinct and irreducible RGC types [17]. As expected, some clusters represent a combination of RGC types that cannot be identified solely by their temporal profiles (Fig. 4b, cluster 8). We next used this classification and looked at the relative surround strength, center size, and asymmetric strength across clusters (Fig. 4c). As with the global pooling of RF shown in Fig. 3, cells in each functionally defined cluster increase their relative surround strengths gradually in the dorsoventral axis and decrease their center sizes accordingly. Moreover, most clusters appear to contribute to the asymmetric streak. This is consistent with the distribution of asymmetries in the dorsal retina, where most cells have a strongly oriented surround with a ventral bias (Fig. 4d, left panel).

Unexpectedly, we observed that a subpopulation of cells has surrounds strongly oriented along the four cardinal directions. This cardinal asymmetry can be visualized by averaging their RFs, which, apart from the relative orientation, are indistinguishable from each other (Fig. 4d, right panels). Interestingly, these cardinal biases were also present in the ventral retina where they were equally represented (Fig. 4e, left panel). As with the dorsal asymmetries, the average RF for each directional group was indistinguishable apart from their relative orientations (Fig. 4e, right panels). Although asymmetric RF structure indicates direction-selective (DS) properties [32,33], these cells do not appear to be just the classical ON-OFF DS cells. We observed no clear correlation between an increase in direction-selectivity and RF asymmetry. Only for RGCs that had DS properties, a correlation for gratings, but not for moving bars, could be determined. Overall, our results suggest that these cells could show enhanced DS responses to stimulus parameters that match their spatio-temporal profile (Suppl. Fig. 8).

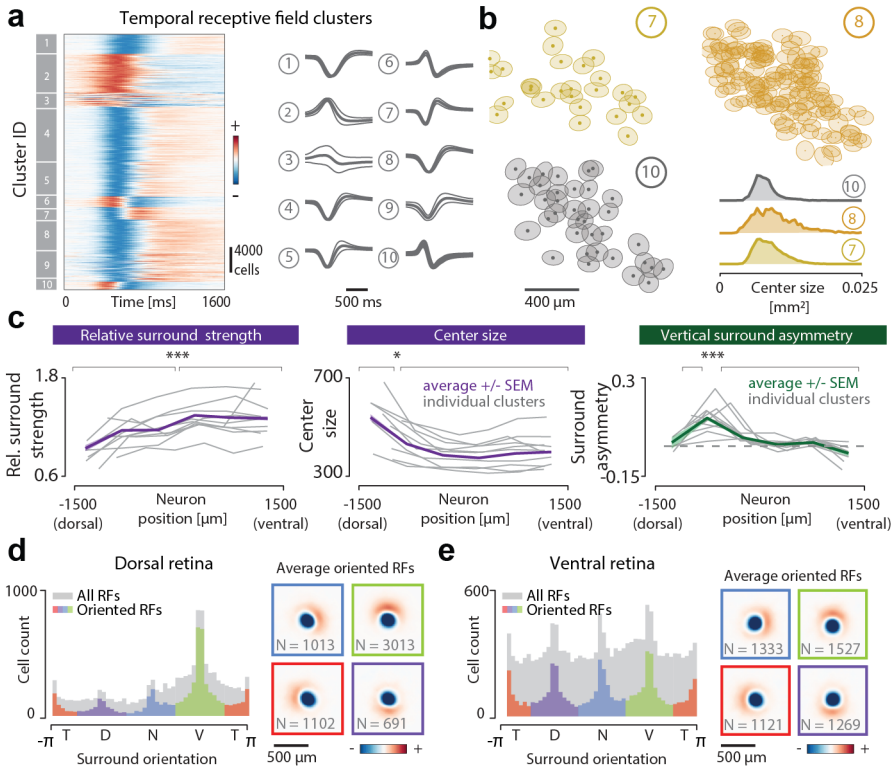


Figure 4. Global adaptation across retinal pathways. a. *Left: Temporal RFs of 26558 RGCs, grouped by cluster membership, as determined by GMM. Right: Mean and standard deviation of temporal RFs in each cluster.* ***b.*** *Locations of RGCs (dots) and their RF centers (filled ellipses) belong to three different clusters in a small region of one retina. Bottom right: Distribution of center sizes for these three clusters.* ***c.*** *Trends for relative surround strength (left), center size (middle) and vertical surround asymmetry (right) for cells within each cluster (grey) and pooled from all clusters (color), binned as in Fig. 3e. (p-values for Kolmogorov-Smirnov test: left: 6.17 x 10^{-5}, middle: 0.01, right: 2.51 x 10^{-5})* ***d****. Left: Distribution of surround orientations in the dorsal retina for all cells (grey, n = 11166) and cells with a large center-surround distance (colored, n = 5819). Note: symmetric cells will also have an orientation due to fitting noise, but negligible center-surround distance. Right: Average spatial RFs where the surround is oriented to one of the four cardinal directions (color-coded as on left).* ***e****. Same as* ***d****, but for the ventral retina. (n = 15392 total RGCs of which n = 5250 had oriented surrounds).*

Discussion

In this study, we leveraged a novel, high-throughput neural imaging setup (Fig. 2) to identify a new kind of adaptation to panoramic scene statistics in the retina. In agreement with theoretical predictions derived from the efficient coding framework (Fig. 1), our experimental results indicate that the global RF architecture is adapted to encode panoramic natural scenes efficiently (Fig. 3 & 4). Classically, RGCs are known to dynamically change the strength of the RF surround in response to varying light levels [34], which is thought to further increase the efficiency of sensory coding [15,35,36]. Our findings demonstrate that, in addition to such dynamic effects, RF shapes are also determined by static factors, namely their position within the visual field. In that way, the

retina simultaneously exploits the large-scale spatial and fine-scale temporal structure of the visual space.

The design principle underlying the retina's organization has been a matter of debate. While the motivation behind our work is the principle of minimizing the metabolic cost of sensory coding and communication [8], a competing view proposes that RGCs immediately extract relevant features of sensory signals [37]. According to this view, features, such as direction-selectivity [38] or color-opponency [39,40], are of immediate behavioral relevance and are relayed by dedicated functional cell types to guide behavior. Although both perspectives can coexist, in our view, results described here bear relevance for this debate. The RF shape partially determines the computation of behaviorally relevant features performed by a neuron. However, we demonstrated that RF shapes are consistently modulated by efficiency constraints across diverse cell types, effectively changing the quality of the encoded feature across the visual field (Fig. 4a-c). We note that rigorous statement and test of the general-purpose efficient coding and feature representation hypotheses (e.g., within the information theoretic framework [2–4,19]) will require a directed effort, and is outside the scope of this work. One feature encoded by RGCs is direction-selectivity (DS). DS encoding can emerge as a consequence of an asymmetric and time-shifted surround, as shown before [32,33]. Our results indicate that the computational goal of the pronounced ventral asymmetric surround at the horizon is consistent with minimizing redundancy and efficient coding, not with upward-motion detection. A distinct, yet related question can be asked about the emergence of DS in retinorecipient areas, such as the Superior Colliculus, where neurons integrate input from the entire retina, including the asymmetric streak. Some studies have described these neurons as sensitive to upward motion, whereas others do not find such specificity [41,42]. Interestingly, all of these studies use different sets of artificial stimuli. The efficient coding interpretation, such as the one adopted here, suggests that the key to resolving this puzzle might be understanding the statistics of what the animal ought to see in Nature.

How could the visual system establish such global RF architecture without fine-tuning each RGC pathway independently? One partial mechanism would be the non-uniform distribution of spectral sensitivity across the retina [31]. Such distribution has been discussed to be relevant for color vision [39], contrast coding [6,43], and the detection of aerial predators in the sky [44], but simultaneously, could influence the static RF adaptations. For example, whereas the mouse retina has green light sensing photoreceptors across the entire retina (M- and Rod-opsin), UV sensitivity follows a sharp dorsoventral gradient (S-opsin) [31]. From the RGC's perspective, both inputs will be added at mesopic conditions. In consequence, RGCs responses change their spectral sensitivity along the same functional axis as our predictions (Fig. 3j). Interestingly, the vertical intensity gradient will flatten at lower ambient light levels, e.g., at dusk and dawn. In conjunction, the relative strength of UV sensitivity would also decrease, the spectral distribution. In these conditions, efficient-coding hypothesis would predict a more homogeneous receptive field distribution across the dorsoventral axis. It would be revealing to test if the retina-wide receptive field organization is dynamically reshaped at scotopic conditions. Other mechanisms might also lead to RF asymmetries. For example, the RF of J-RGCs [32,39], have surrounds that are

asymmetric, ventrally displaced, and M-cone and rod-sensitive. Currently, however, these specific global properties are not testable with our experimental setup and remain a subject of future research. Finally, it would be beneficial to maintain a constant relative position of the horizon within the field of view, additional mechanisms would be required. In agreement with this idea, eye and head movements stabilize the retinal image remarkably well during behavior [45].

The exact pattern of global retinal adaptation to the visual sensory niche should vary across species occupying diverse environments. It has been found that dorsoventral opsin gradients are present in many different mammalian species. For example, the rabbit, Chilean subterranean rodent cururo, European mole, the shrew, and even the spotted hyena [44,46] show higher S-opsin expression in the ventral retina. However, not all ecological niches are identical. For example, in dense forests, the vertical gradients of luminance and contrast are less prominent, and a clear horizon line might not be apparent. Interestingly, forest mice species whose opsin distribution have been described, present a spatially uniform opsin distribution [47]. This further strengthens our hypothesis, which relates the global organization of the retina to the statistics of the ecological visual field. Taken together, these findings stress two critical requirements for building a more general theory of vision: analysis of the animal's natural environment and exploration across the animal kingdom [48].

Methods

Theory

We modeled neural responses r_t as a dot product of the model receptive field $\vec{\phi}$ and noisy stimulus vectors (image patches) $\vec{x}_t : r_t = \vec{\phi}^T s_t$, where T denotes vector transposition. Receptive fields (filters) were optimized to minimize the following cost function:

$$L(\vec{\phi}) = \left\langle \sqrt{r_t^2} \right\rangle_t + \lambda \sum_{i=1}^{N} \phi_i^2 d(i)$$

Where $d(i)$ is the squared distance between the i-th value of the RF and the one with the peak absolute value, and λ is the strength of the spatial locality constraint. This form of the locality constraint has been introduced in [2], and it has been demonstrated that it is consistent with RGC properties [2,4]. We note that the activity-related term in the cost function is equivalent to maximizing the sparsity of the neural activity quantified as the average absolute value of neural responses [14]. Without the spatial locality constraint (i.e., $\lambda = 0$, the receptive field) the optimal RF is an oriented, Gabor-like filter. During optimization, to avoid convergence to trivial solutions, the norm of the RF $\vec{\phi}$ was constant and equal to 1. Overall, this cost-function enforces minimization of activity conveyed downstream, while preserving the information about the image and meeting the locality constraints. Conceptually, this goal is equivalent to that of the PC model [1]. Our model generates predictions consistent with the PC model proposed in [1] (see Suppl. Fig. 1). It is, however, more flexible, and enables us to capture changes in the center size.

We modeled output of photoreceptor cells (stimuli s_t) as natural image patches distorted with the additive Gaussian noise with variance σ^2 i.e.: $s_{t,i} = x_{t,i} + \xi$, where $\xi \sim \mathcal{N}(0, \sigma^2)$ is the noise term. To simulate different SNR conditions, we manipulated the noise variance level, and optimized RFs for each of the noise levels separately.

We optimized RFs by numerically minimizing the cost function L via gradient descent. For training, we used a dataset of 50 000 square image patches 27 x 27 pixels in size taken from a movie of animals in the African Savannah used in [49] . Each image patch was normalized to have a zero mean and unit variance. To simulate the impact of changing SNR, we distorted normalized images with noise. During optimization, the dimensionality of natural image data was reduced with Principal Component Analysis (PCA) to 128 dimensions. For each noise level, dimensionality reduction was performed using the same matrix of PCA components computed on noiseless data. To simulate the impact of changing SNR homogeneity, prior to normalization we multiplied the bottom half of each image by a scaling factor lesser than one. After such scaling, we normalized the data and added noise of constant variance. During optimization of RFs on asymmetric stimuli, we computed PCA for each level of SNR asymmetry separately.

In all cases, prior to optimization, in order to enforce that the RF is centered in the image patch, we initialized RFs with random Gaussian noise with variance equal to 0.1 but set the central pixel value to -1.

To evaluate RF properties, we defined the size of the RF to be the smallest circle which included 90% of energy (i.e. ϕ_i^2) of optimal RFs averaged across all noise levels (Fig. 1 b, c). Within that circle, we defined the center to be all $\vec{\phi}$ values smaller than 0, and the surround to be those larger than or equal to 0. The strength of the surround was thus equal to $\sum_{i:\phi_i \geq 0} |\phi_i|$ and the center to $\sum_{i:\phi_i < 0} |\phi_i|$. Sizes of center and surround were simply numbers of entries that were smaller than or greater-or-equal to 0 respectively.

To characterize changes in contrast and luminance across the visual field we used natural images published in [5]. We limited our analysis to UV images only, however light statistics of the green channel do not differ qualitatively. The images provided in [5] were divided into two classes - upper visual field and lower visual field. To simulate the visual horizon, we concatenated pairs of images randomly selected from the upper and lower visual fields. We created a dataset of 1000 such concatenated images, and used them to compute the mean and variance of light intensity as estimates of local luminance and contrast, as well as to estimate the SNR as a function of elevation. To estimate the vertical asymmetry of the SNR pattern we used a square stimulus window, and a fixed noise variance. We note that key, qualitative aspects of our predictions do not depend on these choices. For each position y of the window along the vertical dimension of the visual field, we computed the vertical SNR asymmetry as: $asym(y) = \frac{m_{up}^y - m_{down}^y}{m_{up}^y + m_{down}^y}$ where m_{up}^y ad m_{down}^y are sums of the SNR value within the stimulus window above and below its midline respectively.

Animals

Mouse protocols were reviewed by the institutional preclinical core facility (PCF) at IST Austria. All breeding and experimentation were performed under a license approved by the Austrian Federal Ministry of Science and Research in accordance with the Austrian and EU animal laws. Triple transgenic male and female mice, age 5-12 weeks were used for this study (Vglut2-ires-cre (JAX 28863), TITL-R-CaMP1.07-D (JAX 030217) and ROSA26-ZtTA (JAX 012266)). The mice were housed in a standard 12-hour day-night cycle and euthanized by cervical dislocation before imaging.

Imaging

The dark-adapted mouse retina was isolated under far-red light (LED peak 735 nm, additionally filtered with a 735nm LP filter eliciting an isomerization rate of ~17 R*/s) in oxygenated Ames' medium (Sigma) with constant bubbling (95% O2, 5% CO2) at room temperature. Left and right retinas were kept separate for identification. Four incisions were made to flat-mount the retina, with ganglion cells facing up, on a 18 mm coverslip (VWR 631-0153), and held down with a filter paper (Merck GSWP01300) having a ~2.5 mm x 2.5 mm imaging window cut out. The preparation was then placed in a heated (32 °C) superfusion chamber on the stage of a custom-built upright fluorescence microscope. The retina was left to recover for a minimum of 10 minutes

with the excitation light of the microscope turned on. An amber LED (Thorlabs M595L4) filtered with a BP filter (Thorlabs FB580-10) was used for excitation and a BP filter (Thorlabs 641-75) in series with a 600nm LP filter (Thorlabs FEL0600) was used for collection. Background excitation light intensity was at a constant mean photopic intensity of 10^5 R^*s^{-1} per rod (at 585 +/- 5 nm). Isomerization rates were determined using opsin templates [50] and assuming that the mouse rod has an optical density at peak absorption wavelength of 0.015 μm^{-1}, a length of 24 µm, a diameter of 1.4 µm and a quantum efficiency of 0.67 [51,52]. Each retina was tiled by recording 3-7 different fields of view (FOV) at 10x magnification (Olympus XLUMPLFLN20XW objective) using a sCMOS camera (Photometrics Prime 95B) at 10 fps and 1.1 µm/pixel resolution.

Visual Stimuli

Light stimuli were delivered from a modified Texas Instruments DLPLCR4500EVM DLP projector through a custom-made lens system and focused onto the photoreceptors (frame rate 60 Hz, magnification 2.5 µm per pixel, total area 3.2 mm x 2 mm). The projector's blue LED was replaced with a high-power UV-LED (ProLight 1 W UV LED, peak 405 nm), to improve the differential stimulation of S pigments. Two SP filters in series (Thorlabs FESH0550) were put in the stimulus path to block green light from entering the camera. Intensities and spectra were measured using a calibrated spectrometer (Thorlabs CCS-100) and a digital power meter (Thorlabs S130C sensor). A shifting spatiotemporal white-noise stimulus was presented using a binary pseudo-random sequence, in which the two primary lights (green and UV) varied dependently. All white-noise stimuli were presented at 6 Hz update for 15 min. The checker size was 100 × 100 μm^2 and the entire grid was shifted by random multiples of 10 µm in both x- and y-axis after every frame. In comparison experiments, static checkers (without shifts) of 100 × 100 μm^2 and 25 x 25 μm^2 were interleaved with the shifting checkers in chunks of 5 min for a total of 20 minutes for each of the three checker types. A 'chirp' stimulus with a 1s bright step, increasing amplitude (0 to 127 over 8 s) and increasing frequency (0 to 4 Hz over 8 s) was repeated for 5 trials to reproduce clustering of responses [20]. Moving square gratings (0.6 cycles/s temporal frequency and 0.025 cycles/µm spatial frequency) or a wide bright bar (1 mm/s speed, 2 mm width) in 8 directions, repeated for 5 trials, were used for assessing direction selectivity.

Histology

After the ex-vivo recordings, some of the retinas were fixed with 4 % PFA for 30 min and stained for S-opsin and RFP. Retinas were incubated for 7 days at 4 °C in PBS, containing 5 % donkey serum, 0.5 % Triton X-100, goat anti S-opsin (1:500, Rockland 600-101-MP7) and rabbit anti-RFP (1:1000, Rockland 600-401-379). After washing in PBS for 15 min thrice, retinas were incubated overnight in secondary antibodies, donkey anti-goat Alexa Fluor 488 (1:1000 Abcam ab150129) and donkey anti-rabbit Alexa Fluor 594 (1:1000, Invitrogen R37119). Retinas were then mounted and imaged with Olympus VS120 Slidescanner with a 20x objective. For cell-type characterization, Vglut2-ires-cre; TITL-R-CaMP1.07-D; ROSA26-ZtTA mice were euthanized and perfused intra-cardially, followed by retina extraction and staining for RBMPS or SMI-32, along with RFP (primary antibodies: guinea pig anti-RBPMS (1:500, Sigma ABN1376), mouse

anti-SMI32 (1:500, BioLegend 801701), rabbit anti-RFP (1:1000, Rockland 600-401-379) or mouse anti-RFP (1:500, MBL M155-3); secondary antibodies: goat anti-Guinea pig Alexa Fluor 647 (Invitrogen A21450), donkey anti-mouse Alexa Fluor 647 (1:1000, Abcam A-31571) and donkey anti-rabbit Alexa Fluor 594 (1:1000, Invitrogen R37119). The staining protocol was the same as above and these retinas were imaged with Leica SP8 confocal microscope.

Preprocessing

ROIs were detected automatically from raw calcium movies using Suite2p [53]. Fluorescence traces, F, were detrended by computing $\Delta F/F$, where the 8th percentile in a 20 s sliding window centered around each time point was taken as the baseline fluorescence. Different fields of view from the same retina were stitched together based on coordinates from the stage motors and repeated ROIs in overlapping regions were manually annotated using an open-source tool [54]. Repeated ROIs with the highest score in Suite2p's built-in classifier were kept for analysis. The deconvolved signal from Suite2p (with *tau* = 1.0 s) was used for calculating receptive fields.

The median $\Delta F/F$ response across trials was taken as the response to the chirp and normalized by dividing by the maximum of the absolute values across time. Quality Index was computed as in [20], and only responses with QI > 0.45 were used for clustering (8019 out of 30798 neurons). For moving gratings and bar, the mean across trials and maximum across time was taken as the response in any one direction. The Direction Selectivity index (DSi) was computed as the normalised vector sum of these responses, as in previous work [27,55].

Receptive Field Mapping

The receptive field for each neuron was computed as a calcium-triggered-average. The spatiotemporal receptive field at latency τ, position (x, y) for neuron i was computed as,

$$RF(i,x,y,\tau) = \sum_{t=1}^{T} s(x,y,t-\tau)\cdot r(i,t) - \sum_{t=1}^{T} s(x,y,t-\tau)\cdot r(i,t)\cdot \sum_{t=1}^{N} r(j,t) \quad ,$$

where $r(i,t)$ is the deconvolved response of neuron i at time t, $s(x,y,t)$ is the white noise stimulus, T is the length of the recording and N is the total number of neurons in the recording. The second term in this equation subtracts away the residual distribution of the stimulus and the contribution of slow bleaching that is common to all neurons in that recording and leads to receptive fields that had noticeably less noise. Only the UV channel of the stimulus was used for RF mapping. The latency τ was varied in increments of 0.025 s (40 Hz), and the stimulus was interpolated by using the last frame before a particular time, $t-\tau$. The RF for each neuron was normalized between [-1,1] by subtracting the mean value of the RF at latencies $\tau < 0$, and dividing by the maximum absolute value of the entire RF.

The location of the center of an RF was estimated by finding the pixel that varied the most across time, $P_{var} = \mathrm{argmax}\ Var^t$, where Var^t is the variance across time. Each neuron's RF was cropped within a square window of edge 1mm centered on this pixel. *SNR* of an RF was computed as the peak-to-noise ratio where the power of noise was estimated in regions with distance >0.5mm from the point P_{var}. Only RFs with peak SNR > 15dB were kept for analysis (26558 selected RFs out of 30798 recorded neurons). The location of the RF in time was found

in a similar way; $T_{var} = \text{argmax } Var^{xy}$, where Var^{xy} is the variance in space.

Parametrization of Receptive Fields

We parametrize spatiotemporal properties of the center and surround of the RFs as a sum of two 2D Gaussian distributions (Gaussians) $\mathcal{G}_1 + \mathcal{G}_2$. The first Gaussian represents the center of the RF; its amplitude can be either negative or positive corresponding to an OFF or an ON cell. The amplitude of the second Gaussian is required to be of the opposite sign to model the properties of the surround. We compute a spatial representation of the RF, denoted as RF_{2D} as the median of the RF within a small time window around T_{var}. To reduce noise, we exclude the pixels that are weakly correlated with P_{var} across time. The sum of 2D Gaussians $\mathcal{G}_1 + \mathcal{G}_2$ is then fitted to RF_{2D}. Each Gaussian is defined by the amplitude A, center (m^x, m^y), width (σ^x, σ^y) and the orientation angle θ. We start by fitting only one Gaussian $\mathcal{G}_1$ to parametrize the location of the center (m_1^x, m_1^y) and to have an initial guess of the size of the center $(\sigma_1^{x0}, \sigma_1^{y0})$. In the next step we fit the sum of Gaussians, where we fix the center of the $\mathcal{G}_1$ to the point (m_1^x, m_1^y) while allowing all other parameters to be fitted anew.

Here we differentiate two types of RFs: a RF with a strong center and a weaker surround that largely overlap, and a RF where both center and surround components are strong and well separated. For the first case, we impose a constraint such that the center of $\mathcal{G}_2$ is within the distance $d = 2\min(\sigma_1^{x0}, \sigma_1^{y0})$ from the center (m_1^x, m_1^y). We remove this constraint for the second type of the RFs. To decide the type of the RF, we find the maximum and minimum points of the RF, and we compute the distance between them and the ratio of their absolute values. If the ratio of the smaller to the bigger values is less than 0.75 or if the distance between the extrema points is less than d, then we classify such a RF as the first type, and as the second type otherwise. Experimentally we found that imposing such a constraint on the location of the second Gaussian leads to a better fitted sum of Gaussians for RFs with largely overlapping center and surround components. All the fitting procedures were implemented using the nonlinear least-squares solver `lsqcurvefit` in MATLAB.

Using the parametrization, we can compute various RF characteristics. We find two sets of pixels corresponding to the center and the surround. Center pixels are the pixels within two standard deviations from the center of the $\mathcal{G}_1$. Surround pixels are the pixels within two standard deviations from the center of the $\mathcal{G}_2$ and that are not center pixels. The *Center Size* is the number of pixels in the center set. The *Relative Surround Strength* is the absolute ratio of the sum of the surround pixels to the sum of center pixels. *Vertical Surround Asymmetry* is defined as $(u - l)/(u + l)$ where u and l denote the sum of absolute values in the upper and lower halves of the surround pixels, respectively. The *distance* between the center and the surround is the distance between the center of mass (COM) of the center pixels and the COM of the surround pixels. The *orientation* is the angle between the horizontal axis and the line connecting the two COMs. *Radial profiles* were computed as the average values of the pixels in RF_{2D} within the rings of the increasing radii centered on the point P_{var}. The average value of the center or surround pixels across time was taken to be the *RF center or surround temporal dynamics*, respectively. The *R2*

goodness of fit is computed as

$$R2 = 1 - \frac{\sum_i (RF_{2Di} - \overline{RF_{2D}})^2}{\sum_i (RF_{2Di} - M_i)^2},$$

where $M = \mathcal{G}_1 + \mathcal{G}_2$ is the RF parametrization. The values of the above RF parameters are reported for a few representative neurons in Suppl. Fig. S4.

Retina alignment

All functional imaging experiments were performed with randomized retina orientations. For the retinas that were co-stained against S-opsin and RCaMP1.07 (n = 4), the direction with the highest density of S-opsin was taken to be the ventral direction [56]. The stitched maximal projection images of the functional imaging experiments were aligned to the RCaMP channel using the retinal vasculature. We noticed that the ventral direction was robustly and consistently aligned to the most common direction of the surround orientations (Fig. 3j,k), thus we used this procedure to align all retinas to the same functional reference frame (Fig. 3a-g, Fig. 4). This automated functional aligning procedure, when applied to the S-opsin retinas, had a mismatch of 8 ± 38 deg (average ± std). Note that while this procedure ensures that the ventral direction has the most number of oriented surrounds, it does not presuppose that these cells must be concentrated in any particular region of the retina, like the dorsal streak. The coordinates of cells from each retina were translated to make the optic nerve the zero of the coordinate system and rotated such that the positive y-axis denoted ventral direction. Left retinas were flipped in the nasotemporal axis such that the positive x-axis denoted nasal direction for all retinas. All spatial receptive fields were also translated and rotated accordingly.

The cartesian retinal coordinates of cells in the stained retinas were converted to spherical visual coordinates using the R-package Retistruct, assuming the optical axis of the mouse eye to be oriented at 64 deg azimuth and 22 deg elevation from the long axis of the animal [30].

Binning of RF properties

For 2D bins, the retinal space from -1500 µm to 1500 µm in both nasotemporal (NT) and dorsoventral (DV) axes was divided into a square grid and all neurons within each bin were collected. Only 2D bins with at least 5 neurons were analyzed to minimize sampling bias. The spatial RF of all neurons within each bin of size 300 µm were averaged and plotted at the location of the bin in (Fig. 3a, Suppl. Fig. S5). The RF parameter values of neurons within each bin of size 50 µm were averaged to yield a 2D map of the parameter, and this map was visualized (without smoothing) in (Fig. 3 d,f,h). Owing to its area-preserving property, the sinusoidal projection of visual space was binned in the same way as the retinal surface, and the fraction of cells in each bin that had ventrally oriented surrounds were plotted in Fig. 3k. For 1D analysis, the bins were defined along the DV or NT axis based on the range of coordinates in a particular group (by retina: Fig. 3 or by cluster: Fig. 4). The range was divided into 6 equally spaced bins and the mean parameter value of neurons within each bin were plotted at the coordinate of the center of the bin. The mean and SEM values for neurons across all groups were computed and plotted in

the same way. A two-sample Kolmogorov-Smirnov test was performed between values of RF parameters in all pairs of bins (6x6).

Clustering into functional types

The Gaussian Mixture Model procedure developed in [20] was used for clustering temporal RFs and chirp responses, separately. In brief, after normalization, the trace was first reduced in dimension using Principal Component Analysis (10 components for temporal RFs and 20 components for chirp responses) and then GMM models with diagonal covariance matrices were fitted while increasing the number of clusters. The numbers of clusters were identified to be 10 for temporal RFs and 20 for chirp responses based on elbow points in the respective Bayesian Information Criteria (BIC) curves. One chirp cluster ($n = 68$ cells) lacked stimulus-evoked responses and was discarded on visual inspection.

Literature

1. Srinivasan, M. V., Laughlin, S. B. & Dubs, A. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* **216**, 427–459 (1982).
2. Doi, E. & Lewicki, M. S. A simple model of optimal population coding for sensory systems. *PLoS Comput. Biol.* **10**, e1003761 (2014).
3. Atick, J. J. & Redlich, A. N. What Does the Retina Know about Natural Scenes? *Neural Comput.* **4**, 196–210 (1992).
4. Karklin, Y. & Simoncelli, E. P. Efficient coding of natural images with a population of noisy Linear-Nonlinear neurons. *Adv. Neural Inf. Process. Syst.* **24**, 999–1007 (2011).
5. Abballe, L. & Asari, H. Natural Image Statistics for Mouse Vision. doi:10.1101/2021.04.08.438953.
6. Qiu, Y. *et al.* Natural environment statistics in the upper and lower visual field are reflected in mouse retinal specializations. *Curr. Biol.* (2021) doi:10.1016/j.cub.2021.05.017.
7. Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423 (1948).
8. Barlow, H. B. & Others. Possible principles underlying the transformation of sensory messages. *Sensory communication* **1**, (1961).
9. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954).
10. Laughlin, S. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch. C* **36**, 910–912 (1981).
11. Nemenman, I., Lewen, G. D., Bialek, W. & de Ruyter van Steveninck, R. R. Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS Comput. Biol.* **4**, e1000025 (2008).
12. Fairhall, A. L., Lewen, G. D., Bialek, W. & de Ruyter Van Steveninck, R. R. Efficiency and ambiguity in an adaptive neural code. *Nature* **412**, 787–792 (2001).
13. Rieke, F., Bodnar, D. A. & Bialek, W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. Biol. Sci.* **262**, 259–265 (1995).
14. 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).
15. Rieke, F. & Rudd, M. E. The challenges natural images pose for visual adaptation. *Neuron* **64**, 605–616 (2009).
16. Balasubramanian, V. & Sterling, P. Receptive fields and functional architecture in the retina. *J. Physiol.* **587**, 2753–2767 (2009).

17. Roy, S., Jun, N. Y., Davis, E. L., Pearson, J. & Field, G. D. Inter-mosaic coordination of retinal receptive fields. *Nature* **592**, 409–413 (2021).

18. Ratliff, C. P., Borghuis, B. G., Kao, Y.-H., Sterling, P. & Balasubramanian, V. Retina is structured to process an excess of darkness in natural scenes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17368–17373 (2010).

19. Chalk, M., Marre, O. & Tkačik, G. Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 186–191 (2018).

20. Baden, T. *et al.* The functional diversity of retinal ganglion cells in the mouse. *Nature* **529**, 345–350 (2016).

21. Sümbül, U. *et al.* A genetic and computational approach to structurally classify neuronal types. *Nat. Commun.* **5**, 1–12 (2014).

22. Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168–174 (2013).

23. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

24. Masland, R. H. The neuronal organization of the retina. *Neuron* **76**, 266–280 (2012).

25. Baylor, D. A., Nunn, B. J. & Schnapf, J. L. The photocurrent, noise and spectral sensitivity of rods of the monkey Macaca fascicularis. *J. Physiol.* **357**, 575–607 (1984).

26. Schneeweis, D. M. & Schnapf, J. L. The photovoltage of macaque cone photoreceptors: adaptation, noise, and kinetics. *J. Neurosci.* **19**, 1203–1216 (1999).

27. Sabbah, S. *et al.* A retinal code for motion along the gravitational and body axes. *Nature* **546**, 492–497 (2017).

28. Bethge, P. *et al.* An R-CaMP1.07 reporter mouse for cell-type-specific expression of a sensitive red fluorescent calcium indicator. *PLoS One* **12**, e0179460 (2017).

29. Tran, N. M. *et al.* Single-Cell Profiles of Retinal Ganglion Cells Differing in Resilience to Injury Reveal Neuroprotective Genes. *Neuron* **104**, 1039–1055.e12 (2019).

30. Sterratt, D. C., Lyngholm, D., Willshaw, D. J. & Thompson, I. D. Standard anatomical and visual space for the mouse retina: computational reconstruction and transformation of flattened retinae with the Retistruct package. *PLoS Comput. Biol.* **9**, e1002921 (2013).

31. Nadal-Nicolás, F. M. *et al.* True S-cones are concentrated in the ventral mouse retina and wired for color detection in the upper visual field. *Elife* **9**, (2020).

32. Kim, I.-J., Zhang, Y., Yamagata, M., Meister, M. & Sanes, J. R. Molecular identification of a retinal cell type that responds to upward motion. *Nature* **452**, 478–482 (2008).

33. Rossi, L. F., Harris, K. D. & Carandini, M. Spatial connectivity matches direction selectivity in visual cortex. *Nature* **588**, 648–652 (2020).

34. Barlow, H. B., Fitzhugh, R. & Kuffler, S. W. Change of organization in the receptive fields of the cat's retina during dark adaptation. *J. Physiol.* **137**, 338–354 (1957).

35. Młynarski, W. F. & Hermundstad, A. M. Efficient and adaptive sensory codes. *Nat. Neurosci.* (2021) doi:10.1038/s41593-021-00846-0.

36. Laughlin, S. B. The role of sensory adaptation in the retina. *J. Exp. Biol.* **146**, 39–62 (1989).

37. Lettvin, J. Y., Maturana, H. R., McCulloch, W. S. & Pitts, W. H. What the Frog's Eye Tells the Frog's Brain. *Proceedings of the IRE* **47**, 1940–1951 (1959).

38. Gollisch, T. & Meister, M. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* **65**, 150–164 (2010).

39. Joesch, M. & Meister, M. A neuronal circuit for colour vision based on rod–cone opponency. *Nature* **532**, 236–239 (2016).

40. Szatko, K. P. *et al.* Neural circuits in the mouse retina support color vision in the upper visual field. *Nat. Commun.* **11**, 1–14 (2020).

41. de Malmazet, D., Kühn, N. K. & Farrow, K. Retinotopic Separation of Nasal and Temporal Motion Selectivity in the Mouse Superior Colliculus. *Curr. Biol.* **28**, 2961–2969.e4 (2018).

42. Chen, H., Savier, E. L., DePiero, V. J. & Cang, J. Lack of Evidence for Stereotypical Direction Columns in the Mouse Superior Colliculus. *J. Neurosci.* **41**, 461–473 (2021).

43. Baden, T. *et al.* A tale of two retinal domains: near-optimal sampling of achromatic contrasts in natural scenes through asymmetric photoreceptor distribution. *Neuron* **80**, 1206–1217 (2013).

44. Peichl, L., Künzle, H. & Vogel, P. Photoreceptor types and distributions in the retinae of insectivores. *Vis. Neurosci.* **17**, 937–948 (2000).

45. Holmgren, C. D. *et al.* Visual pursuit behavior in mice maintains the pursued prey on the retinal region with least optic flow. *Elife* **10**, (2021).

46. Heukamp, A. S., Warwick, R. A. & Rivlin-Etzion, M. Topographic Variations in Retinal Encoding of Visual Space. *Annu Rev Vis Sci* (2020) doi:10.1146/annurev-vision-121219-081831.

47. Szél, Á. *et al.* Different patterns of retinal cone topography in two genera of rodents, Mus and Apodemus. *Cell Tissue Res.* **276**, 143–150 (1994).

48. Baden, T., Euler, T. & Berens, P. Understanding the retinal basis of vision across species. *Nat. Rev. Neurosci.*

21, 5–20 (2020).

49. Cadieu, C. F. & Olshausen, B. A. Learning intermediate-level representations of form and motion from natural movies. *Neural Comput.* **24**, 827–866 (2012).
50. Govardovskii, V. I., Fyhrquist, N., Reuter, T., Kuzmin, D. G. & Donner, K. In search of the visual pigment template. *Vis. Neurosci.* **17**, 509–528 (2000).
51. Carter-Dawson, L. D. & LaVail, M. M. Rods and cones in the mouse retina. I. Structural analysis using light and electron microscopy. *J. Comp. Neurol.* **188**, 245–262 (1979).
52. Penn, J. S. & Williams, T. P. A new microspectrophotometric method for measuring absorbance of rat photoreceptors. *Vision Res.* **24**, 1673–1676 (1984).
53. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* 061507 (2017).
54. *ROIMatchPub: Match suite2p ROIs between imaging sessions*. (Github).
55. Mazurek, M., Kager, M. & Van Hooser, S. D. Robust quantification of orientation selectivity and direction selectivity. *Front. Neural Circuits* **8**, (2014).
56. Sondereker, K. B., Stabio, M. E., Jamil, J. R., Tarchick, M. J. & Renna, J. M. Where You Cut Matters: A Dissection and Analysis Guide for the Spatial Orientation of the Mouse Retina from Ocular Landmarks. *J. Vis. Exp.* (2018) doi:10.3791/57861.

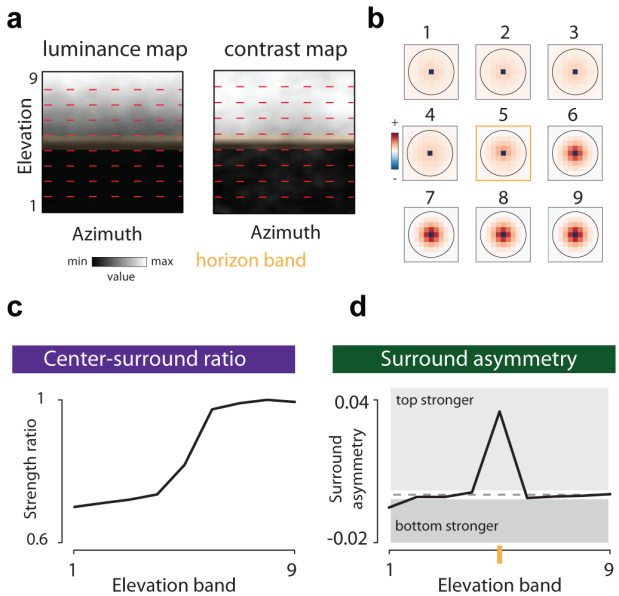
Acknowledgments

We thank Hiroki Asari for sharing the dataset of naturalistic images, Anton Sumser for sharing visual stimulus code, Yoav Ben Simon for initial explorative work with the generation of AAVs, and Tomas Vega-Zuñiga for help with immunostainings. We also thank Gasper Tkacik and members of the Neuroethology group for comments on the manuscript. This research was supported by the Scientific Service Units of IST Austria through resources provided by Scientific Computing, the Preclinical Facility, the Lab Support Facility, and the Imaging and Optics Facility. This work was supported by the European Research Council Starting Grant (N° 756502) to M.J. and Marie Skłodowska-Curie Grant (N° 665385) to D.G.

Author Contributions

D.G., W.M. and M.J. designed the study. D.G. and J.S. performed the calcium imaging experiments with help from M.J.. D.G. and O.S. analysed the experimental data with inputs from W.M. and M.J.. W.M. analysed natural scene statistics and developed the normative model. D.G., W.M. and M.J. wrote the manuscript.

Supplemental Information



Suppl. Fig. 1. Predictive coding predictions derived from a previously proposed model. a. *Average luminance (left panel) and local contrats (right panel) in the mouse field of view in the ultraviolet range. Red dashed lines separate elevation bands. Orange opaque rectangle denotes the horizon band. Natural image data - courtesy of Hiroki Asari [5]. **b.** Receptive fields optimized for each band in **a**. **c.** Relative surround-to-center strength as a function of the elevation band. **d.** Surround asymmetry as a function of elevation band. Orange mark denotes the horizon band.*

Here we follow the derivation and assumptions proposed in [1]. The receptive field model in [1] assumes that the center pixel $s_{t,cent}$ of the t-th stimulus $\vec{s}_t$ is subtracted from its linear prediction computed from the surround $\vec{s}_{t,sur}$. Instead of encoding the raw value of the central pixel, the model RGC encodes the difference between this prediction and the center in order to minimize the dynamic range of its output. The optimal prediction weights $\vec{w}$ are optimized to minimize the mean squared error:

$$E(w) = \left\langle (s_{t,cent} - \vec{w}^T \vec{s}_{t,sur})^2 \right\rangle_t$$

where T denotes vector transposition.

The optimal vector of surround weights $\vec{w}$ is a solution to the following equation:

$$R_{-c}\vec{w} = \vec{R}_c$$

where $R_{i,j} =< s_{t,i} s_{t,j} >_t$ is the spatial autocorrelation of natural images, and i, j index pixels within an image patch, $\vec{R}_c$ is the autocorrelation vector of the center pixel with all other pixels, and R_{-c} is the square correlation matrix of all pixels without the center pixel.

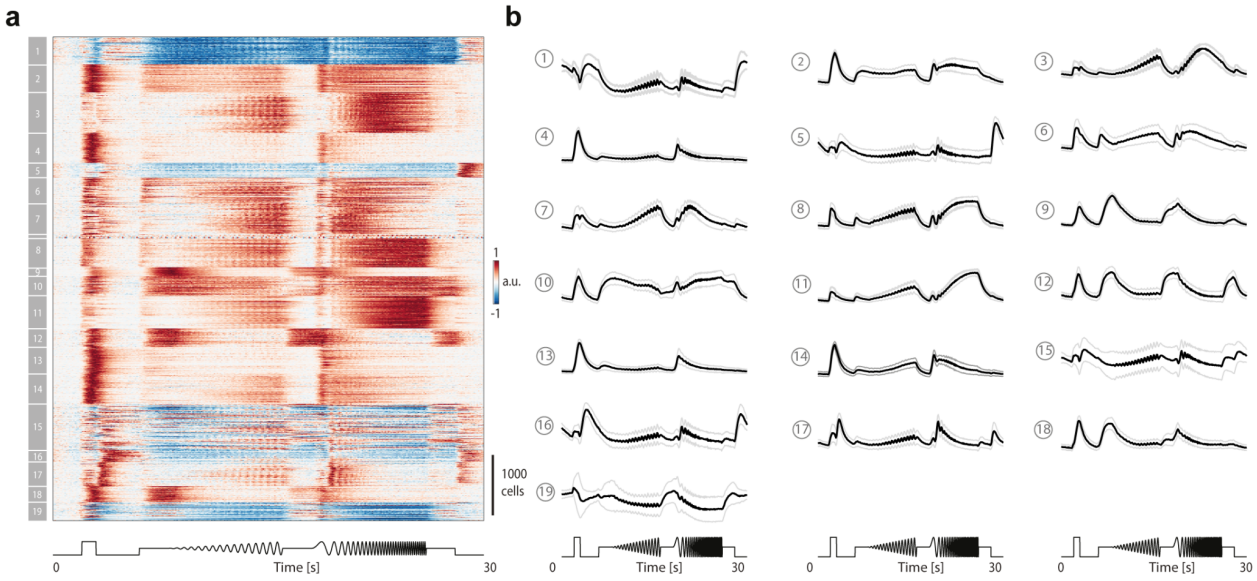
The correlation function $R_{i,j}$ is approximated analytically as:

$R_{i,j} = M_i M_j + S_i S_j \exp[-\frac{d(i,j)}{D}]$, for $i \neq j$ and,

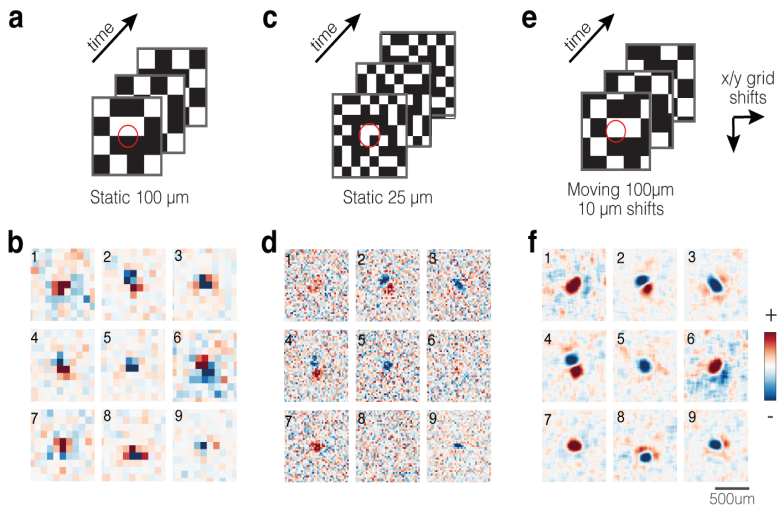
$R_{i,i} = M_i^2 + S_i^2 + N^2$, for $i = j$,

Where M_i and S_i are mean and standard deviation of the i-th entry of the image intensity respectively, $d(i,j)$ is the Euclidean, spatial distance between entries labeled i and j, D is a constant controlling the decay of the correlation, and N is the standard deviation of the noise. The term $N_i N_j$ vanishes for $i \neq j$ because noise is assumed to be uncorrelated.

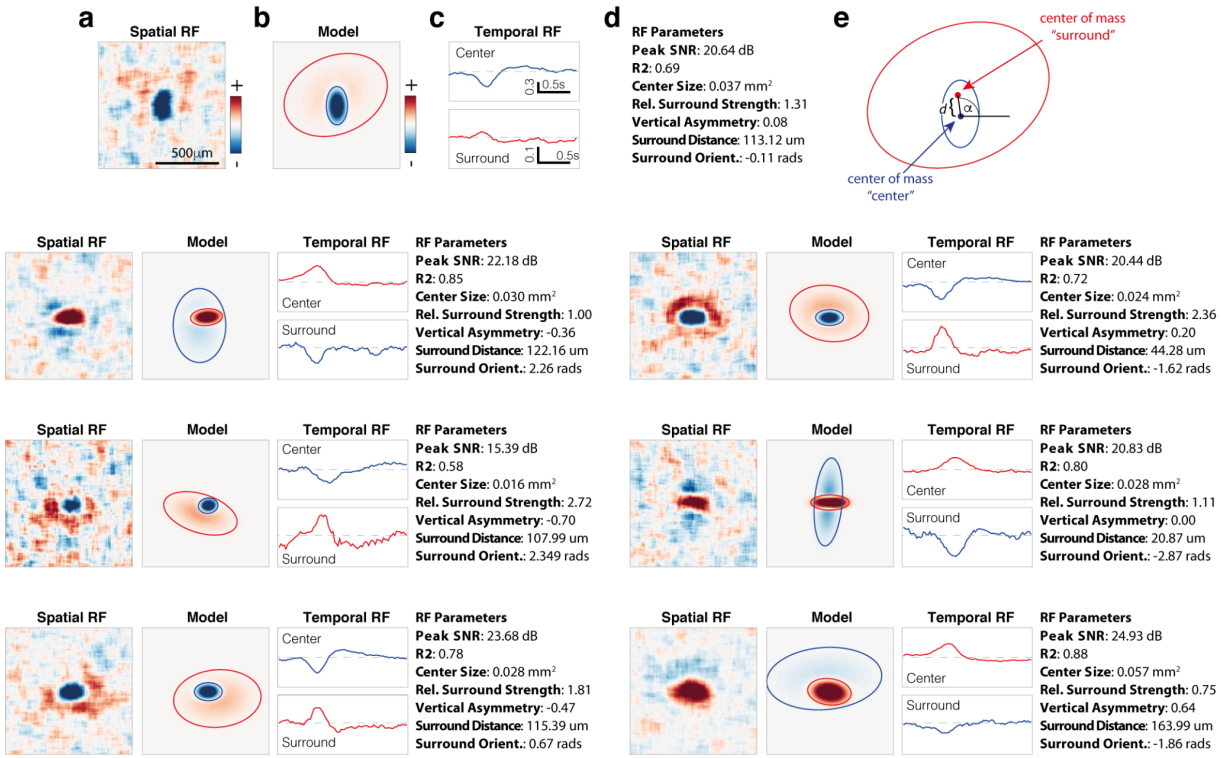
To approximate the spatial autocorrelation function of natural images as a function of elevation within the visual field, we created a dataset of images with simulated horizon as described in the Methods. We then divided each of these images into uniformly separated horizontal bands. We sampled square image patches within each band. We then computed mean vectors $\vec{M}^y$ and standard deviation vectors $\vec{S}^y$, where the upper index y indicates the elevation band. Individual entries of these vectors corresponded to mean and variance of pixel values within the elevation band y respectively. We assumed constant values of the decay constant D and noise standard deviation N. Ratio of surround-to-center strength and surround asymmetry were computed as described in the Methods. We note that our results do not depend qualitatively on parameter choice and reveal similar trends across a broad range of parameter values.



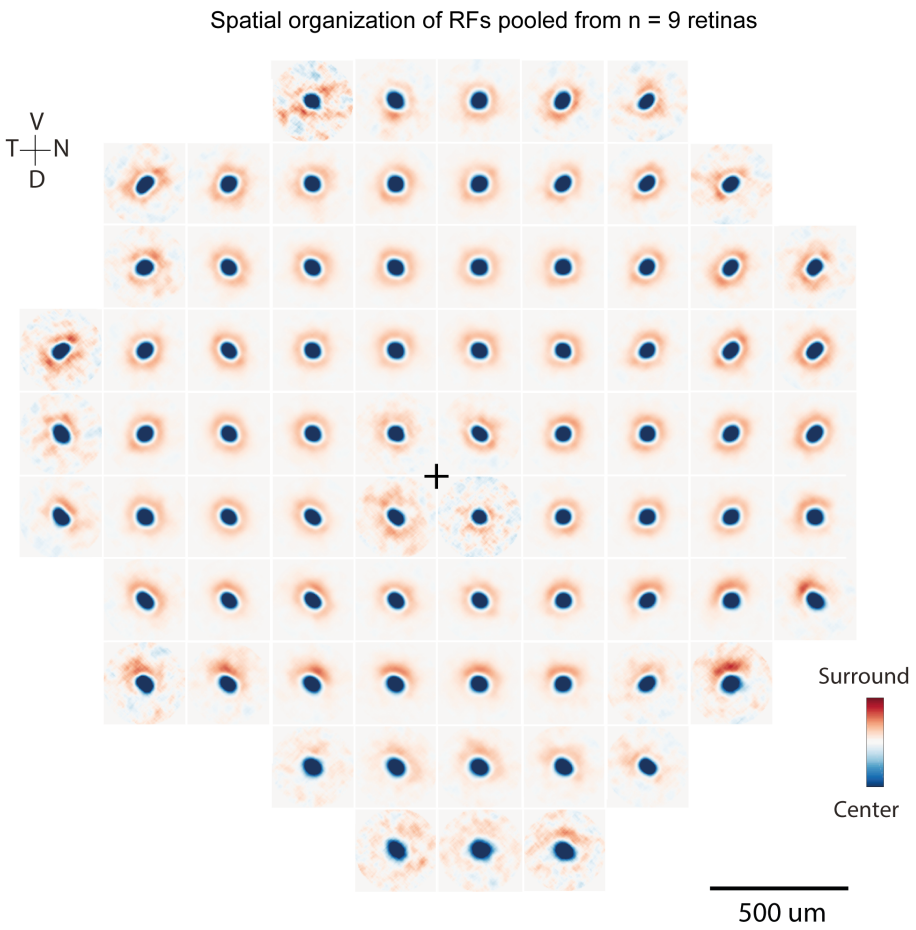
Suppl. Fig. 2. Clustering of Ca^{2+} signals to "chirp" stimulus shows segregation into functional types. ***a****. normalized Ca^{2+} responses to changes in frequency, contrast, and luminance, known as "chirp" stimulus from RGCs, sorted by cluster ids, as determined by GMM.* ***b****. Mean and standard deviation of each cluster reveal ON (cluster 2, 4, 8, 9, 10, 11, 13, 14, 18), OFF (5, 12, 15, 16), ON-OFF (3, 6, 7, 17), suppressed-by contrast (1, 19) RGC-types, as well as differences in frequency tuning to slow (e.g., cluster 4), mid (e.g., cluster 3), fast (e.g., cluster 11) and all frequency modulations (e.g., cluster 2) or difference in luminance sensitivity (e.g., compare ON responsive cluster 14 and 18). Note: clusters 15 and 19 are not homogeneous, as seen in their large standard deviations.*



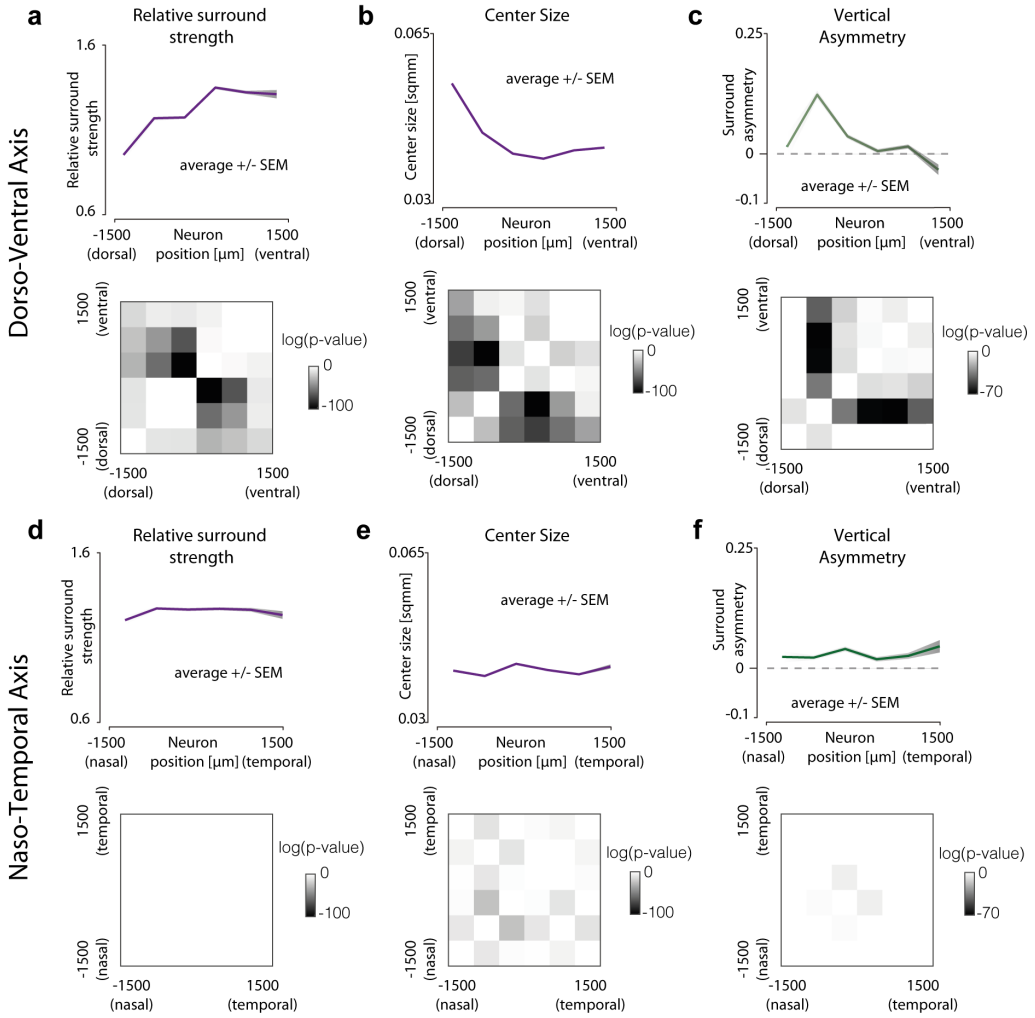
Suppl. Fig. 3. Comparison of RFs recovered from three different white-noise stimuli a. *Schematic of binary white noise stimuli used for recovering RFs during imaging experiments for static checkers with grid size 100 x 100 um^2.* ***b****. Spatial RFs of 9 representative neurons were generated from the stimuli above.* ***c-d.*** *Same as* ***a-b*** *for the same neurons, but for static 25 µm sized checkers.* ***e-f****. Same as in* ***a-b*** *for the same neurons but for a shifting grating - moving checkers with grid size 100 x 100 µm^2, where the entire grid was shifted by random multiples of 10 µm in both x- and y-axis. While the static 100 µm checkers were too low resolution for automatic analysis and the static 25 µm checkers were unable to drive many neurons strongly enough to elicit sufficient responses for RF reconstruction, the moving white noise stimulus was able to unambiguously recover the most detail in the center-surround structure of RFs.*



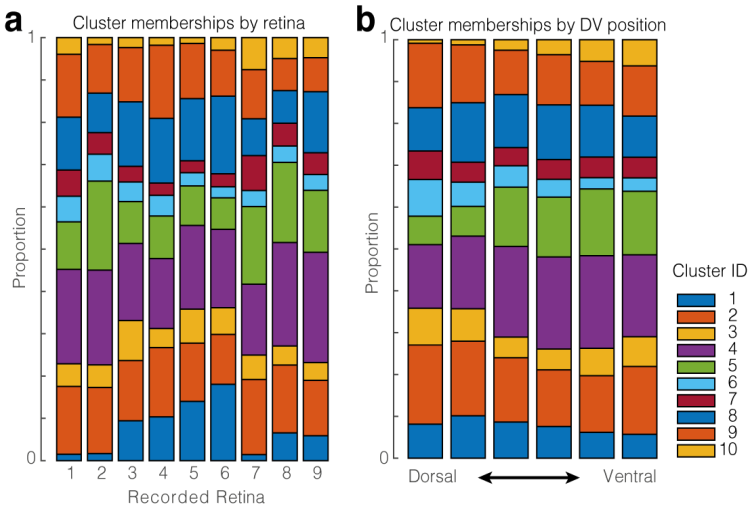
Suppl. Fig. 4. Receptive field parametrizations for representative neurons. a. *Spatial receptive field snapshot at the peak center strength.* ***b****. Difference of Gaussians model fitted to the spatial RF. Ellipses represent 2SD of the two Gaussians.* ***c****. Temporal trace of the mean value of pixels within the respective Gaussians (top: center, bottom: surround). Dashed line represents no correlation between stimulus and response.* ***d****. Values of different parameters of the RF reconstruction (Peak SNR), the goodness of fit (R2), center-surround structure (Center Size, Rel. Surround Strength) and eccentricity of surround (Vertical Asymmetry, Surround Distance, Surround Orientation). All neurons are plotted with the same scale and limits (after normalization).* ***e.*** *Schematic depicting the parametrization of the receptive field in (****a****), showing the surround distance (d) and orientation (****α****).*



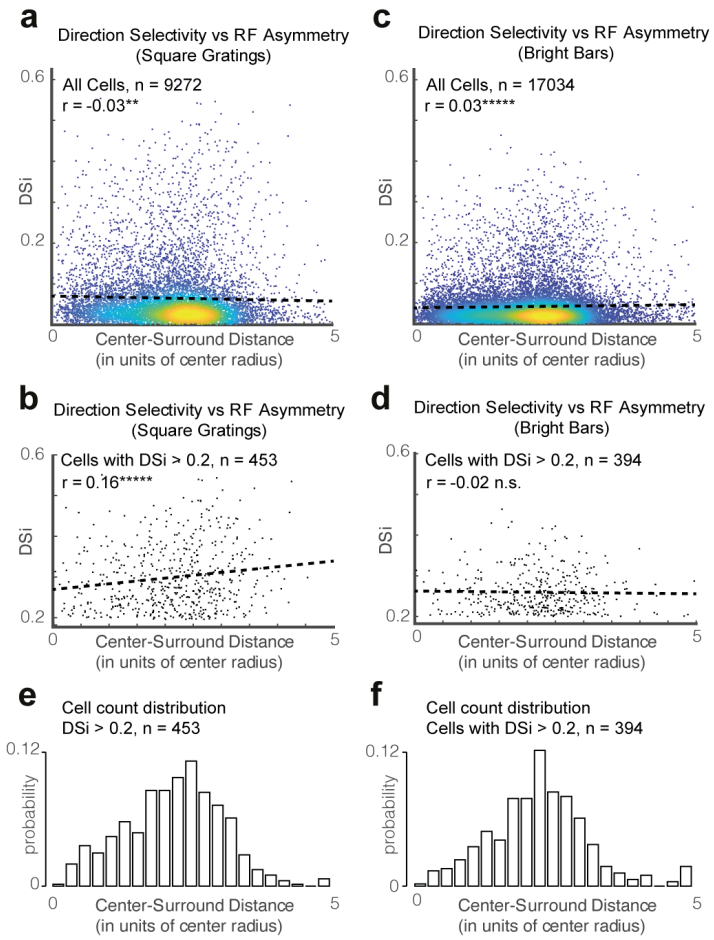
Suppl. Fig. 5. Spatial structure of average RFs across the retina. *Average spatial RFs of all RGCs in square bins of size 300 μm at different positions of the retinal surface. As in Fig. 3a, but including cells from all 9 retinas. N = 344 ± 288 cells per bin. Black cross: optic nerve head position.*



Suppl. Fig. 6. Homogeneity of receptive field architecture in the temporal-nasal axis. a. *Top: Mean relative surround strength at 6 different dorsoventral positions. Bottom: p-values of 2 sample Kolmogorov-Smirnov tests (with Bonferroni correction) between all pairs of bins. Darker colors represent higher significance levels that the cells in the two corresponding bins have different surround strengths.* ***b****. Same as a, but for trends and significance levels across the naso-temporal axis.* ***c*** *and* ***d.*** *Same as* ***a*** *and* ***b****, respectively, but for center sizes.* ***e*** *and* ***f****. Same as* ***a*** *and* ***b****, respectively, but for Vertical surround asymmetry.*



Suppl. Fig. 7. Proportions of cluster membership. a*. Fraction of cells from each retina that were classified into each of the 10 temporal RF clusters (from Fig. 4).* ***b****. Fraction of cells across RF clusters within bins of width 500 μm.*



Suppl. Fig. 8. Correlation between direction selectivity and surround asymmetry. a. Scatter plot of Direction selectivity index (DSi) for drifting square gratings stimulus vs bright bar. Center-Surround distance measured from spatial RFs. Points are color-coded by local density to visualize overlapping points and the dashed black line shows the linear least-squares fit between the two quantities for all cells (p-value for correlation coefficient: 0.0058). **b**. Same as **a**, but for cells with DSi > 0.2 (p-value: 2.82 x 10^{-4}). **c**. same as **b**, but for the density distribution. **d**-**f** Same as **a-c**, but for a moving bright bar. (p-values: (**d**) 3.3 x 10^{-4}, (**e**) 0.65).