

Distinct Laminar Processing of Local and Global Context in Primate Primary Visual Cortex

Maryam Bijanzadeh^{1,2}, Lauri Nurminen¹, Sam Merlin^{1,3}, Andrew M. Clark¹
and Alessandra Angelucci^{1,*}

¹Dept. of Ophthalmology and Visual Science
Moran Eye Institute
University of Utah, Salt Lake City, UT 84132, USA

²Present address: Department of Neurological Surgery, UCSF, CA 94143, USA.

³Present address: Medical Science, School of Science and Health, Western Sydney
University, Campbelltown, NSW 2560 Australia

* Corresponding author and Lead Contact.

Corresponding author's address:

65 Mario Capecchi Drive
Salt Lake City, UT 84132, USA

Tel: (801) 5857489

Email: alessandra.angelucci@hsc.utah.edu

SUMMARY

Visual perception is affected by spatial context. In visual cortex, neuronal responses to stimuli inside the receptive field (RF) are suppressed by stimuli in the RF surround. To understand the circuits and cortical layers processing spatial context, we simultaneously recorded across all layers of macaque primary visual cortex, while presenting stimuli at increasing distances from the recorded cells' RF. We find that near vs. far surround stimuli activate distinct layers, thus revealing unique laminar contributions to the processing of local and global spatial context. Stimuli in the near-surround evoke the earliest subthreshold responses in superficial and upper-deep layers, and earliest suppression of spiking responses in superficial layers. Conversely, far-surround stimuli evoke the earliest subthreshold responses in feedback-recipient layer 1 and lower-deep layers, and earliest suppression of spiking responses almost simultaneously in all layers, except 4C, where suppression emerges last. Our results, suggest distinct circuits for local and global signal integration.

KEYWORDS: V1, macaque, cortical layers, feedback, horizontal connections, lateral interactions, surround suppression, LFP, CSD, linear arrays.

INTRODUCTION

The mammalian neocortex consists of six interconnected layers with distinct functional properties and input/output connections. This architecture is a defining feature of all neocortex, and, thus, likely critical to cortical information processing. To understand laminar processing, here we study a canonical cortical computation found across sensory modalities and species, surround suppression (SS) (reviewed in: Allman et al., 1985; Angelucci and Shushruth, 2013), in a cortical area whose laminar connectivity and neuronal response properties are well understood, i.e. the macaque primary visual cortex (V1) (Callaway, 2014; Hubel and Wiesel, 2004). Understanding the role of V1 layers in SS has the potential to reveal generalizable principles of laminar computation.

SS is a form of contextual modulation, whereby neurons change their responses to stimuli inside their receptive field (RF) depending on spatial context, i.e. stimuli simultaneously presented outside the RF (Angelucci and Shushruth, 2013; Blakemore and Tobin, 1972; Hubel and Wiesel, 1965). These modulatory effects are typically suppressive when the stimuli in the RF and surround have similar properties, e.g. stimulus orientation. In visual cortex, SS is thought to contribute to segmentation of object boundaries and visual saliency (Knierim and Van Essen, 1992; Nothdurft et al., 2000; Nurminen and Angelucci, 2014; Petrov and McKee, 2006).

Single-unit electrophysiology has revealed laminar differences in the properties of SS (Henry et al., 2013; Ichida et al., 2007; Sceniak et al., 2001; Shushruth et al., 2009, 2013). However, in these previous studies neural responses were not simultaneously recorded across layers. Thus, it remains unknown how neurons in different cortical layers integrate visual signals from the surround, and in which layers SS first emerges. Because cortical layers exhibit different

afferent connectivity, understanding which layers first integrate signals from the RF surround and generate SS can reveal the circuitry underlying contextual integration. Specifically, in macaque V1, feedforward afferents from the lateral geniculate nucleus (LGN) terminate primarily in layer (L)4C (Blasdel and Lund, 1983); millimeters-long horizontal connections are present in all V1 layers, except mid- and lower 4C, but are most prominent in layers 2/3 and 5, and weaker in 4B-upper4C α and 6 (Lund et al., 2003; Rockland and Lund, 1983); feedback afferents from higher visual areas terminate primarily in V1 layers 1-2A and 5B-6 (Federer et al., 2015; Rockland and Pandya, 1979) (**Fig. 1A**). This laminar specificity of afferents to V1 offers the possibility of gaining insights into the circuits that initiate SS, as the circuits carrying surround signals to a V1 column must evoke the earliest pre- and post-synaptic depolarization in the V1 layers where they terminate.

Current evidence suggests that feedforward, horizontal and feedback connections all contribute to contextual integration and SS, but at different spatial scales and in different cortical layers (Angelucci et al., 2017). We hypothesized that three connection types, and therefore multiple layers, contribute to the processing of local context, i.e. the influence of stimuli adjacent to the RF (the near surround). Instead, only feedback connections and, therefore, feedback-recipient layers, should underlie the processing of global context, i.e. the influence of stimuli in the far RF surround (**Fig. 1B**).

To understand the contribution of different V1 layers and circuits to spatial integration and SS, we recorded simultaneously through all layers of macaque V1 the local field potential (LFP) and multiunit spiking activity (MUA) evoked by visual stimuli presented at increasing distances from the recorded neurons' RFs, and measured the onset latency of subthreshold

responses and of SS of spiking responses across layers. We found that distinct layers, and therefore distinct circuits, underlie the processing of local and global spatial context.

Figure 1 about here

RESULTS

We recorded visually evoked LFP and MUA using 24-channel linear electrode arrays (100 μ m electrode spacing) oriented perpendicular to the cortical surface of area V1 in anesthetized macaque monkeys (see STAR Methods). Verticality of the array was verified *in vivo* by the alignment of RF location and similarity of orientation tuning functions across layers (**Fig. 2A**), and confirmed by postmortem histology (**Fig. 2B**). We present results from neural recordings in 4 macaques from 10 penetrations deemed to be perpendicular to the V1 surface by these criteria.

We used current source density (CSD) analysis of LFP signals (Mitzdorf, 1985) to identify laminar boundaries, in particular the location of V1 input L4C, aided by spiking activity for identifying the top and bottom of the cortex. CSD, defined as the second spatial derivative of LFP signals, produces a map of current sinks (negative voltage fluctuations) and sources (positive fluctuations) as a function of time; input L4C can be identified as the location of the earliest current sink followed by a reversal to current source in response to RF stimulation (**Fig. 2C**) (Schroeder et al., 1998). The top of the cortex (L1) was identified as one contact above the uppermost contact exhibiting visually-driven spiking responses, and the L6/white matter boundary as the deepest contact at which visually-driven spike rates dropped by >50%

compared to responses at contacts immediately above (**Fig. 2A**). The boundaries between other layers were estimated based on previous anatomical studies (Lund, 1973) and postmortem histology, therefore, these boundaries are necessarily tentative. Specifically, we identified upper-L4 as extending from the uppermost part of L4C to approximately 1-2 contacts above the top of 4C; this region likely encompassed layers upper-4C α and 4B, and possibly L4A (**Figs. 1, 2C**). L2/3 were the layers between upper-L4 and L1, and the deep layers those below L4C, within which we identified upper (U) and lower (L) halves as approximations for L5 and L6, respectively (**Figs. 1, 2B-C**). We also refer to all layers above 4C as superficial.

Figure 2 about here

To investigate the laminar processing of surround signals, we recorded neuronal responses through a V1 column to visual stimuli presented at increasing distances from the columnar RF; this experimental design was motivated by previous studies implicating different circuits in the processing of near vs. far surround stimuli (Angelucci et al., 2002, 2017) (**Fig. 1B**). To identify the circuits carrying visual signals from the surround to the recorded column, we calculated the onset latency of current sinks in the CSD across layers, following stimulation of the surround only. In the absence of direct RF stimulation, surround stimuli do not evoke significant spiking responses, therefore the LFP/CSD largely reflects presynaptic activity and postsynaptic subthreshold responses evoked by surround stimulation (Buzsaki et al., 2012). This approach, i.e. measuring the onset latency of current sinks that reflect the net post-synaptic potentials of local neurons (Schroeder et al., 1991; Tenke et al., 1993), is better suited to localize laminar activation than measuring onset latency of spiking responses; this is because the

dendritic arbors of cortical neurons, where most synaptic integration occurs, rarely co-localize with the soma, where spikes are initiated (Callaway, 2014).

Near and far surround stimuli evoke distinct laminar patterns of CSD signals

Figure 3 about here

After mapping the location and size of the minimum response field (mRF) and the stimulus preferences of neurons across the recorded V1 column (see STAR Methods), we recorded LFP and MUA in response to 0.5° black square stimuli flashed at increasing distances from the columnar mRF. Distances from the mRF center to the inner edge (i.e. the edge closest to the mRF center) of the square stimulus ranged from 0° to 1.25° , and the mRF surround was systematically probed. Since these small stimuli failed to evoke reliable LFP signals when presented beyond these distances, we also used larger stimuli, i.e. sinusoidal annular gratings (2° in width) flashed at distances from the mRF center ranging from 0.2° to 6° (measured from the mRF center to the inner edge of the annulus). This larger stimulus evoked robust LFP signals from the far surround, but, unlike the small square stimulus, often also evoked spiking responses across the column when presented $<1.2^\circ$ from the mRF center. Therefore, we restricted our analysis to annular stimuli that did not encroach into the recorded column RF, i.e. those from 1.2° to 6° from the mRF center. To localize the LFP signals to specific layers, we performed CSD analysis and computed the onset latency of current sinks across layers (see STAR Methods).

Figure 3A-D shows the CSD and MUA from an example penetration. Small square stimuli inside the columnar mRF (**Fig. 3A** top) evoked the fastest CSD current sink in L4C, followed by sinks in deep and then superficial layers (**Fig. 3A** middle). The earliest evoked spiking activity, in response to this stimulus, occurred in L4C and deep layers (**Fig. 3A** bottom). Early activation of L4C by RF stimulation can be explained by feedforward activation of this layer, where geniculocortical afferents predominantly terminate (**Fig. 1A**), and is consistent with previous findings (Schroeder et al., 1998).

As the square stimulus was progressively moved away from the columnar mRF, spiking activity ceased first in L4C, then in superficial layers, and finally in deep layers, consistent with previous reports of larger mRFs in deep layers (Gilbert, 1977; Hubel and Wiesel, 1977). When this stimulus reached a distance of just $>1^\circ$ from the mRF center (e.g. **Fig. 3B** top), CSD signals in L4C were delayed, and the earliest current sinks were observed in L2/3 and deep layers almost at the same time (**Fig. 3B** middle). To ascertain that these early CSD responses were evoked by stimulation of the near surround, rather than from direct activation of the larger RFs in superficial and deep layers, we examined how the location of the surround stimuli related to the boundaries of the RF of neurons across the recorded column. To this purpose, and because RF size varies depending on the methods used to map it (Angelucci and Bressloff, 2006; Angelucci et al., 2002), we measured the size of the mRF as well as of the summation RF (sRF: see STAR Methods for definitions) across contacts. In the example recording of **Fig. 3B**, the inner edge of the square stimulus was located 1.25° from the center of the columnar RF, corresponding to 0.5° and 0.25° outside the edge of the largest mRF and sRF, respectively, recorded across the column. This near surround stimulus did not evoke significant spiking responses across the column (**Fig. 3B** bottom), therefore the CSD sinks it evoked reflect subthreshold responses.

In the same example penetration, an annular grating located 3° from the mRF center (2.25° and 2° outside the edge of the largest mRF and sRF, respectively; **Fig. 3C** top) evoked a similar laminar pattern of CSD signals as that evoked by the small square stimulus, i.e. earliest activation of superficial and deep layers, and delayed activation of L4C. However, unlike the small square, the annulus evoked more robust activation of upper-L4. Early CSD signals in all layers, but 4C, evoked by stimuli in the near surround suggests involvement of multiple circuits in the processing of local context, including horizontal connections in superficial and deep layers, and possibly feedback connections to L1 and deep layers (**Fig. 1A**; see Discussion).

Stimulation of the far surround with an annular grating located 5.5° from the mRF center (4.75° and 4.5° outside the outer edge of the largest mRF and sRF, respectively) evoked the earliest CSD signals in feedback-recipient layers, i.e. L1 and the lower deep layers, with large increases in signal latency in the remaining layers (**Fig. 3D** middle). This suggests that visual signals in the far surround are relayed to the recorded (center) V1 column by feedback connections (**Fig. 1A**; see Discussion).

The CSD grand averages for the population of penetrations (**Fig. 3E-H**; see STAR Methods) resembled the profiles of the example case in **Fig. 3A-D**.

Figure 4 about here

We quantified the onset latency of current sinks across layers for the population. To determine which layers were first activated by each stimulus, for each stimulus condition we computed the shortest onset latency of current sinks in each layer for each penetration, and then generated layer-by-layer distributions of these latencies across penetrations and stimulus

conditions (**Fig. 4A-D** left column). **Table S1** reports population medians and means for this data. As absolute latencies can be variable, and because we wished to determine which layers show the earliest current sinks, we also computed for each stimulus condition a relative latency (ΔL) as the difference between the shortest onset latency of currents sinks within each layer and the shortest latency across all layers in the same penetration for that condition. Thus, lower values of ΔL for a layer indicate shorter onset latency in that layer. The layer-by-layer distribution of ΔL values for the population is shown in **Fig. 4A-D** (middle column).

Stimuli inside the RF ($n = 17$ conditions) evoked the earliest current sinks in L4C (mean $\Delta L \pm \text{s.e.m.} = 0 \pm 0 \text{ms}$); onset latency in this layer was significantly shorter than in L1 ($23.44 \pm 1.55 \text{ms}$, $p < 0.0001$; Kruskal-Wallis test with Bonferroni correction), L2/3 ($23.44 \pm 3.09 \text{ms}$, $p < 0.0001$) and the upper and lower halves of the deep layers (Deep_U: $14.29 \pm 1.96 \text{ms}$, $p = 0.0021$; Deep_L: $61 \pm 6.68 \text{ms}$, $p < 0.0001$), but it was not significantly different from latency in upper-L4 ($16.91 \pm 6.96 \text{ms}$, $p = 0.43$) (**Fig. 4A**; **Table S1**).

Small square stimuli in the near surround ($n = 13$ located 0.75° - 1.25° from the mRF center), instead, evoked delayed current sinks in L4C (mean $\Delta L \pm \text{s.e.m.} = 71.58 \pm 8.89 \text{ms}$); onset latency in this layer was significantly longer than in all other layers (mean $\Delta L \pm \text{s.e.m.}$, L1: $10.88 \pm 2.21 \text{ms}$, $p = 0.0262$; L2/3: $6.07 \pm 1.39 \text{ms}$, $p = 0.0005$; Deep_U: $3.65 \pm 1.68 \text{ms}$, $p < 0.0001$; Deep_L: $21.63 \pm 6.25 \text{ms}$, $p = 0.0262$), except upper-L4 ($73.54 \pm 12.24 \text{ms}$, $p = 1.0$) (**Fig. 4B** and **Table S1**), but onset latency in these layers was only significantly earlier than in L4C ($p < 0.0001$, $p = 0.0005$, respectively) and upper-L4 ($p < 0.001$, $p = 0.0017$, respectively), but not significantly earlier than in the remaining layers.

To ascertain that the stimuli intended to activate only the surround did not also directly activate the RFs' fringes across the column, we estimated the location of the square stimuli in the near surround relative to the edge of the mRF of neurons in different layers (**Fig. 4B** right panel). This was expressed as “normalized mRF-surround gap”, defined as

$$G/mRF \text{ radius}$$

where G is the gap between the inner edge of the surround square and the outer edge of the mRF, and $mRF \text{ radius}$ is the radius of the mRF. Thus, a normalized mRF-surround gap equal to zero indicates that the inner edge of the surround stimulus abutted the mRF edge, and negative (positive) values that the stimulus was inside (outside) the mRF. The distribution of normalized mRF-surround gaps in **Fig. 4B** ranges from 0 to 4 (median=1 for all layers), indicating that the surround stimuli were located outside the columnar mRF. We calculated a similar metric (“normalized sRF-surround gap”) quantifying the location of near surround stimuli relative to the edge of the columnar sRF (**Fig. S1A**). This analysis revealed that, for most contacts the near surround stimuli were also located outside the columnar sRF, but for a few contacts (8/138) in the superficial and deep layers they did activate the sRF fringes. However, repeating the same analysis after discarding stimulus conditions with negative values of normalized sRF-surround gap produced similar CSD profiles and quantitative results as in **Figs. 3F and 4B (Fig. S1B-C)**, indicating that these profiles were indeed evoked only by near surround stimuli rather than by direct activation of few neuronal RFs.

Similar to the square stimuli, annular stimuli in the near surround (n=7 located 1.2° to 2° from the mRF center, except for 3° in one case) evoked earliest activation of superficial and deep

layers and delayed activation of L4C (**Fig. 4C** left and middle, and **Table S1**), but in contrast to the square stimuli, the annuli also evoked early current sinks in upper-L4 (mean $\Delta L \pm \text{s.e.m} = 5.16 \pm 2.15 \text{ms}$). Upper-L4 was activated nearly as early as L2/3 ($7.64 \pm 4.43 \text{ms}$), and both layers were activated significantly earlier than L4C ($54.12 \pm 6.98 \text{ms}$, $p=0.0245$ and 0.0189 , respectively), but not than the remainder of layers. The normalized mRF-surround and sRF-surround gaps for these stimuli were all >0.1 (**Fig. 4C** right and **Fig. S2**), indicating all surround annuli were located outside the columnar mRF and sRF.

Stimuli in the far surround ($n=11$ located 2.3° - 5.5° from the mRF center, except for 1.42° in one case) evoked the earliest current sinks in L1 (mean $\Delta L \pm \text{s.e.m} = 12.54 \pm 3.69 \text{ms}$), and the lower deep layers ($1.09 \pm 0.73 \text{ms}$); latencies in these layers were significantly shorter than in L4C ($74.15 \pm 6.16 \text{ms}$, $p=0.0025$ and <0.0001 , respectively) and upper-L4 ($75.65 \pm 6.2 \text{ms}$, $p=0.0016$ and <0.0001 , respectively), and latency in Deep_L was also significantly shorter than in Deep_U ($60.5 \pm 6.31 \text{ms}$, $p=0.0008$) and L2/3 ($44.05 \pm 8.97 \text{ms}$, $p=0.0176$) (**Fig. 4D** and **Table S1**). The normalized mRF-surround and sRF-surround gaps for these stimuli (**Fig. 4D** right and **Fig. S3**) indicated that all surround annuli were located outside both the columnar mRF and sRF.

Figure 5 about here

To facilitate comparison of the laminar latency profiles across different stimulus conditions, in **Fig. 5** we plot the absolute onset latencies of current sinks for each layer in response to stimuli in the RF, near and far surround (only for 6 penetrations for which we had data from a full stimulus set). **Table S2** reports mean and median onset latencies for this dataset, as well as the significance values of a pairwise Ranksum test performed across stimulus

conditions for each layer pair. The results in **Fig. 5** and **Table S2** revealed several key facets of the data. First, L4C was activated significantly faster by a stimulus inside the RF compared to any surround stimulus. Second, L2/3 and Deep_U were activated significantly faster by near than far surround stimuli. Third, upper-L4 was activated significantly faster by near annuli than near squares or far annuli; however, all other layers were activated at similar latencies by near-surround squares and annuli. Fourth, L1 and Deep_L were activated at similar latencies by stimuli in the RF, near or far surround, suggesting that activation of L1 and Deep_L by these stimuli may occur via similar circuits (see Discussion). In summary, these data demonstrate that, in V1, stimuli in the near and far surround evoked distinct laminar patterns of CSD signals, suggesting involvement of different circuits and layers in the processing of local and global context (see Discussion).

For the data in **Fig. 5** and **Table S2**, we also performed a pairwise Ranksum test across layers and stimulus conditions (**Tables S3-5**), in order to understand which stimulus condition evoked the fastest current sinks, independent of the layer where these sinks occurred. This analysis also resulted in several interesting findings. First, a current sink in L4C following stimulation of the RF was the fastest occurring evoked event in V1 (**Table S3**). Second, the earliest current sinks in L2/3 and Deep_U evoked by near squares occurred significantly later (by about 13-16 ms; **Tables S2-S3**) than the earliest current sinks in L4C evoked by RF stimulation. Third, the earliest current sinks in L2/3 and upper-L4 evoked by near annuli were delayed by about 10 ms relative to the earliest activation of L4C by RF stimulation, but this time difference was not statistically significant (perhaps due to our small sample size for near-surround annuli). In contrast, current sinks evoked by near annuli in deep layers and L1 occurred significantly later (by about 34-35 ms) than sinks in L4C evoked by RF stimulation (**Tables S2-S3**). Fourth, the

earliest current sinks in L1 and Deep_L evoked by far annuli occurred significantly later (by about 22-28 ms) than sinks in L4C evoked by RF stimulation (**Fig. 5, Tables S2-S3**), and later (by 6-15 ms) than the earliest sinks in L2/3 and Deep_U evoked by near square stimuli (**Fig. 5, Table S2 and S4**). The earliest current sinks in L1 and Deep_L evoked by far annuli were also delayed by about 12-18 ms relative to sinks in superficial layers evoked by near annuli (**Table S2**); however, this difference in latency was not statistically significant (**Table S4**), perhaps indicating that beyond a very local distance from the RF, surround signals are integrated at similar latencies. Finally, there was no statistically significant difference in latency between the earliest sinks evoked by near squares and those evoked by near annuli (**Table S5**).

These results indicate that not only is the processing of near and far surround stimuli initiated in different V1 layers, but it is also temporally distinct, with RF signals processed faster than surround signals, and near surround signals processed faster than far surround signals.

Onset latency of near and far surround suppression across V1 layers

Figure 6 about here

The results presented above point to the circuits that carry near and far surround signals to the V1 center column. SS is likely to be initiated in the layers where these circuits terminate and evoke the earliest postsynaptic responses. To understand how these surround pathways contribute to SS in different layers, we presented visual stimuli simultaneously in the RF and surround, and measured the onset latency of SS of spiking responses (MUA) across layers (see STAR Methods). Here, we used only gratings of different orientations, because the small square

stimuli did not evoke sufficiently strong suppression to allow for reliable latency measurements, and gratings also allowed us to study both orientation tuned and untuned suppression. We used two kinds of gratings (see STAR Methods): a 20° diameter grating patch centered on the columnar RF of optimal orientation for the recorded column (**Fig. 6A**, top inset); and 2°-width annular gratings of optimal or orthogonal-to-optimal orientations flashed at increasing distances from the RF (as above) and presented together with center grating patches matched for orientation and size to the columnar RF (**Fig. 6B-C** top insets).

For the example penetration in **Figure 6**, the peristimulus time histograms (PSTHs) obtained in response to center-only gratings, and iso- or orthogonally-oriented center-surround gratings confirm several previous findings on the strength and orientation tuning of SS. Specifically, the strength of iso-oriented SS (measured as SI) decreases with increasing distance of the surround stimulus from the RF (Sceniak et al., 2001; Shushruth et al., 2009). Moreover, the SI is generally greater for iso-oriented than for orthogonally-oriented gratings, i.e. SS is orientation-tuned, and this is true for both near-SS and far-SS, although far-SS is less tuned than near-SS (Shushruth et al., 2013). Finally, SS is more sharply tuned above L4C (Henry et al., 2013; Shushruth et al., 2013), albeit in the case of **Fig. 6C** tuning of far-SS in the lower deep layers resembled that in L2/3.

For each contact, we measured the onset latency of SS as the time point at which the MUA profile in response to the center-surround stimuli diverged significantly from the MUA profile in response to the center-only stimulus (see STAR Methods). For the example penetration, in response to the 20°-diameter grating patch, the earliest suppression occurred in layers Deep_U, 4C and upper-4, and at the same time as the onset of the center-only response (**Fig. 6A**). In L2/3, suppression occurred later, but it still coincided with the onset latency of the

center-only response in L2/3. The co-occurrence in time of SS with visually evoked responses in V1 strongly suggests that the origin of this suppression is subcortical, and therefore the earliest suppression in V1 is inherited from the LGN as reduced feedforward excitation.

In L4C, when an iso-oriented grating was presented in the near surround (**Fig. 6B**), but did not directly about the RF (thus, presumably located beyond the spread of geniculocortical afferents –**Fig. 1**), the onset of suppression was markedly delayed relative to the onset of the center-only response as well as relative to the suppression onset caused by the large patch stimulus. Moreover, near-SS in L4C was also delayed relative to the onset of near-SS in all other layers, except Deep_L; the earliest near-SS occurred in superficial layers and upper deep layers. These observations suggest that suppression evoked by near surround stimuli located beyond the spread of geniculocortical afferents is not inherited from the LGN, because suppression occurred later than visually evoked responses in geniculate input L4C. Rather this form of suppression is likely generated within V1, outside L4C.

We also measured the onset latency of tuned SS as the time point at which the response curves to iso- and ortho-oriented center-surround stimuli diverged significantly (**Fig. 6B**; see STAR Methods). Tuned near-SS in all layers, except upper-4, appeared at about the same time as the earliest suppression (the point at which the response curves to the center-only and either the iso- or ortho-oriented stimuli diverged), suggesting that the earliest suppression generated within cortex is orientation-tuned.

Far surround annuli evoked the earliest suppression in deep layers, and latest suppression in L4C (**Fig. 6C**). Far suppression in all layers was significantly delayed relative to the onset of center-only responses in the same layers; moreover, in all layers, but Deep_L, far suppression was also delayed relative to suppression caused by large grating patches or near surround annuli

within the same layers. These results suggest that far-SS is also generated intracortically outside L4C, and is delayed relative to near suppression.

Figure 7 about here

We quantified the onset latency of center-only responses and SS across layers for the population, as described above (**Fig. 7 and Table S6**), including only MUA responses that showed suppression ($SI > 0.15$) (see STAR Methods). For each contact in each penetration and stimulus condition we estimated an absolute latency (time after stimulus onset), as well as a relative latency (ΔL), the latter defined as the difference between the absolute onset latency of suppression at that contact and the shortest latency across all contacts in the same penetration.

RF stimulation evoked the earliest spiking in L4C (mean $\Delta L \pm \text{s.e.m.} = 3.5 \pm 1.35 \text{ms}$, $n=12$ contacts), with a significantly shorter latency than in all other layers ($p \leq 0.04$, Kruskal-Wallis test with Fisher's least significant difference correction; **Fig. 7A; Table S6**).

Large (20°) grating patches encompassing the RF and the full extent of the surround evoked the earliest suppression in L4C (mean $\Delta L \pm \text{s.e.m.} = 4 \pm 1.15 \text{ms}$, $n=12$ contacts), followed by upper-L4 ($13.8 \pm 2.48 \text{ms}$, $n=9$) with no significant difference in latency of suppression between these layers (**Fig. 7B; Table S6**). Latency in L4C was significantly faster than in all remaining layers (L2/3: mean $\Delta L \pm \text{s.e.m.} = 22.71 \pm 2.07 \text{ms}$, $n=26$, $p=0.0000$; Deep_U: $16.81 \pm 2.97 \text{ms}$, $n=15$, $p=0.0055$; Deep_L: $32.17 \pm 4.06 \text{ms}$, $n=14$; $p=0.0000$), again suggesting that the earliest suppression is inherited from LGN.

As for the example penetration, for the population iso-oriented near-surround annuli not directly abutting the RF evoked the earliest suppression in upper-L4 (mean $\Delta L \pm \text{s.e.m.} = 8.84 \pm 2 \text{ms}$,

n=12), followed by Deep_U (19.55±5.6ms, n=13), with no significant difference in suppression onset latency between these layers (**Fig. 7C**; **Table S6**), and then by L2/3 (25.46±3 ms, n=25). Suppression in upper-L4 was significantly faster than in L2/3 (p=0.0005), L4C (34.21±7.1 ms, n=11, p=0.0003), and Deep_L (28.22±8.05 ms, n=10, p=0.0081). L4C showed delayed suppression, whereas suppression latency was variable in deep layers. Again, delayed suppression in L4C points to an intracortical origin, outside L4C, of near-SS caused by stimuli located beyond the extent of geniculocortical afferents.

The onset latency of tuned near-SS showed a similar laminar profile as that of iso-oriented near-SS, emerging first in upper-L4 (mean $\Delta L \pm s.e.m. = 12.01 \pm 3.07$ ms, n=12), followed by L2/3 (37.15±7.26ms, n=25), with latency in upper-L4 being significantly faster than in all layers (L2/3: p=0.013; L4C: 38.58±6.38ms, n=11, p=0.0068; Deep_L: 57.61±17.83ms, n=9, p=0.0045), except Deep_U (28.18±6.48ms, n=13) (**Fig. 7D**; **Table S6**).

The deep layers showed the shortest absolute onset latency of far-SS (**Fig. 7E left**; **Table S6**), but their relative onset latency (ΔL) resembled that in superficial layers (**Fig. 7E right**), suggesting that far-SS emerges almost simultaneously in superficial and deep layers. Far-SS was significantly delayed in L4C (mean $\Delta L \pm s.e.m. = 42.48 \pm 12.66$ ms, n=9) compared to all other layers (upper-L4: 14.18±3.7ms, n=12, p=0.0057; L2/3: 14.53±1.83ms, n=23, p=0.008; Deep_U: 8.23±2.67ms, n=9, p=0.0003; Deep_L: 15.76±6.94ms, n=8, p=0.0064). Notably, many contacts in L4C did not show significant far-SS, therefore they did not contribute to the distribution of (or average) onset latency in this layer; however, lack of suppression in L4C further supports the notion that far-SS is generated outside the geniculocortical recipient layer, which lacks feedback and long-range horizontal connections. We did not measure the onset latency of tuned far-SS as significant tuning of far-SS was generally observed only in L2/3 and upper-L4.

As MUA recorded at contacts bordering two cortical layers could be contaminated by spikes from units in adjacent layers, we performed the same analysis as in **Fig. 7**, but excluding contacts at laminar borders. Moreover, we performed a similar analysis of onset latency on spike sorted single units. Both analyses produced qualitatively similar results as those in **Fig. 7 (Figs. S4-S5)**.

In summary, responses to large gratings covering the RF and full surround, near-SS and far-SS showed distinct laminar profiles of onset latencies.

To compare the latency of SS in each layer evoked by different kinds of center-surround stimuli, for the data in **Fig. 7 and Table S6**, we performed a pairwise Ranksum test across stimulus conditions for each layer pair (**Table S6**). **Figure 7 and Table S6** reveal several key results. First, in L4C SS caused by the large grating (termed full-SS) occurred significantly faster than SS caused by near or far annuli, and at the same time as the onset of visual responses to RF stimulation (20.7ms, $p=0.242$). For the reasons explained above, this finding strongly suggests that SS consists of a fast component inherited from the LGN, and slower components generated intracortically outside L4C. Second, in upper-L4 full-SS and near-SS occurred significantly faster than far-SS. Third, although, latencies of tuned near-SS were generally longer than latencies of iso-oriented near-SS, these differences were not significant in any layer. This observation suggests that intracortically generated SS is already orientation-tuned when it first emerges. Finally, in deep layers full-, near- and far-SS occurred at similar latencies.

To understand differences in onset latency of SS caused by different kinds of stimuli irrespective of layer of earliest emergence, for the data in **Fig. 7 and Table S6** we performed a pairwise Ranksum test across layers and stimulus conditions (**Tables S7-S10**). This analysis led to several important results. First, the onset of spiking responses in L4C following RF

stimulation was the fastest evoked event in V1, and this event occurred at a similar latency as full-SS in L4C (**Tables S6-S7**), suggesting an LGN origin of the earliest SS. Second, full-SS in L4C occurred significantly faster than near-SS, tuned-SS or far-SS in any layer (**Table S8**), suggesting the latter three forms of suppression are generated intracortically. Third, near-SS occurred first in upper-L4, about 14 ms after onset in L4C of responses to RF stimulation (**Table S7**), and about 8 ms after the onset of full-SS in L4C (**Table S8**). However, near-SS in upper-L4 was not significantly faster than the earliest far-SS in deep layers (**Table S9**); this suggests that the earliest SS caused by near and far annuli located $>1.2^\circ$ from the RF center occurs at similar latencies, albeit in distinct layers. Fourth, there was no significant difference in the earliest onset latency of suppression between near-SS and tuned near-SS (**Table S10**). Fifth, far-SS occurred first in deep layers, about 16-19 ms after onset in L4C of responses to RF stimulation (**Table S7**), and about 10-13 ms after the onset in L4C of full-SS (**Table S8**).

Lastly, it is important to note that the absolute latencies of CSD signals (**Fig. 4A-D** left and **Tables S1-S2**) cannot be directly compared to the absolute latencies of spiking responses and of SS (**Fig. 7A-E** left and **Table S6**), due to the fundamentally different nature of CSD signals and spikes, and the different methods we used to compute onset latency of the two signals. For example, in L4C the average onset latency of responses to RF stimulation was about 39 ms, for CSD signals (**Table S1**), but 23 ms for spiking responses (**Table S6**). However, when related to the onset latency of center-only responses measured with the same method, the onset latencies of CSD signals evoked by surround stimuli are consistent with the latencies of SS of spiking responses. For example, near annuli evoked the earliest CSD signals in upper-L4 about 12ms after onset in L4C of CSD responses to RF stimulation (**Table S1**); similarly, near-SS occurred first in upper-L4, about 14 ms after onset in L4C of spiking responses to RF stimulation

(**Table S6**). Likewise, the earliest CSD responses to far annuli and far-SS both occurred about 20 ms after onset in L4C of CSD and spiking responses, respectively, to RF stimulation.

DISCUSSION

Figure 8 about here

Laminar Processing and Circuits Underlying Local Context

Stimuli inside the RF evoked the earliest current sinks and spiking responses in L4C (**Fig. 8A**), consistent with initial activation of this layer by LGN afferents, which terminate predominantly in L4C (Blasdel and Lund, 1983) (**Fig. 1A**), and are spatially restricted to the RF size of recipient V1 neurons (Angelucci and Sainsbury, 2006) (**Fig. 1B**).

Stimulation of the surround region just outside the columnar RF, i.e. the near surround, without simultaneous RF stimulation, evoked the earliest current sinks in superficial and deep layers almost at the same time, but only in superficial layers and the upper deep layers was onset latency of near surround signals significantly faster than onset latency of far surround signals in any layer. Current sinks in L4C evoked by near surround stimuli were delayed relative to those in other layers and those in L4C evoked by RF stimulation, thus ruling out any involvement of LGN afferents to L4C in the processing of local spatial context. These results, rather, suggest that near surround signals are conveyed to the center column via intracortical connections outside L4C. Horizontal connections are prominent in L2/3 and 5, and absent in L1 and mid-to-lower L4C, while feedback connections dominate in L1 and the lower deep layers, and absent in L4C (**Fig. 1**). Thus, the earlier activation of L2/3 and upper deep layers by near stimuli located

<1.2° from the RF center compared to far stimuli suggests involvement of horizontal connections in the processing of local context (**Fig. 8C**). Near surround stimuli also evoked early subthreshold responses in L1, however, latencies did not differ from those evoked by far surround stimuli. Early subthreshold activation of L1 by near stimuli suggests involvement of circuits other than horizontal connections (as these are absent in L1), possibly inter-areal feedback (**Fig. 8C**), as also suggested by early and strong activation of lower deep layers where feedback terminations are denser than horizontal connections (Federer et al., 2015; Lund et al., 2003; Rockland and Pandya, 1979). This interpretation implies that feedback connections are very fast, conducting signals to the center column almost as fast as monosynaptic V1 horizontal connections. Indeed, beyond a distance of 1.2° from the RF center, the earliest subthreshold responses and SS evoked by near (1.2°-2°) and far (2.3-5.5°) stimuli occurred at similar latencies, albeit in different layers. This is consistent with previous studies reporting that beyond a local distance, the onset latency of SS is independent of the distance of the surround stimulus from the RF (Bair et al., 2003). However, it is unclear whether the small 0.5° square stimuli were able to significantly activate extrastriate cortex, and therefore feedback neurons.

Alternatively, or in addition, early current sinks in L1 following near surround stimulation could result from recruitment of K1-K2 koniocellular LGN afferents (**Fig. 8C**), whose terminations in L1 are more widespread (up to 1 mm) than those of magno- or parvocellular afferents in L4C and 6, or those of K3-K6 koniocellular afferents in L2/3 (Angelucci and Sainsbury, 2006; Casagrande et al., 2007; Lund, 1973).

SS of spiking responses induced by large (20°) gratings encompassing the RF and the full extent of the surround first emerged in L4C, at the same time as responses to RF stimulation, and much earlier than near- and far-SS evoked by annuli in any layer. This strongly suggests that

stimuli encompassing the extent of LGN afferents in V1 first suppress LGN cells, which show untuned SS (Alitto and Usrey, 2008; Sceniak et al., 2006), resulting in withdrawal of feedforward excitation to V1 cells (**Fig. 8B**). Together with previous findings that untuned suppression in V1 occurs as fast as responses to RF stimulation (Henry et al., 2013; Webb et al., 2005), our results suggest LGN afferents as the source of early untuned suppression in V1.

Introducing a small gap between the RF and near surround stimuli, so that the surround stimulus was likely located beyond the anatomical spread of LGN afferents to the center V1 column, led to delayed SS in L4C relative to other layers and the onset of visual signals in L4C; this points to an intracortical origin of the suppression, outside L4C. Under this stimulus condition, both the earliest SS and orientation-tuned SS first emerged in superficial layers, particularly in upper-L4 (encompassing upper-L4C α to L4A) (**Fig. 8D**). In V1, along the flow of visual information exiting L4C, long-range horizontal connections first appear in L4B/upper-L4C α (Lund et al., 2003). Therefore, it is conceivable that horizontal connections in upper-L4 are activated by near surround stimuli earlier than horizontal connections in downstream layers. Notably, we only observed consistent early current sinks in upper-L4 when presenting the larger near surround annuli, but not the small squares, perhaps because only larger stimuli can reliably activate the weaker horizontal (and feedback) connections in upper-L4.

That orientation-tuned SS also emerges first in superficial layers is consistent with the known orientation-specific organization of horizontal connections in L2/3 (Malach et al., 1993) (albeit we lack this information for horizontal connections in upper-L4), and suggests that the deeper layers may, perhaps, inherit tuned-SS from the superficial layers. Moreover, the similarity in the onset latency of iso-orientation-SS and tuned-SS caused by near annuli not abutting the RF suggests that intracortically-generated near-SS is tuned from its emergence.

In summary, our results suggest that near surround signals arising from beyond the extent of LGN afferents are processed by multiple connections types, including horizontal and, possibly, feedback and/or koniocellular geniculocortical connections (**Fig. 8C-D**). Two recent studies support this interpretation. Optogenetic activation of horizontal connections in mouse V1 mimicked the effects of surround stimulation (Adesnik et al., 2012; Sato et al., 2014), and optogenetic inactivation of V2 feedback connections to V1 reduced suppression in the proximal region of the RF surround (Nurminen et al., 2018).

In conclusion, our results suggest that the processing of local spatial context is initiated in superficial and deep layers via mechanisms involving primarily feedforward connections from LGN and horizontal connections in superficial layers and L5, but possibly also feedback connections (**Fig. 8A-D**).

Laminar Processing and Circuits Underlying Global Context

Stimulation of the far surround without simultaneous RF stimulation evoked the earliest current sinks in L1 (and 2A) and the lower deep layers, where feedback terminations are densest (**Fig. 1**). This suggests that far surround signals are relayed to these layers in the center V1 column via feedback connections, and therefore that far-SS is initiated by feedback (**Fig. 8E**). Far-SS emerged first and almost simultaneously in superficial and deep layers, and latest in L4C, a layer that lacks both horizontal and feedback connections, and whose neurons confine their dendrites within 4C (**Fig. 8F**). Since far surround stimuli evoked the earliest subthreshold responses in L1/2A and lower deep layers, early far-SS throughout the superficial and deep layers could be initiated by feedback contacts with inhibitory cells in L1/2A and L5/6. Inhibitory neurons in L1-2A could suppress pyramidal cells in most layers by contacting their apical

dendrites. Most L2-4B pyramids and many, but not all, L5/6 pyramids, have apical dendrites ascending to L1, while L4C cells' dendrites are confined to 4C (Callaway and Wiser, 1996; Lund, 1973). Inhibitory neurons in L5/6, thus, could suppress neurons in these layers whose apical dendrites do not reach L1/2A, while L4C could inherit late far-SS from other layers (**Fig. 8F**). The synaptic mechanisms generating SS are reviewed in detail in Angelucci et al. (2017).

A role for feedback connections in global contextual integration and far-SS is consistent with evidence that feedback, but not monosynaptic horizontal, connections encompass the full spatial extent of the RF and surround of V1 neurons (Angelucci et al., 2002) (**Fig. 1B**), and conduct signals 10 times faster than horizontal axons (Girard et al., 2001). The slower conduction velocity and narrower spatial extent of horizontal connections, instead, seems inadequate to mediate fast far-SS (for reviews see: Angelucci and Bressloff, 2006; Angelucci and Shushruth, 2013). However, studies in which feedback activity was abolished by cooling or pharmacologically blocking an entire extrastriate area have provided contrasting results regarding the role of feedback in SS. Some studies observed weak reduction of SS after cooling primate area MT (Hupé et al., 1998) or areas V2 and V3 (Nassi et al., 2013), or cat postero-medial temporal visual cortex (Bardy et al., 2009). Other studies, instead, found general reduction in response amplitude, but no change in SS after pharmacologically silencing primate V2 (Hupé et al., 2001), cooling cat postero-medial temporal visual cortex (Wang et al., 2010) or optogenetically silencing mouse cingulate cortex (Zhang et al., 2014). A recent study reconciled these discrepancies; optogenetically reducing V2 feedback activity with varying intensity resulted in decreased SS, increased RF size, and decreased response amplitude in V1, but the magnitude of these effects depended on the degree of feedback inactivation (Nurminen et al., 2018).

In conclusion, our results suggest that the processing of global spatial context is initiated in L1/2A and the lower deep layers via mechanisms involving feedback connections (**Fig. 8E-F**).

The laminar pattern of neuronal activity we observed in response to far surround stimulation resembles that seen in other forms of top-down modulations of V1 activity, such as attentional modulation, working memory and figure-ground segregation (Self et al., 2013; van Kerkoerle et al., 2017), and may, thus, represent a signature of feedback processing. Similar to far-SS in our study, modulation of V1 spiking activity by attention, working memory or figure-ground is more pronounced in the superficial layers, upper L4, and deep layers compared to L4C (van Kerkoerle et al., 2017). However, while far surround stimuli in our study evoked the earliest current sinks in L1 (and upper 2) and lower deep layers (likely lower L5 and L6), attentional modulation and working memory were found to evoke stronger current sinks throughout the superficial layers and L5. These differences may be due to the different time windows analyzed in the two studies; van Kerkoerle et al. (2017) analyzed CSD signals between 200-750ms after stimulus onset, whereas we examined CSD signals occurring <200ms after stimulus onset. Instead, figure-ground modulation resulted in strongest and earliest CSD signals in L1/2A (as in our study) and L5 (unlike in our study, where far-surround stimuli evoked earliest CSD signals in lower deep layers –i.e. L5B and 6) (Self et al., 2013). Feedback activation of both L5B and 6, rather than just L5, is more consistent with the anatomy of feedback connections (Federer et al., 2015) (**Fig. 1A**). These laminar differences between studies could depend on differences in layer estimation. Self et al. (2013) overestimated the thickness of L4C and underestimated that of the superficial layers, compared to our layer estimates, which were based on anatomy and postmortem verification of laminar thickness, unlike studies in awake animals. Moreover, the laminar analysis in Self et al. (2013) was based on averaged CSD signals across many aligned

penetrations, which can lead to layer misalignments, while our quantitative analysis was performed on individual penetrations. Nevertheless, despite these subtle differences, all this evidence suggests involvement of feedback connections in different forms of top-down influences, and that subthreshold signals in L1/2A and deep layers represents a signature of feedback processing.

Effects of Anesthesia

Our study was performed under anesthesia. In mouse visual cortex isoflurane and urethane anesthesia is known to reduce inhibition and SS (Adesnik et al., 2012; Haider et al., 2013). However, no such effects have been demonstrated under sufentanil-anesthesia. For example, Shushruth et al. (2009) reported SS in 100% of recorded V1 cells, and up to 85% suppression strength, in lightly sufentanil-anesthetized macaques. Moreover, SS measurements in awake monkeys resemble those in anesthetized monkeys (Rossi et al., 2001). While anesthesia may affect absolute onset latencies of CSD signals and of SS, it is unlikely to affect the relative latency across layers and stimulus conditions, as all measurements were obtained under the same level of anesthesia, and stimuli were presented in interleaved trials. Finally, while feedback activity is reduced in extrastriate areas that are weakly active under anesthesia (e.g. area V4), this is unlikely for feedback arising from areas such as V2, V3 and MT, which are highly responsive under anesthesia (e.g. Gegenfurtner et al., 1997; Raignel et al., 1999; Shushruth et al., 2009). Moreover, a role for V2 feedback connections in SS in V1 has been demonstrated in a recent optogenetic study in sufentanil-anesthetized primates (Nurminen et al., 2018).

Conclusions

We found distinct laminar processing of local vs. global contextual signals, suggesting involvement of distinct circuits. Our study provides a template for how contextual modulation influences laminar information processing that may generalize to other sensory modalities.

Acknowledgments

We thank Kesi Sainsbury for technical assistance, Drs. Frederick Federer, Jennifer Ichida and Jeff Yarch for help with experiments. Supported by grants from NIH (R01 EY026812, R01 EY019743, BRAIN U01 NS099702), NSF (IOS 1355075, EAGER 1649923), University of Utah Research Foundation and Neuroscience Initiative, to A.A., Research to Prevent Blindness, Inc. and a core grant from NIH (EY014800) to the Department of Ophthalmology, University of Utah, a Neuroscience achievement fellowship to M.B. from the University of Utah Neuroscience Graduate Program, and postdoctoral fellowships to L.N. from the Ella and Georg Ehrnrooth Foundation., and to A.M.C. from an NIH training grant (EY024234) to the Department of Ophthalmology, University of Utah.

Author Contributions

Conceptualization, M.B., L.N., A.A. Methodology, M.B., A.A., A.M.C. Software, M.B., A.M.C. Validation, M.B., A.A. Formal Analysis, M.B. Investigation, M.B., L.N., S.M., A.A. Writing-Original Draft, M.B., A.A. Writing-Review/Editing, L.N., S.M. A.M.C. Visualization, M.B., S.M., A.A.; Supervision & Funding Acquisition, A.A.

Declaration of Interest

The authors declare no competing interests.

REFERENCES

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z.J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature* 490, 226-231.

Alitto, H.J., and Usrey, W.M. (2008). Origin and dynamics of extraclassical suppression in the lateral geniculate nucleus of the macaque monkey. *Neuron* 57, 135-146.

Allman, J., Miezin, F., and Mc Guinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Ann Rev Neurosci* 8, 407-430.

Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., and Bressloff, P.C. (2017). Circuits and mechanisms for surround modulation in visual cortex. *Ann Rev Neurosci* 40, 425-451.

Angelucci, A., and Bressloff, P.C. (2006). The contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog Brain Res* 154, 93-121.

Angelucci, A., Levitt, J.B., Walton, E., Hupé, J.M., Bullier, J., and Lund, J.S. (2002). Circuits for local and global signal integration in primary visual cortex. *J Neurosci* 22, 8633-8646.

Angelucci, A., and Sainsbury, K. (2006). Contribution of feedforward thalamic afferents and corticogeniculate feedback to the spatial summation area of macaque V1 and LGN. *J Comp Neurol* 498, 330-351.

- Angelucci, A., and Shushruth, S. (2013). Beyond the classical receptive field: surround modulation in primary visual cortex. In *The new visual neurosciences*, L.M. Chalupa, and J.S. Werner, eds. (MIT press), pp. 425-444.
- Bair, W., Cavanaugh, J.R., and Movshon, J.A. (2003). Time Course and Time–Distance Relationships for Surround Suppression in Macaque V1 Neurons. *J Neurosci* 23, 7690-7701.
- Bardy, C., Huang, J.Y., Wang, C., Fitzgibbon, T., and Dreher, B. (2009). "Top-down" influences of ipsilateral or contralateral postero-temporal visual cortices on the extra-classical receptive fields of neurons in cat's striate cortex. *Neurosci* 158, 951-968.
- Blakemore, C., and Tobin, E.A. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Exp Brain Res* 15, 439-440.
- Blasdel, G.G., and Lund, J.S. (1983). Terminations of afferent axons in macaque striate cortex. *J Neurosci* 3, 1389-1413.
- Buzsaki, G., Anastassiou, C.A., and Koch, C. (2012). The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* 13, 407-420.
- Callaway, E.M. (2014). Cell types and local circuits in the primary visual cortex of the macaque monkey. In *The New Visual Neurosciences*, L.M. Chalupa, and J.S. Werner, eds. (MIT Press), pp. 353-366.
- Callaway, E.M., and Wiser, A.K. (1996). Contributions of individual layer 2-5 spiny neurons to local circuits in macaque primary visual cortex. *Vis Neurosci* 13, 907-922.

- Casagrande, V.A., Yazar, F., Jones, K.D., and Ding, Y. (2007). The morphology of the koniocellular axon pathway in the macaque monkey. *Cereb Cortex* 17, 2334-2345.
- Federer, F., Merlin, S., and Angelucci, A. (2015). Anatomical and functional specificity of V2-to-V1 feedback circuits in the primate visual cortex. *Soc Neurosci Abstr Online*, 699.602.
- Gegenfurtner, K.R., Kiper, D.C., and Levitt, J.B. (1997). Functional properties of neurons in macaque area V3. *J Neurophysiol* 77, 1906-1923.
- Gilbert, C.D. (1977). Laminar differences in receptive field properties of cells in cat primary visual cortex. *J Physiol (Lond)* 268, 391-421.
- Girard, P., Hupé, J.M., and Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J Neurophysiol* 85, 1328-1331.
- Haider, B., Hausser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature* 493, 97-100.
- Henry, C.A., Joshi, S., Xing, D., Shapley, R.M., and Hawken, M.J. (2013). Functional characterization of the extraclassical receptive field in macaque V1: contrast, orientation, and temporal dynamics. *J Neurosci* 33, 6230-6242.
- Hubel, D.H., and Wiesel, T.N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol* 28, 229-289.
- Hubel, D.H., and Wiesel, T.N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B Biol Sci* 198, 1-59.

Hubel, D.H., and Wiesel, T.N. (2004). *Brain and Visual Perception* (New York: Oxford University Press).

Hupé, J.M., James, A.C., Girard, P., and Bullier, J. (2001). Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2. *J Neurophysiol* 85, 146-163.

Hupé, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394, 784-787.

Ichida, J.M., Schwabe, L., Bressloff, P.C., and Angelucci, A. (2007). Response facilitation from the "suppressive" receptive field surround of macaque V1 neurons. *J Neurophysiol* 98, 2168-2181.

Knierim, J.J., and Van Essen, D. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J Neurophysiol* 67, 961-980.

Logothetis, N.K., Kayser, C., and Oeltermann, A. (2007). In vivo measurement of cortical impedance spectrum in monkeys: implications for signal propagation. *Neuron* 55, 809-823.

Lund, J.S. (1973). Organization of neurons in the visual cortex, area 17, of the monkey (*Macaca mulatta*). *J Comp Neurol* 147, 455-496.

Lund, J.S., Angelucci, A., and Bressloff, P.C. (2003). Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cereb Cortex* 13, 15-24.

- Malach, R., Amir, Y., Harel, M., and Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proc Natl Acad Sci USA* 90, 10469-10473.
- Mitra, P., and Bokil, H. (2008). *Observed Brain Dynamics* (New York: Oxford University Press).
- Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol Rev* 65, 37-100.
- Mitzdorf, U., and Singer, W. (1979). Excitatory synaptic ensemble properties in the visual cortex of the macaque monkey: a current source density analysis of electrically evoked potentials. *J Comp Neurol* 187, 71-83.
- Nassi, J.J., Lomber, S.G., and Born, R.T. (2013). Corticocortical feedback contributes to surround suppression in V1 of the alert primate. *J Neurosci* 33, 8504-8517.
- Nicholson, C., and Freeman, J.A. (1975). Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. *J Neurophysiol* 38, 356-368.
- Nothdurft, H.C., Gallant, J.L., and Van Essen, D.C. (2000). Response profiles to texture border patterns in area V1. *Vis Neurosci* 17, 421-436.
- Nurminen, L., and Angelucci, A. (2014). Multiple components of surround modulation in primary visual cortex: multiple neural circuits with multiple functions? *Vision Res* 104, 47-56.

- Nurminen, L., Merlin, S., Bijanzadeh, M., Federer, F., and Angelucci, A. (2018). Top-down feedback controls spatial summation and response amplitude in primate visual cortex. *Nature Commun* 9, 2281.
- Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Harris, K.D. (2016). Kilosort: realtime spike sorting for extracellular electrophysiology with hundreds of channels. *BioRxiv*, <https://doi.org/10.1101.061481>.
- Petrov, Y., and McKee, S.P. (2006). The effect of spatial configuration on surround suppression of contrast sensitivity. *J Vis* 6, 224-238.
- Potworowski, J., Jakuczun, W., Leski, S., and Wojcik, D. (2012). Kernel current source density method. *Neural Comput* 24, 541-575.
- Raiguel, S.E., Xiao, D.K., Marcar, V.L., and Orban, G.A. (1999). Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *J Neurophysiol* 82, 1944-1956.
- Rockland, K.S., and Lund, J.S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *J Comp Neurol* 216, 303-318.
- Rockland, K.S., and Pandya, D.N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the Rhesus monkey. *Brain Res* 179, 3-20.
- Rossi, A.F., Desimone, R., and Ungerleider, L.G. (2001). Contextual modulation in primary visual cortex of macaques. *J Neurosci* 21, 1698-1709.
- Sato, T.K., Hausser, M., and Carandini, M. (2014). Distal connectivity causes summation and division across mouse visual cortex. *Nat Neurosci* 17, 30-32.

Sceniak, M.P., Chatterjee, S., and Callaway, E.M. (2006). Visual spatial summation in macaque geniculocortical afferents. *J Neurophysiol* 96, 3474-3484.

Sceniak, M.P., Hawken, M.J., and Shapley, R.M. (2001). Visual spatial characterization of macaque V1 neurons. *J Neurophysiol* 85, 1873-1887.

Schroeder, C.E., Mehta, A.D., and Givre, S.J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cereb Cortex* 8, 575-592.

Schroeder, C.E., Tenke, C.E., Givre, S.J., Arezzo, J.C., and Vaughan, H.G., Jr. (1991). Striate cortical contribution to the surface-recorded pattern-reversal VEP in the alert monkey. *Vision Res* 31, 1143-1157.

Self, M.W., van Kerkoerle, T., Super, H., and Roelfsema, P.R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Curr Biol* 23, 2121-2129.

Shushruth, S., Ichida, J.M., Levitt, J.B., and Angelucci, A. (2009). Comparison of spatial summation properties of neurons in macaque V1 and V2. *J Neurophysiol* 102, 2069-2083.

Shushruth, S., Nurminen, L., Bijanzadeh, M., Ichida, J.M., Vanni, S., and Angelucci, A. (2013). Different orientation-tuning of near and far surround suppression in macaque primary visual cortex mirrors their tuning in human perception. *J Neurosci* 33, 106-119.

Tenke, C.E., Schroeder, C.E., Arezzo, J.C., and Vaughan, H.G., Jr. (1993). Interpretation of high-resolution current source density profiles: a simulation of sublaminal contributions to the visual evoked potential. *Exp Brain Res* 94, 183-192.

van Kerkoerle, T., Self, M.W., and Roelfsema, P.R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat Commun* 8, 13804.

Wang, C., Huang, J.Y., Bardy, C., FitzGibbon, T., and Dreher, B. (2010). Influence of 'feedback' signals on spatial integration in receptive fields of cat area 17 neurons. *Brain Res* 1328, 34-48.

Webb, B.S., Dhruv, N.T., Solomon, S.G., Talib, C., and Lennie, P. (2005). Early and late mechanisms of surround suppression in striate cortex of macaque. *J Neurosci* 25, 11666-11675.

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660-665.

FIGURE LEGENDS

Figure 1. Hypothetical circuits for contextual integration in V1, and their laminar specificity. (A) V1 laminar terminations of geniculocortical (*green arrows*), intra-V1 horizontal (*red arrows*) and inter-areal feedback (*blue arrows*) projections, shown on a pia-to-white matter (*WM*) section of macaque V1 stained for the metabolic enzyme cytochrome oxidase (CO) to reveal layers. Arrow thickness represents density of projections. *White dashed contours*: laminar boundaries (layers are indicated). The terminology used to indicate layers above and below 4C is shown to the left; specifically, *Deep_U*, *Deep_L* indicate the upper and lower half of the deep layers, respectively, and *Upper-L4* encompasses upper L4C α , L4B and L4A. *M*, *P*: magnocellular and parvocellular LGN inputs, respectively. (B) Hypothetical circuits for contextual interactions in V1. Feedforward (*green*), horizontal (*red*) and feedback (*blue*) connections all contribute to the RF and the “near surround” (*orange annulus*), but only feedback contributes to the “far surround” (*gray annulus*). *White area*: RF. Figure modified from Angelucci et al. (2017).

Figure 2. Identification of laminar boundaries and RF alignment in a V1 column.

(A) **Left columns**: Minimum response field (mRF) mapping (see STAR Methods) shown as heat maps for every other contact through the depth of V1, obtained by averaging the MUA response (0-200ms after stimulus onset) to 0.5° black square stimuli flashed in a 6x6 grid centered on the hand-mapped mRF. Good alignment of mRFs (hot spots) across contacts indicates perpendicular penetration. **Right columns**: direction tuning curves obtained at each contact in response to 1° diameter grating of varying orientation and direction centered on the aggregate mRF of the column. Similarity of orientation preference across contacts indicates perpendicular penetration.

Red curves: fits to the data (*blue dots*) using the sum of two Gaussian functions. Normalized cortical depths of 0 and 1.0 indicate top and bottom of the cortex, respectively. **(B) Left**: CO staining of pia-to-white matter section of V1 showing the array track as a discoloration (*arrows*). **Right**: Same section viewed under fluorescence, showing DiI staining of the track (*arrows*). **(C) Left**: Stimulus-evoked LFP profile from same array penetration as in (A-B) obtained in response to a flashing 0.5° black square centered on the columnar mRF. *Gray shading*: L4C. **Right**: Baseline-corrected (z-scored; see STAR Methods) CSD calculated from the LFPs and displayed as a color map. *Black contour* on the CSD map: estimated onset latency of current sinks. *Solid and dashed horizontal lines*: main cortical boundaries and their subdivisions, respectively. *Blue and purple vertical bars* on the LFP and CSD profiles: stimulus onset (0 ms) and 50 ms after stimulus onset, respectively. Our definition of upper-L4 is indicated to the right of the CSD map.

Figure 3. Laminar patterns of CSD and MUA signals evoked by stimulation of the RF, near or far surround.

(A-D) Top panels: Location of the visual stimuli (0.5° black squares or 2°-width annular gratings) relative to the mRF and sRF of neurons recorded in one example penetration. *Solid and dashed red circles*: largest sRF size measured across all layers and in L4C, respectively; *black circle*: largest mRF size measured across all layers. Distances of the surround stimuli relative to the center (*ctr*) and edges of the columnar RF are indicated. **Middle panels**: Baseline-corrected and half-wave rectified (positive values are set to zero) CSD signals recorded in the example penetration across V1 layers in response to presentation of the stimulus shown in each respective top panel. The CSD profile in panel (B) is the average of 6 CSD profiles evoked by the small square stimulus presented at each of the indicated positions around the RF. We performed half-

wave rectification of the z-scored CSD in order to eliminate current sources (which mostly reflect passive return currents) and focus on the time-varying sinks that indicate changes in synaptic activation. Conventions are as in **Fig. 2C**. **Bottom panels:** Color maps of MUA activity in response to the same stimuli for the same penetration. Color scale applies to panels (A-D). **(E-H)** Population averages of z-scored and half-wave rectified CSD evoked by stimulation of the RF (E), near surround using a small square (F), or the annulus (G), and far surround (H), using similar stimuli as shown in (A-D). Near surround square stimuli encompassed distances from the mRF center between 0.75° and 1.25°, near surround annuli between 1.2° and 2° (3° in one case), and far surround annuli between 2.3° (1.42° in one case) and 5.5°.

Figure 4. Onset latency of CSD signals across layers evoked by stimulation of the RF, near or far surround: population data.

(A-D) Left panels: Box plots of absolute onset latency (time after stimulus onset) in response to stimuli in the RF (A), near surround (B-C) or far surround (D) for the population (n=10 penetrations and 48 stimulus conditions). **Middle panels:** Box plots of relative onset latency for the population. *Red vertical lines:* median values. *Black boxes:* layers where earliest responses occurred. Dashed vertical line in (D) indicates p=0.06. **(B-D) Right panels:** Distributions of normalized mRF-surround gaps for near surround square stimuli (B; n= 149 contacts), near surround annular stimuli (C; n= 88 contacts) and far surround annuli (D; n= 127 contacts). Bars of different gray scale indicate different layers. *Arrowheads:* median latency. (B) For all layers: median= 1, range= 0-4. (C) Superficial layers: median= 1.45, range=0.13-11; L4C: median= 1.58, range=1-11; Deep layers: median= 2.8, range=0.2-11. (D) Superficial layers: median= 6.66,

range=1.6-21; L4C: median= 7.8, range=4.3-21; Deep layers: median= 6.36, range= 1.6-21. See also **Table S1** and **Figs. S1-S3**.

Figure 5. RF, near and far surround stimuli engage distinct layers with distinct temporal profiles. Laminar profiles of absolute onset latency of CSD signals evoked by stimuli in the RF, near and far surround. *Dots*: absolute onset latency of current sinks in a given layer in response to a single stimulus condition. Different colors indicate different stimulus conditions (as per legend), and different dots of the same color are data from different penetrations. For penetrations for which we had multiple data from the same stimulus conditions (e.g. annuli of different inner diameter in the far surround), we plot onset latencies for each condition (therefore, the numbers of data point for RF, near and far surround do not necessarily match). *Solid lines*: medians. The statistical analyses for this data are reported in **Tables S2-S5**.

Figure 6. Laminar patterns of MUA responses evoked by center-only and center-surround gratings: example case.

(A-C) Top insets: Center-surround grating stimuli used to probe near-SS and far-SS across the recorded V1 column. *Red circle* outlines the size of the aggregate sRF of neurons in a column to which the center grating was matched. **Bottom rows:** PSTHs in response to the center-only stimulus (*red*) and the respective center-surround stimuli shown at the top of each column, for one example penetration (same penetration as in **Fig. 3A-D**). *Black curves*: responses to iso-oriented center-surround gratings; *cyan curves*: responses to orthogonally-oriented center-surround gratings. The response at each contact was normalized to the peak of the center-only response in the first 300ms. *Shade*: ± 1 sem. *Vertical lines* mark onset latencies (*red*: center-only;

black: iso-oriented suppression; *orange*: tuned suppression). Suppression Index (SI; see STAR Methods) values indicate the strength of suppression (estimated from the average response across trials) caused by iso-oriented and orthogonally-oriented surround stimuli (ranging from 0 –no suppression – to 1 –complete suppression –). OSI is the orientation-selectivity index of the suppression (see STAR Methods) measured from the average response and ranging from 0 (unselective) to 1 (highly selective). The center-only responses differ in (A) (B) and (C) because they are responses to the center-only stimulus presented closest in time to each respective center-surround stimulus.

Figure 7. Onset latency of spiking responses to RF stimulation and of surround suppression across layers: population data.

(A) Layer-by-layer distribution of absolute (left panel) and relative (right panel) onset latency of spiking responses to gratings in the RF, across the population of penetrations (n=6) and contacts (n=79). **(B-E)** Layer-by-layer distribution of absolute (left panels) and relative (right panels) onset latency of SS evoked by a 20° diameter grating patch encompassing the RF center and the full extent of the surround (B), iso-oriented center + near surround gratings (C), orthogonally-oriented center + near surround gratings (orientation-tuned SS; D), or iso-oriented center + far surround gratings (E). *Rightmost insets*: visual stimuli used. Other conventions are as in **Fig. 4**. Dashed vertical lines in (A,C) indicate $p=0.059$, 0.053 and 0.053 , respectively. See also **Fig. S4** and **Tables S6-S10**.

Figure 8. Laminar processing and circuits underlying RF responses, and their suppression by local and global context.

Schematics of our results on the circuits and V1 layers that initiate responses to stimuli inside the RF **(A)**, and their suppression by local **(C-D)** and global **(E-F)** contextual stimuli in the RF surround. The stack of layers represents the center V1 column from which we performed laminar recordings. Panels (A,C,E) indicate the circuits carrying visual signals to the center column following stimulation of the RF (A), near-surround only (C) or far-surround only (E), and the layers where these circuits evoke the earliest subthreshold responses (*orange shading*). *Darker orange shading* indicates layers activated significantly faster by near than far surround stimuli. *Red cells*: laminar location of earliest spiking responses to the stimulus. *M, P, K*: magno-, parvo- and konio-cellular LGN afferents. Panels (B,D,F) indicate the layers where SS emerges earliest (*black shading*) in response to large visual stimuli encompassing the RF and either the full surround (B), or the near (D) or far surround (F). *Orange cells*: laminar location of earliest surround suppressed cells as a result of the circuits depicted in each respective panel above. **(A)** Stimuli inside the RF evoke the earliest CSD and spiking signals in L4C, which are relayed to the V1 center column by faster M and P (relative to K) geniculate pathways (*green arrows*). **(B)** A large stimulus encompassing the RF and full surround causes SS of LGN afferents to the center V1 column, resulting in withdrawal of afferent excitation to L4C, where suppression first emerges in V1. **(C)** Visual stimuli in the near surround (beyond the extent of M and P geniculocortical afferents), in the absence of RF stimulation, activate neurons outside the center hypercolumn in V1 (*red shading*), in extrastriate cortical areas (*blue shading*), and possibly LGN K-layers (*green shading*); these surround signals are conveyed to the superficial and deep layers of the center V1 column by multiple long-range connections, including V1 horizontal connections (*red arrows*), feedback connections (*blue arrows*) and K1-2 afferents (*green arrow*). **(D)** SS in response to center and near-surround stimuli first emerges in superficial layers,

particularly upper-L4. **(E)** Visual stimuli in the far surround (beyond the extent of monosynaptic V1 horizontal connections) activate neurons in both V1 and extrastriate areas located several millimeters away from the center V1 column. However, faster conducting and spatially more widespread feedback connections convey these signals to L1/2A and Deep_L of the center V1 column much faster than multisynaptic chains of V1 horizontal connections. **(F)** Feedback inputs to L1 and Deep_L can cause SS in all layers except L4C, which lacks direct feedback connections and whose cells confine their dendrites to L4C.

STAR METHODS

Experimental Model and Subject Details

Macaque monkeys (*Macaca fascicularis*) were purchased from a commercial breeder, quarantined for 6 weeks and group-housed at the University of Utah prior to being used for experimentation. Linear array recordings (total of 22 penetrations from 7 animals) were made in the parafoveal representation of the primary visual cortex (V1) in adult macaque monkeys (3-4 kg). We selected for analysis only those penetrations that were deemed to be perpendicular to the surface of V1 according to the criteria described below (n=10 penetrations, total 162 contacts, from 4 macaques, 2 males and 2 females). All experimental procedures were in accordance with protocols approved by the University of Utah Institutional Animal Care and Use Committee and with NIH guidelines.

Methods Details

Surgical Procedures

Animals were pre-anesthetized with ketamine (25 mg/kg, i.m.), intubated, artificially ventilated with a 70:30 mixture of N₂O and O₂, and their head was fixed by positioning in a stereotaxic apparatus. During surgery, anesthesia was maintained with isoflurane (2%), and end-tidal CO₂, blood O₂ saturation, electrocardiogram, blood pressure, lung pressure, and body temperature were monitored continuously. A small craniotomy and durotomy were performed over the opercular regions of V1 and a PVC chamber glued to the skull surrounding the craniotomy. On

completion of surgery, isoflurane was turned off, anesthesia maintained with sufentanil citrate (6-12 $\mu\text{g/kg/h}$, i.v.), and paralysis was induced by continuous i.v. infusion of vecuronium bromide (0.3 $\mu\text{g/kg/h}$) to prevent eye movements. The pupils were dilated with topical atropine, and the corneas were protected with gas-permeable contact lenses. The eyes were refracted using corrective lenses, and the foveae were plotted on a tangent screen using a reverse ophthalmoscope, and periodically remapped throughout the experiment.

Electrophysiological Recordings

Extracellular recordings (MUA and LFP) were made in parafoveal V1 (4-8° eccentricities) using 24-channel linear electrode arrays (100 μm inter-contact spacing, 20 μm contact diameter; V-Probe, Plexon, Texas). One penetration was performed using a 32-channel linear probe (100 μm spacing; NeuroNexus A32, Michigan). A custom-made guide tube provided mechanical stability for the V-probe recordings. At the beginning of each recording session, the array was positioned normal to the cortical surface under visual guidance using triangulation, and the recording site was stabilized by half filling the chamber with agar; the agar was then covered with silicon oil or saline to prevent it from drying out. The probe was then slowly advanced through the cortical thickness to a depth of 2.0-2.2mm, over a 60-90 minute period, or until LFP signals and spikes could be recorded from the bottom contact through the top third or fourth contact (from the pial surface). At the end of each recording, new craniotomies and durotomies were performed and the recordings targeted to a new cortical site. To facilitate post-mortem visualization of the lesion tracks, the probes were coated with DiI (Molecular Probes, Eugene, OR) prior to insertion.

Data was collected (30kHz sampling rate) and amplified using a 128-channel system (Cerebus, 16-bit A-D, Blackrock Microsystems, Salt Lake City, UT). To obtain LFPs, the raw

voltage recordings were band-pass filtered (1-100Hz, 2nd-order Butterworth filter) and down sampled to 2kHz. MUA was obtained by band-pass filtering (250 Hz-7.5 kHz) the raw signal continuously recorded at a sampling rate of 30 kHz. MUA was thresholded based on signal energy, using the built-in Cerebus program.

Receptive Field Mapping and Array Verticality

After manually locating the receptive fields (RFs) of neurons across the column, their aggregate minimum response field (mRF) was mapped quantitatively using 0.5° black square stimuli flashed systematically over a 3x3° visual field area (500ms, 5-15 trials, interleaved with 500ms blanks of mean luminance gray). The aggregate spatial mRF of the column was defined as the visual field region in which the square stimulus evoked a mean response (-2 SD of the stimulus evoked response) that was > 2 SD above mean spontaneous activity, and the geometric center of this region was taken as the RF center. All subsequent stimuli were centered on this field. We then determined orientation, eye, spatial and temporal frequency preferences of cells across contacts using 1-1.5° diameter drifting sinusoidal grating patches of 100% contrast presented monocularly. Subsequent stimuli were presented at the optimal parameters for most contacts across the column (unless otherwise specified), and in cases when different contacts showed significantly different stimulus preferences, the experiments were run multiple times using each preferred stimulus. We also measured size tuning across the column using 100% contrast drifting grating patches of increasing size (from 0.1-26°) centered over the aggregate mRF of the column. From these tuning curves we extracted the summation RF (sRF) diameter as the grating diameter at peak response. The latter was later used to create center and annular

surround stimuli used to probe SS. To monitor eye movements, the RFs were remapped by hand approximately every 10-20 minutes and stimuli re-centered on the RF if necessary.

For the analysis of sRF sizes shown in **Figs. S1A, S2-S3** size tuning curves were generated by plotting for each stimulus size the mean- 2SD of the evoked response (in the first 200 ms) and subtracting from this the mean+2SD of the spontaneous activity; the peak of this curve was chosen as the sRF size.

To ensure that the array was positioned orthogonal to the cortical surface, we used as criteria the vertical alignment of the mapped mRF at each contact, and the similarity in the orientation tuning curves across contacts (e.g. **Fig. 2A**). If RFs were misaligned across contacts, the array was retracted and repositioned. Moreover, during offline analysis, we excluded from further analysis all penetrations that were deemed to be non-vertical.

Visual Stimuli

Visual stimuli were generated using Matlab (Mathworks Inc., Natick, MA; RRID:SCR_001622) and presented on a calibrated CRT monitor (Sony, GDM-C520K, 600x800 pixels, 100Hz frame rate, mean luminance 45.7cd/m², at 57cm viewing distance), and their timing was controlled using the ViSaGe system (Cambridge Research Systems, Cambridge, UK; RRID:SCR_000749). All stimuli were displayed for 500ms, followed by 500 or 750ms interstimulus interval.

To characterize the onset latency of CSD signals evoked by stimuli in the RF and surround, we used two kinds of stimuli: 1. a black square of 0.5° side systematically flashed over a 3x3° visual field areas centered on the columnar RF; and 2. static and dynamic annular gratings of optimal parameters for the neurons in the column, 2° in width, presented at distances from the

RF center ranging from 0.2° to 6° (measured to the inner edge of the annulus), and presented without a stimulus in the RF.

To characterize the temporal emergence of tuned and untuned SS of spiking responses, we used two kinds of grating stimuli: 1. a 20° diameter drifting grating patch centered on the columnar RF, and presented at the optimal parameters for neurons in the column; and 2. a 2°-width annular static grating of optimal or of orthogonal-to-optimal orientation, flashed at increasing distances from the RF center (from 0.25° outside the edge of the columnar sRF to 6° from the RF center), and presented together with a static grating patch of optimal parameters centered on the columnar RF and matched to the columnar sRF size. The latter stimulus, thus, differed from the former stimulus in that there was always a gap (of the same mean luminance as the grating) between the center and surround gratings. Stimulus presentation was randomized and all grating stimuli were interleaved with presentation of center-only grating patches to activate the RF in isolation.

Quantification and Statistical Analysis

Current Source Density Analysis

Current source density (CSD) was applied to the band-pass filtered (1-100Hz) and trial averaged LFP using the kernel CSD toolbox (kCSD_Matlab, RRID:SCR_016424) (Potworowski et al., 2012). CSD was calculated as the second spatial derivative of the LFP signal, which reflects the net local transmembrane currents that generate the LFP.

Specifically, CSD was computed as:

$$\text{CSD}(x) = -\sigma * \frac{v(x-h) - 2v(x) + v(x+h)}{h^2} \quad \text{eq.1}$$

where, v is the voltage (μV), x the point in the extracellular medium at which CSD is calculated, h is the spacing between recording contacts on the linear probe (here $100\mu m$), and σ is the conductivity of the cortical tissue ($0.4 S/m$)(Potworowski et al., 2012). To estimate CSD across layers, we interpolated the CSD every $10\mu m$. The CSD was baseline corrected (Z-scored). In particular, we normalized the CSD of each profile to the standard deviation (SD) of the baseline (defined as 200ms prior to stimulus onset) after subtraction of the baseline mean, as in equation 2:

$$Z\text{-CSD}(t) = \frac{CSD_{st}(t) - \text{mean}(CSD_b)}{\text{std}(CSD_b)} \quad \text{eq.2}$$

where, $CSD_{st}(t)$ is the computed CSD at each time point (every 0.5ms) after stimulus onset and CSD_b is the computed CSD from 200ms prior to stimulus onset to stimulus onset.

CSD provides information about the current flow in the extracellular medium, and is better suited than LFP to localize input activity (Logothetis et al., 2007). Current sinks (negative voltage deflections, visualized in blue in our CSD maps) in the extracellular medium are thought to reflect integrated subthreshold inputs at postsynaptic dendrites (Mitzdorf, 1985; Mitzdorf and Singer, 1979; Nicholson and Freeman, 1975). We used CSD responses to small stimuli flashed inside the RFs to identify laminar borders (as detailed in the Results). We also used CSD analysis to localize surround-evoked input activity to specific cortical layers. In particular, since surround stimuli presented alone in the absence of direct RF stimulation do not cause significant spiking activity from the recorded cells, the CSD sinks evoked by these stimuli reflect the laminar location of the subthreshold inputs. Measuring the onset latency of these CSD sinks, thus provides us with information about which layers are first activated by surround stimuli.

Alignment of Penetrations

To generate average CSDs across layers, we aligned the different penetrations using their individual CSDs (as well as the other criteria described in the Results) to identify layers. All our penetrations spanned cortical depths of 1.5-1.7mm, and the bottom of L4C across penetrations was consistently located at depths of 0.95-1.1mm from the top. This allowed us to align penetrations at the location of the lower border of L4C.

To obtain grand average CSDs (e.g. **Fig. 3E-H**), we half-wave-rectified the individual Z-scored CSD profiles (discarding the positive source values), normalized them between -1 and 0, aligned the individual CSD profiles, and finally averaged the CSD values across aligned penetrations.

Latency Analysis of CSD

The onset latency of current sinks in the CSD for the quantitative analysis shown in **Fig. 4** was measured at each interpolated depth as the earliest time bin after stimulus onset in which the CSD amplitude was 3-5SD below baseline for three consecutive bins. The time of the first bin was taken as the signal onset. For the grand averages shown in **Fig. 3E-H**, onset latency was computed using a SD of 7, as the average rectified CSD showed lower variation of baseline activity compared to the non-rectified signal.

Suppression and Orientation Specificity Indices

The strength of suppression induced by the surround stimuli was measured as a suppression index (SI), which was computed as:

$$SI(\theta) = 1 - \frac{R_{CS}}{R_C} \quad \text{eq. 3}$$

where θ is the surround orientation for the condition, R_c is the mean MUA response to the center-only stimulus (in the first 150ms after stimulus onset), and R_{CS} is the mean MUA response to the center+surround stimulus (in the first 150ms). A SI=1 indicates that the surround stimulus completely abolished the response to the center-only stimulus, SI=0 indicates no SS, and SI<1 indicates that the surround stimulus increased the response to the center stimulus alone.

The orientation specificity of SS was measured as an orientation selectivity index (OSI), which was computed as:

$$OSI = \frac{R_{cross} - R_{iso}}{R_{cross} + R_{iso}} \quad \text{eq.4}$$

where, R_{cross} and R_{iso} are the response to an iso-oriented and cross-oriented center-surround stimulus, respectively. An OSI=0 indicates no selectivity, while an OSI=1 indicates maximal selectivity.

Latency Analysis of Surround Suppression

Peristimulus time histograms (PSTHs) were obtained by convolving a Gaussian filter (10ms bandwidth) with the MUA raster plots, using the chronux toolbox (<http://chronux.org>; RRID:SCR_005547) (Mitra and Bokil, 2008), providing spikes with 2ms resolution. Then, the mean baseline response was subtracted from the stimulus-evoked response. The onset latency of SS caused by iso-oriented or cross-oriented surround stimuli was estimated to be the time point at which the MUA PSTH in response to the center-surround stimulus diverged significantly from the MUA PSTH in response to the center-only stimulus. This was estimated as follows. First the average MUA response to the center-surround stimulus was subtracted from the average MUA response to the center-only stimulus; second the local minima and maxima of the absolute value of this response difference were found (minima=0 indicates that the two curves intersected);

third, the algorithm searched forward in time (up to 300ms after stimulus onset) starting from the first local extrema occurring after the onset of the center-only response, until it found a bin (20ms width) at which the SI reached 0.15 and was followed by a bin with a larger area under the curve (i.e. for which the difference between the two curves was larger). If these criteria held for 3 consecutive bins, then the first bin was chosen as the time of suppression onset. Essentially, based on these constraints, the time of suppression could only be equal to, or larger than, the response onset to the center-only stimulus. The response onset to the center-only stimulus was taken to be the time point at which the PSTH reached 10% of the maximum value (Bair et al., 2003). For the boxplots of **Fig. 7B,C,E** this analysis was performed only on MUA responses that showed an $SI \geq 0.15$ (measured over 150ms after stimulus onset, because after this time window the response was typically back to baseline).

The onset latency of tuned SS was measured, using the same approach as described above, as the time point at which the PSTHs in response to iso-oriented and cross-oriented center-surround stimuli diverged significantly (i.e. the absolute OSI value was >0.1 for 3 consecutive bins). An additional constraint, here, was that the onset of tuned suppression could not precede in time the onset of untuned suppression (i.e. the onset of suppression caused by either an iso-oriented or a cross-oriented center-surround stimulus, whichever was faster). For the boxplots of **Fig. 7D** this analysis was performed only on MUA responses that showed an $SI \geq 0.15$ for either iso- or cross-oriented center-surround stimuli (over the 150ms after stimulus onset).

We also compared our method to one that computes the latency by taking into account the trial variability, using an approach similar to that used in Henry et al (Henry et al., 2013); this method compares the cumulative spike counts during stimulus presentation in the center-only vs.

the center-surround stimuli. The cumulative spike counts were generated by bootstrap resampling (with replacement, 5000 iteration) from the population of trials at each time bin (10ms width). The time bin was chosen as the time of onset latency, if the spike count in response to the center-only stimulus was larger than the spike count in response to the center-surround stimulus for at least 95% of the time. We found this method was only effective to measure latency under conditions that evoked strong SS or in which SS was sharply tuned [as also discussed in Henry et al. (2013)]. However, with this method the latency estimate increases as the change in response is scaled down, i.e. for weaker SS or more weakly tuned suppression, as is the case for far-SS. We do not report the results of this analysis, as we found it to be ineffective to measure onset latency for far-SS.

Spike Sorting

To verify that the results of the MUA analysis were consistent with single unit responses in V1, we first identified single units using the automated spike sorting software Kilosort (RRID: SCR_016422) (Pachitariu et al., 2016). Importantly, in addition to the temporal features of distinct spike waveforms from individual units, Kilosort also takes into account the spatial distribution of these events across multiple channels on the array. This ensures the same spike event recorded across multiple channels is not erroneously classified as belonging to separate units on different channels. The output of the Kilosort algorithm was manually checked, by comparing identified units' spike waveforms, principal component features and cross-correlograms, to determine whether units needed to be merged or further split to ensure good isolation. Finally, putative single-unit isolations were assessed on the basis of waveform

amplitude, drift, and ISI distribution before final classification. Once extracted, single units that showed visual responses were selected for the latency analysis. The peristimulus time histogram, latency of SS and tuned SS were all computed in the same way as for the MUA data described above.

Histology and Track Reconstruction

On completion of the recording session, the animal was perfused transcardially with 4% paraformaldehyde in 0.1M phosphate buffer. The occipital pole was frozen-sectioned at 40 μ m sagittally. DiI-labeled tracks were visualized under fluorescence to ascertain verticality of the array and verify cortical layer assignment (e.g. **Fig. 2B**). Adjacent tissue sections were counterstained for cytochrome oxidase for identifying cortical layers as well as the location of the electrode track (visible as a discoloration in staining).

TABLE FOR AUTHOR TO COMPLETE

Please upload the completed table as a separate document. **Please do not add subheadings to the Key Resources Table.** If you wish to make an entry that does not fall into one of the subheadings below, please contact your handling editor. (NOTE: For authors publishing in Current Biology, please note that references within the KRT should be in numbered style, rather than Harvard.)

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Bacterial and Virus Strains		
Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
Critical Commercial Assays		
Deposited Data		
Experimental Models: Cell Lines		

Experimental Models: Organisms/Strains		
Oligonucleotides		
Recombinant DNA		
Software and Algorithms		
kCSD	Potworowski et al., 2012	RRID:SCR_016424
Kilosort	Pachitariu et al., 2016	RRID:SCR_016422
Custom Matlab scripts	This Paper	RRID:SCR_001622
Blackrock Matlab toolkits	Blackrock Microsystems	http://blackrockmicro.com/neuroscience-research-products/ephys-analysis-software/matlab-development-kits/
ViSaGe toolbox (Matlab scripts)	Cambridge research system	RRID:SCR_000749
Chronux	Mitra and Bokil, 2008	RRID:SCR_005547
Other		
A32 laminar microelectrode probe	NeuroNexus	http://neuronexus.com/
V-Probe laminar microelectrode array	Plexon	https://plexon.com/

Figure 1

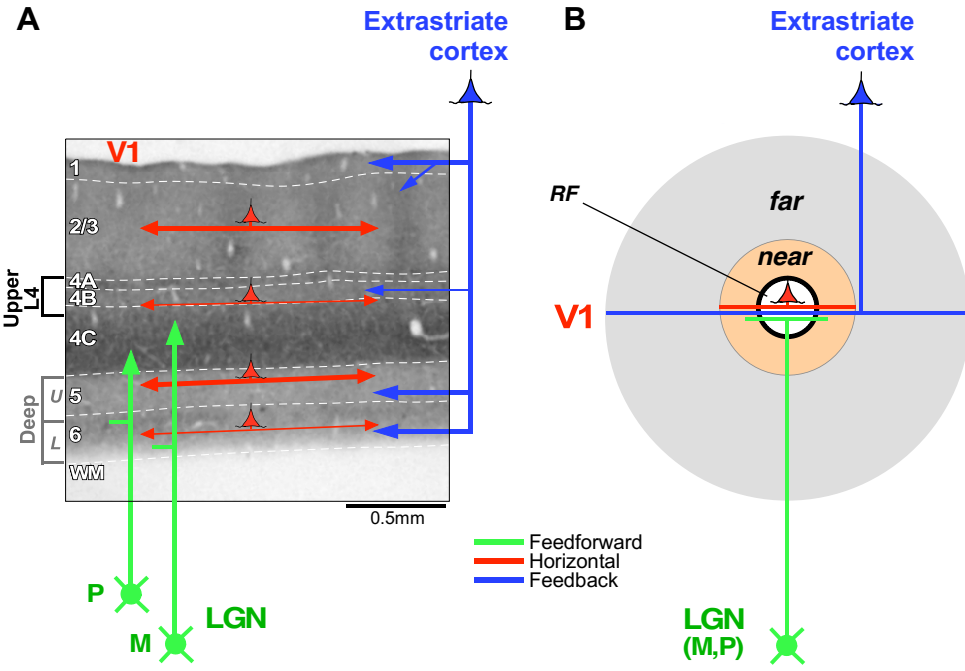


Figure 2

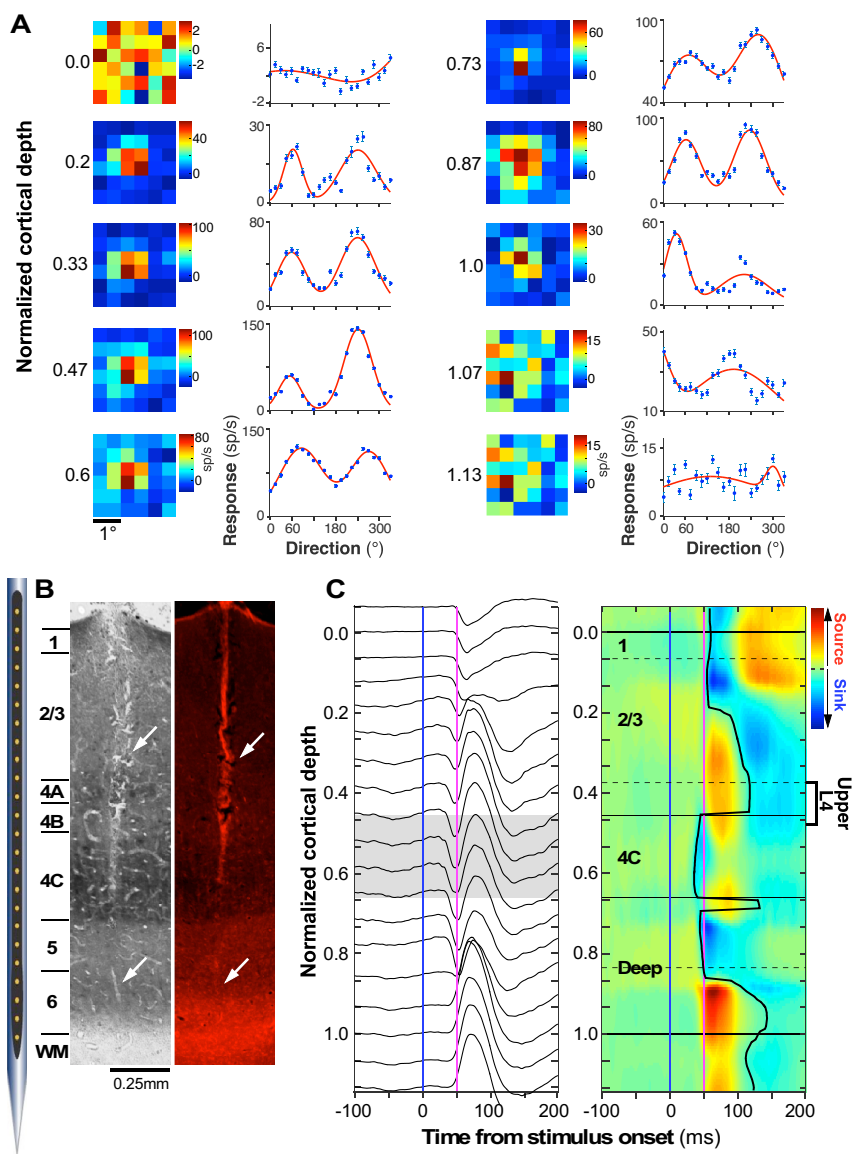


Figure 3

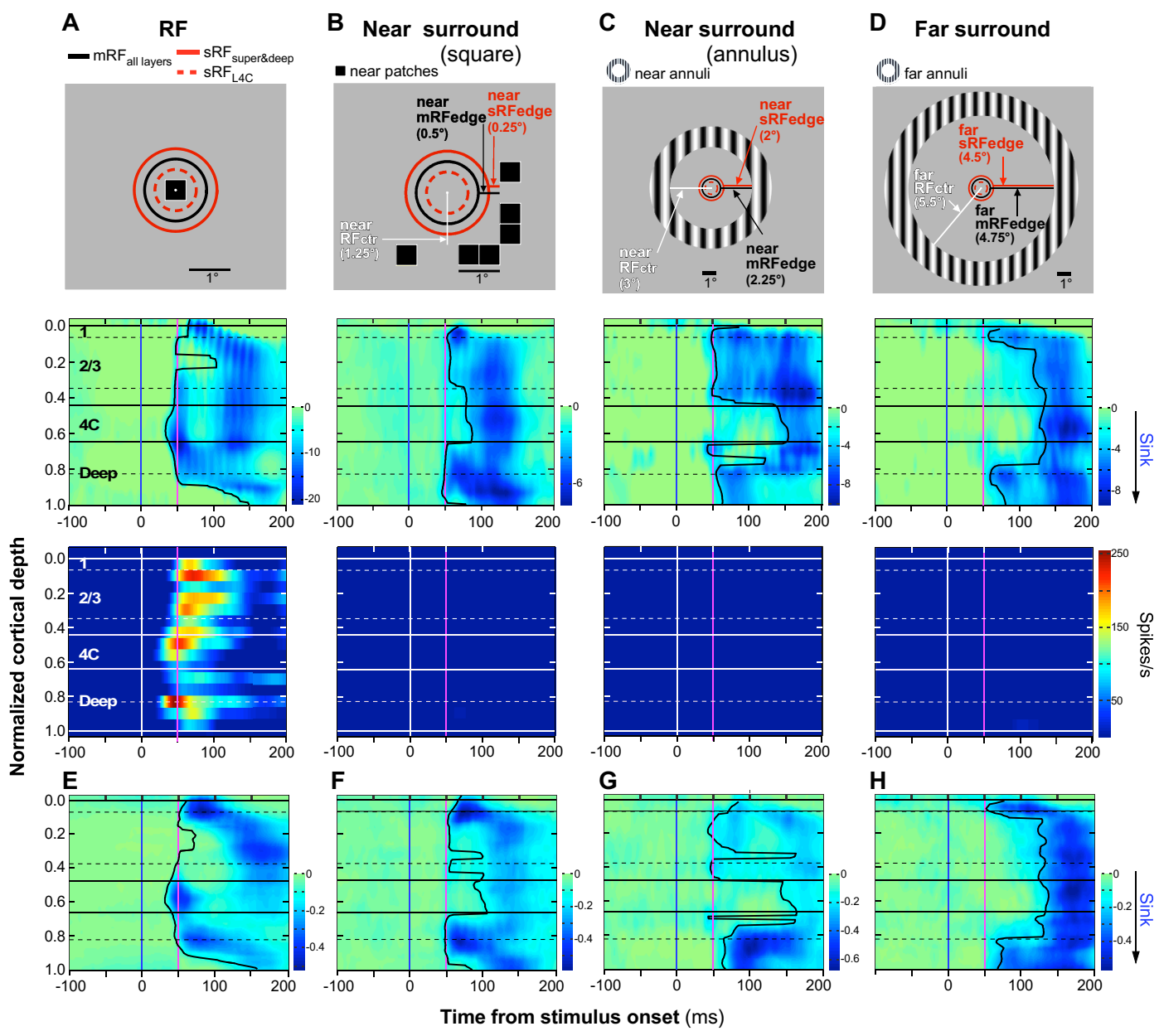


Figure 4

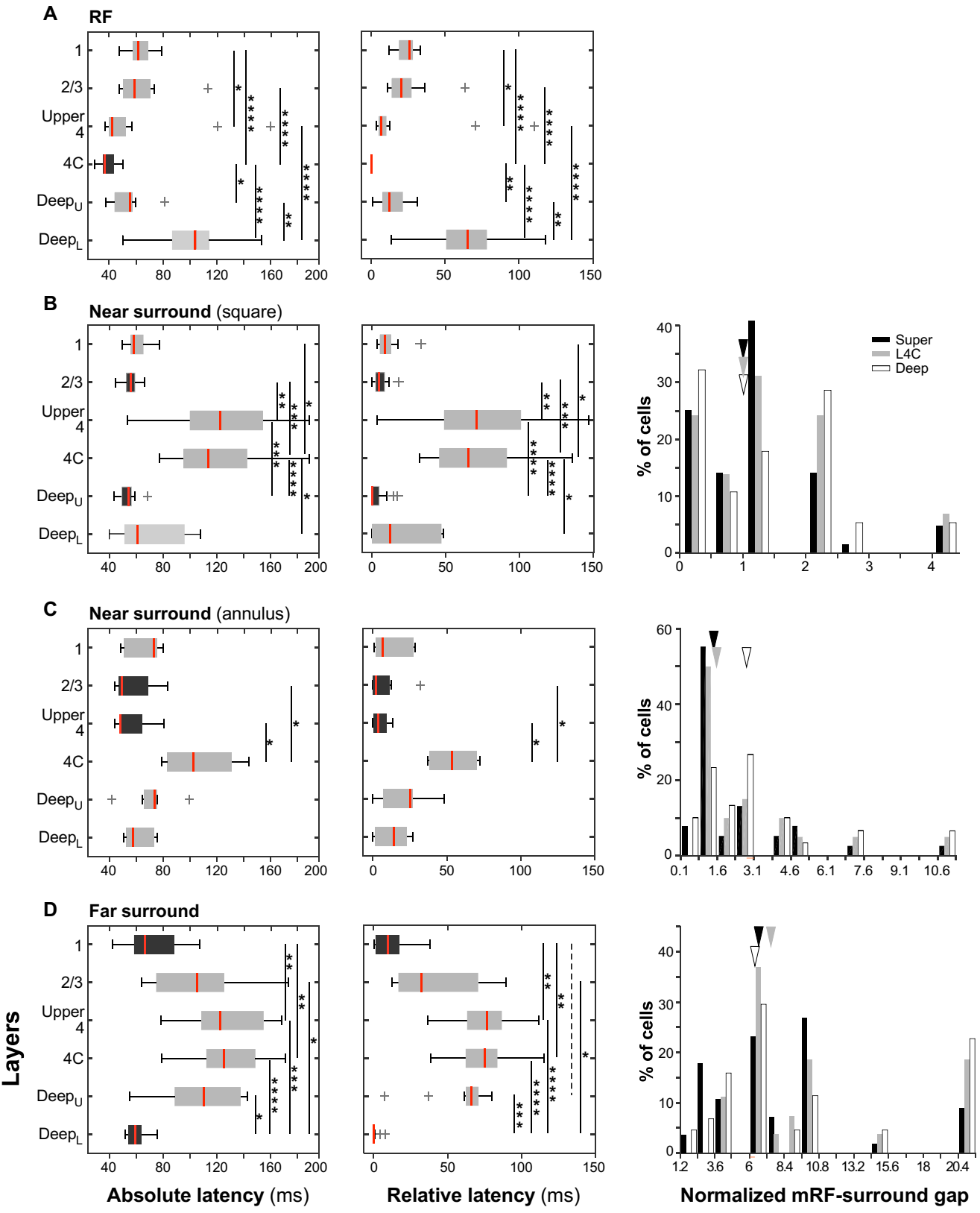


Figure 5

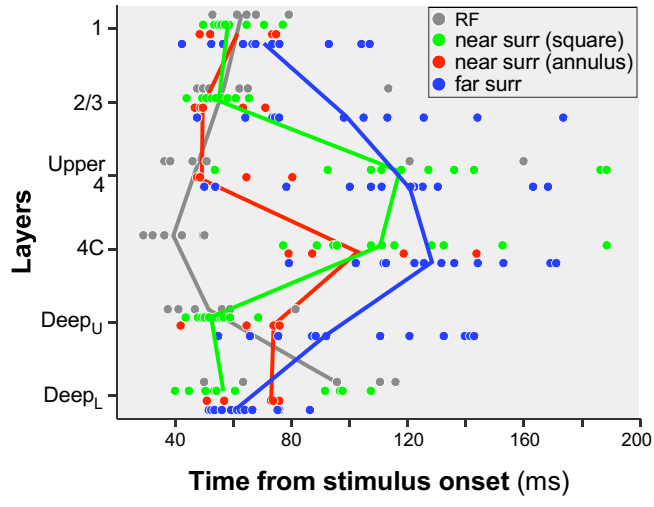


Figure 6

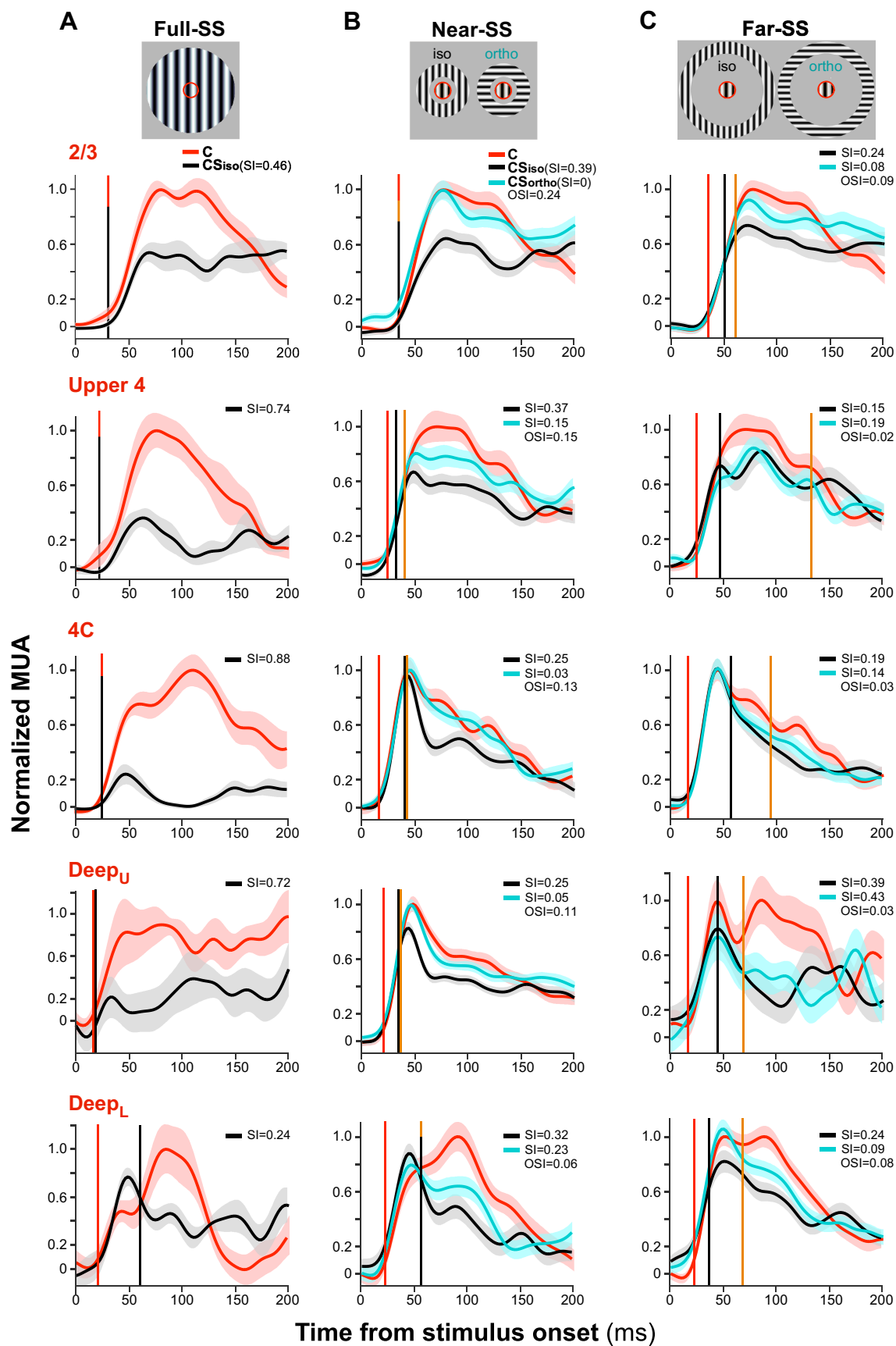


Figure 7

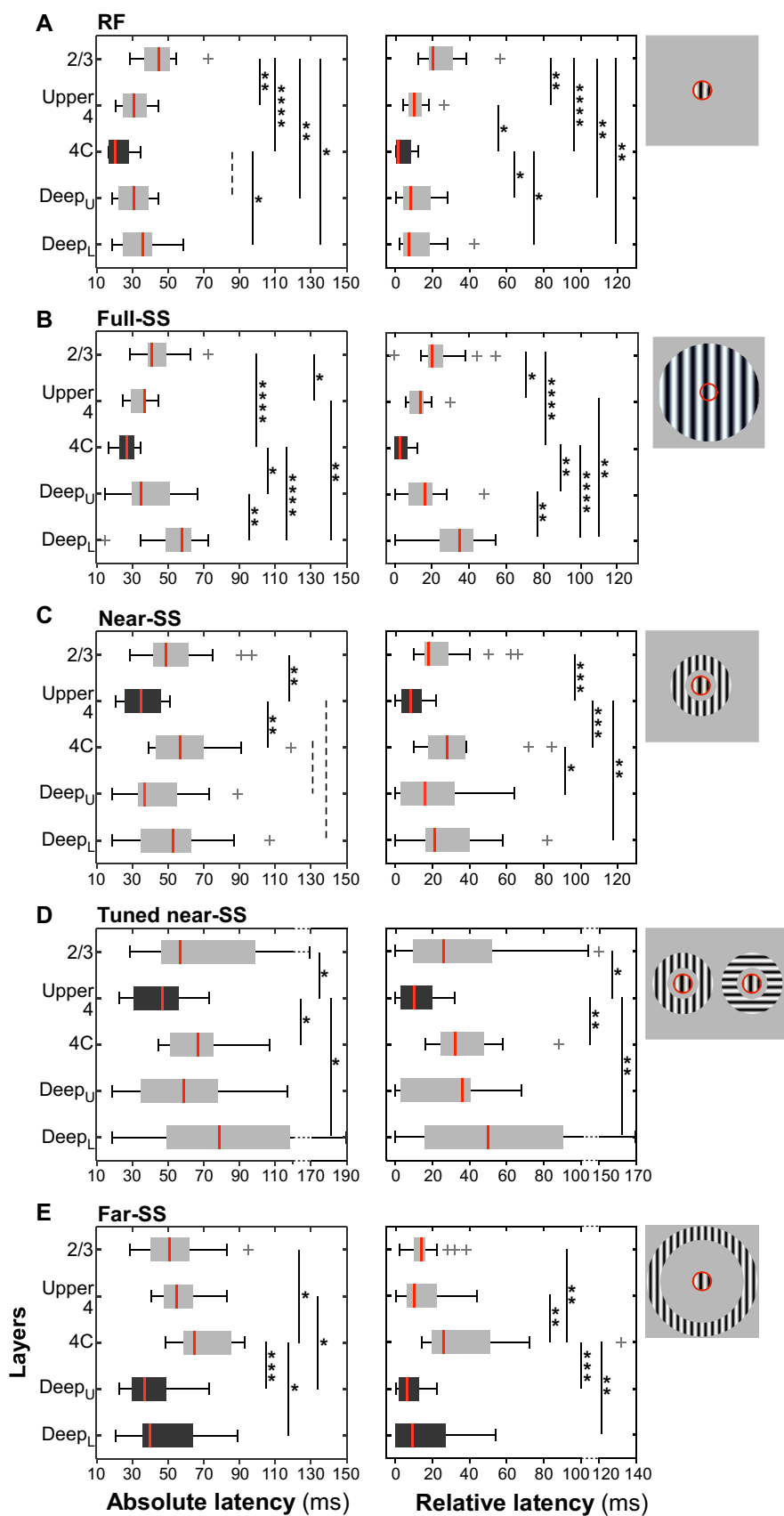
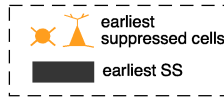


Figure 8



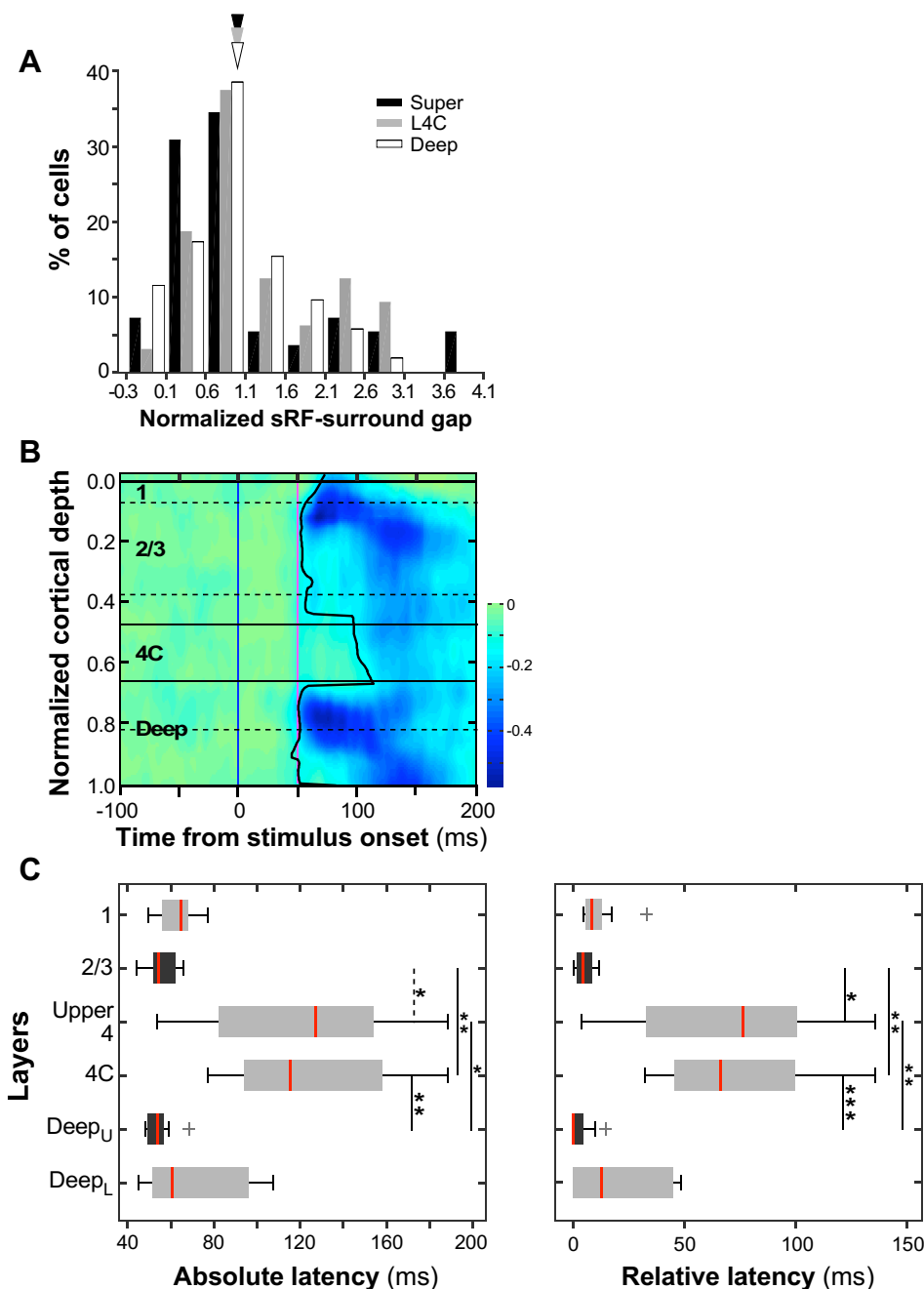


Figure S1 (Related to Fig. 3F and 4B)

(A) Distribution of normalized sRF-surround gaps for square stimuli in the near surround ($n=139$ contacts). *Arrowheads*: medians=1 for all layers. Superficial layers: range=-0.22-4; L4C: range=0-3; Deep layers: range= -0.33-3. (B-C) Grand average CSD analysis (B) and onset latency of CSD signals (C) performed as in **Figures 3F,4B**, respectively, but excluding conditions ($n=9$) with negative values of normalized sRF-surround gap. Conventions are as in **Figs. 3,4**. (C) Left panel: L2/3 vs upper-L4, $p=0.058$; L2/3 vs L4C, $p=0.0035$; upper-L4 vs. Deep_U: $p=0.023$; L4C vs. Deep_U: $p=0.0011$. Right panel: L2/3 vs upper-L4, $p=0.027$; L2/3 vs L4C, $p=0.003$; upper-L4 vs. Deep_U: $p=0.0029$; L4C vs. Deep_U: $p=0.0002$.

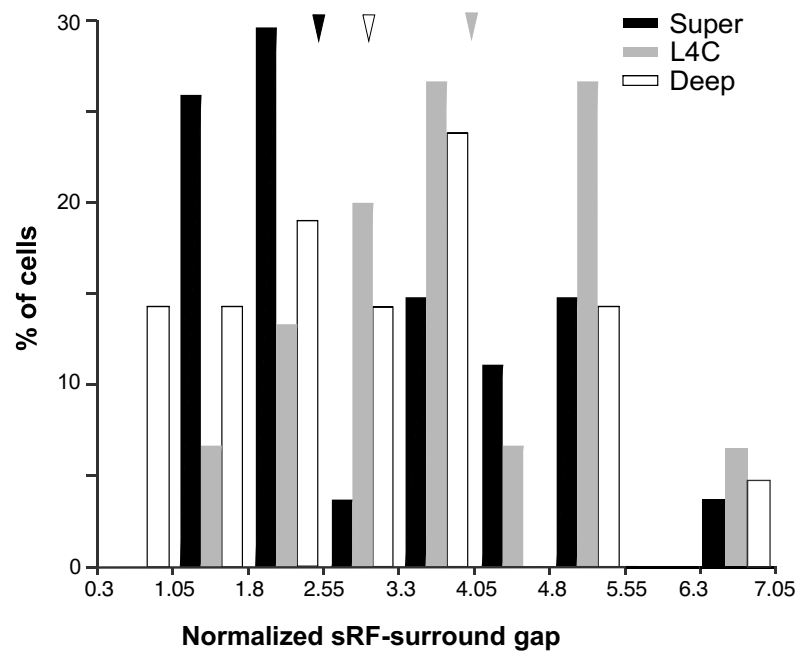


Figure S2 (Related to Fig. 4C)

Distribution of normalized sRF-surround gaps for annular stimuli in the near surround (n=66 contacts). *Arrowheads*: medians. Superficial layers: median=2.5, range=1.22-6.5; L4C: median= 3.92, range=1.33-6.5; Deep layers: median=3, range= 0.4-6.5.

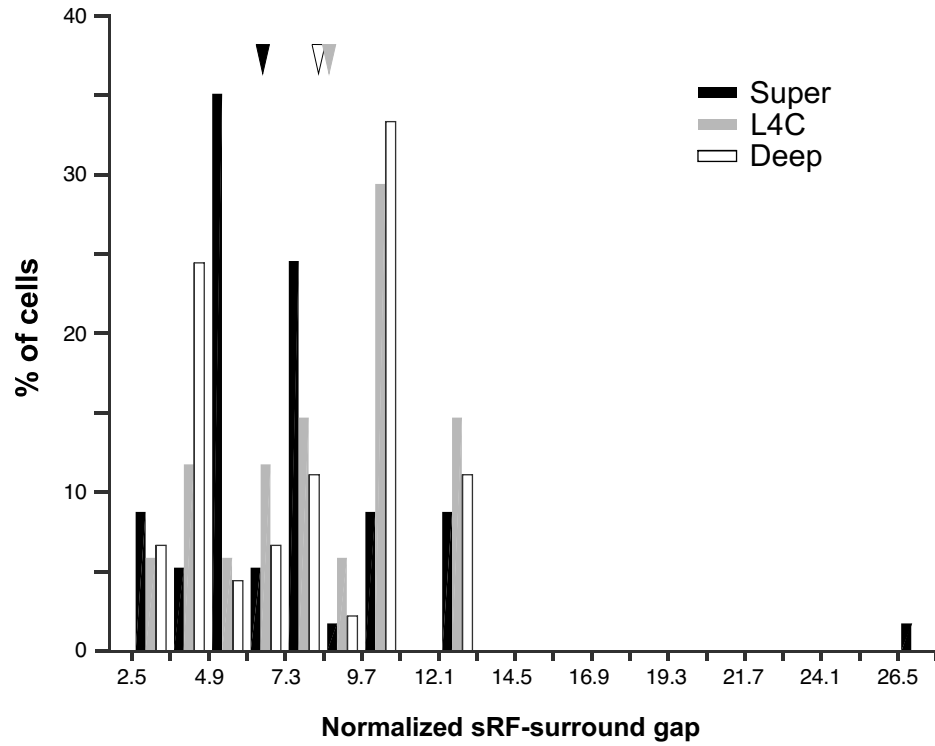


Figure S3 (Related to Fig. 4D)

Distribution of normalized sRF-surround gaps for stimuli in the far surround (n=136 contacts). *Arrowheads*: medians. Superficial layers: median=6.66, range=3.41-27; L4C: median= 8.66, range=3.6-12.75; Deep layers: median=8.33, range= 2.53-12.75.

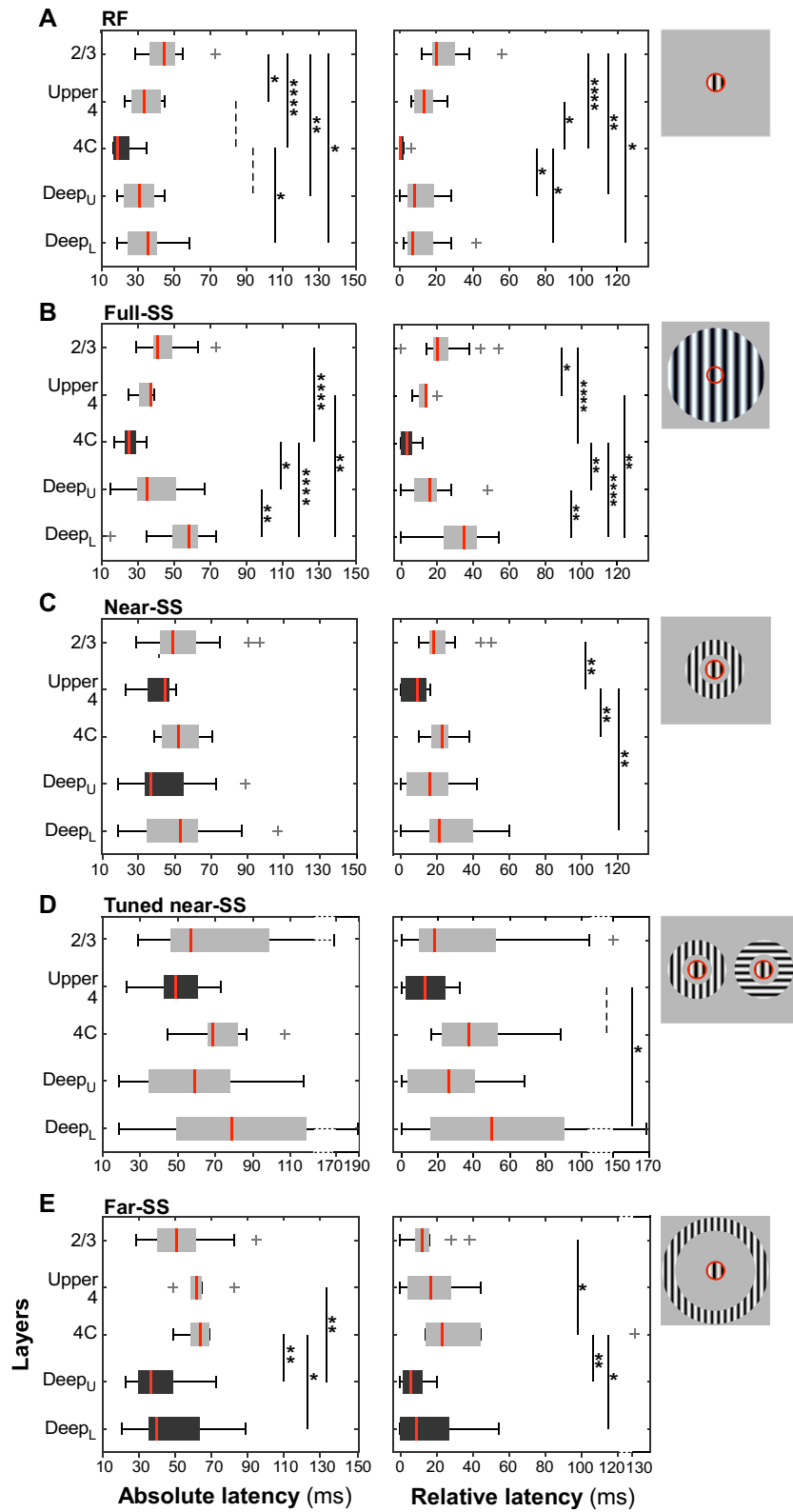


Figure S4 (Related to Fig. 7)

Layer-by-layer distributions of absolute and relative onset latency in response to the stimuli indicated to the right of each row. Same data and analysis as in **Fig. 7**, but excluding contacts at laminar borders. Conventions are as in **Fig. 7**.

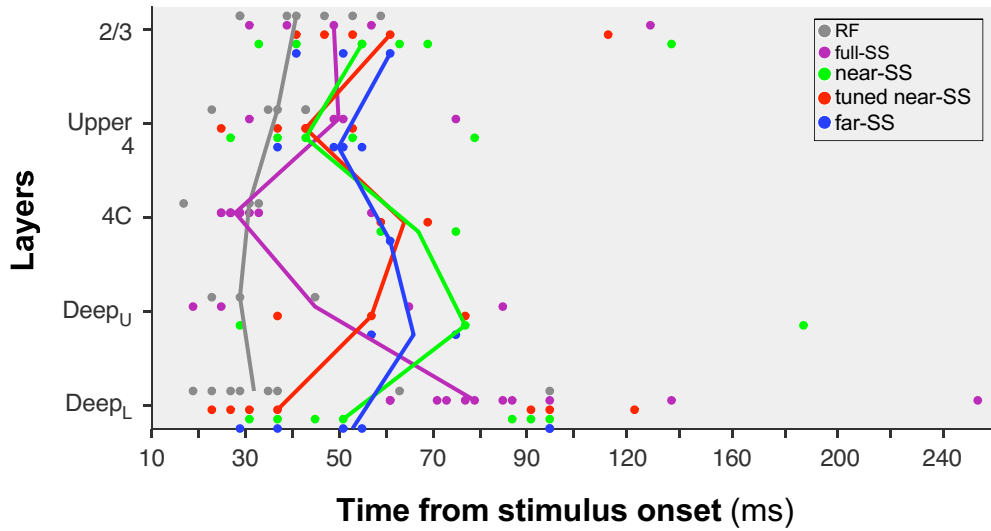


Figure S5 (Related to Fig. 7)

Singel Unit (SU) data. Laminar profile of absolute onset latency of spiking responses to stimuli in the receptive field (RF), and of suppression of these responses by stimuli in the surround. Each *dot* represents the absolute onset latency of a sorted SU in a given layer in response to a single stimulus condition. Different colors represent different stimulus conditions (as per legend). SUs are pooled across penetrations. *Solid lines*: medians.

Table S1

Mean and median absolute onset latency of CSD sinks across layers
(Related to data in Fig. 4)

RF				near _{square}				near _{annulus}				far			
Layer	n	Latency		Layer	n	Latency		Layer	n	Latency		Layer	n	Latency	
	17	Mean±sem (ms)	Median (ms)		13	Mean±sem (ms)	Median (ms)		7	Mean±sem (ms)	Median (ms)		11	Mean±sem (ms)	Median (ms)
1		62.2 ± 1.99	61.5	1		60.69 ± 2.11	58	1		65.7 ± 5.43	73.5	1		72.13 ± 6.27	66.5
2/3		62.2 ± 3.9	58.5	2/3		55.88 ± 1.7	56	2/3		58.21 ± 5.6	49.5	2/3		106.54 ± 10.27	105
Upp-4		55.67 ± 8.03	42	Upp-4		123.5 ± 12.3	122.5	Upp-4		55.58 ± 5.33	48.25	Upp-4		126.36 ± 8.58	122.5
4C		38.76 ± 1.46	36.5	4C		122.2 ± 9.58	113.25	4C		107 ± 11.23	102.75	4C		128.54 ± 8.44	125.5
Deep _U		53.05 ± 2.45	55.5	Deep_U		53.46 ± 1.74	54	Deep _U		71.35 ± 6.4	74	Deep _U		107.04 ± 9.38	110.5
Deep _L		98.85 ± 6.57	104	Deep _L		71.68 ± 6.96	60.5	Deep _L		62.92 ± 4.16	58	Deep_L		60.22 ± 2.11	59.5

Gray shading indicates the shortest latency across layers for a given stimulus condition

n= number of stimulus conditions

Table S2

Mean and median absolute onset latency of CSD sinks across layers
(Related to data in Fig. 5)

RF				near _{square}				near _{annulus}				far			
Layer	n	Latency		Layer	n	Latency		Layer	n	Latency		Layer	n	Latency	
	6	Mean±sem (ms)	Median (ms)		12	Mean±sem (ms)	Median (ms)		5	Mean±sem (ms)	Median (ms)		13	Mean±sem (ms)	Median (ms)
1		64.58 ± 3.52	63	1		60.16 ± 2.22	57.75	1		62.12 ± 6.16	62.75	1		72.45 ± 5.51	67.25
2/3		64.83 ± 10.14	56.75	2/3		55.04 ± 1.6	55.25	2/3		56 ± 4.76	49.5	2/3		99.65 ± 9.94	98
Upp-4		75.41 ± 21.23	48.5	Upp-4		119.72 ± 12.84	118	Upp-4		57.8 ± 6.51	49	Upp-4		114.88 ± 10.6	120.5
4C		40 ± 3.5	39.5	4C		117.5 ± 9.22	111	4C		107 ± 13.28	102.75	4C		129.83 ± 7.5	128.5
Deep _U		53.66 ± 6.51	51.5	Deep _U		53.25 ± 1.87	52.75	Deep _U		66.1 ± 6.34	74	Deep _U		103.07 ± 8.34	92
Deep _L		87.1 ± 11.87	95.5	Deep _L		69.7 ± 7.37	57.25	Deep _L		66.2 ± 5.09	73	Deep _L		63.46 ± 2.88	61.5

Gray shading indicates the shortest latency across layers for a given stimulus condition
n= number of stimulus conditions

Pairwise Ranksum Test across conditions within each layer
(Related to data in Fig. 5 and Table above)

Layer	RF vs near _{square}	RF vs near _{annulus}	RF vs far	near _{square} vs near _{annulus}	near _{square} vs far	near _{annulus} vs far
1	p=0.279	p=0.724	p=0.451	p= 0.930	p=0.1	p=0.47
2/3	p=0.768	p=0.571	p=0.030	p=0.739	p=0.0003	p=0.004
Upper 4	p=0.073	p=0.753	p=0.06	p=0.0046	p=0.772	p=0.004
4C	p=0.00016	p=0.0095	p=0.0001	p=0.661	p=0.207	p=0.228
Deep _U	p=0.733	p=0.225	p=0.0013	p=0.076	p=0.00008	p=0.013
Deep _L	p=0.254	p=0.421	p=0.168	p=0.953	p=0.951	p=0.793

Gray shading indicates statistically significant difference

Table S3

Pairwise Ranksum Test across layers and conditions: RF vs. surround
(Related to data in Fig. 5 and Table S2)

RF	near_{square}		near_{annulus}		far	
Layer	Layer	p	Layer	p	Layer	p
1	1	0.2788	1	0.7238	1	0.4507
	2/3	0.0191	2/3	0.2294	2/3	0.0204
	Upp-4	0.0184	Upp-4	0.3333	Upp-4	0.0154
	4C	0.0003	4C	0.019	4C	0.0002
	Deep _U	0.0127	Deep _U	0.5584	Deep _U	0.0044
	Deep _L	0.7722	Deep _L	0.8918	Deep _L	0.5649
2/3	1	0.7012	1	0.7619	1	0.2048
	2/3	0.7677	2/3	0.5714	2/3	0.0304
	Upp-4	0.0133	Upp-4	0.5325	Upp-4	0.0124
	4C	0.0048	4C	0.0381	4C	0.0013
	Deep _U	0.4776	Deep _U	0.5022	Deep _U	0.0086
	Deep _L	0.9578	Deep _L	0.4286	Deep _L	0.4284
Upper4	1	0.3231	1	0.6095	1	0.3727
	2/3	0.4523	2/3	0.7922	2/3	0.1479
	Upp-4	0.073	Upp-4	0.7532	Upp-4	0.06
	4C	0.1215	4C	0.3524	4C	0.0415
	Deep _U	0.6006	Deep _U	0.632	Deep _U	0.1333
	Deep _L	0.6354	Deep _L	0.4589	Deep _L	0.2729
4C	1	0.0003	1	0.0381	1	0.0005
	2/3	0.0008	2/3	0.0909	2/3	0.0002
	Upp-4	0.00016	Upp-4	0.1169	Upp-4	0.0001
	4C	0.00016	4C	0.0095	4C	0.0001
	Deep _U	0.0064	Deep _U	0.026	Deep _U	0.00007
	Deep _L	0.0047	Deep _L	0.0043	Deep _L	0.00007
Deep_U	1	0.2586	1	0.4762	1	0.0504
	2/3	0.5685	2/3	0.5671	2/3	0.0047
	Upp-4	0.0044	Upp-4	0.5022	Upp-4	0.0032
	4C	0.0003	4C	0.019	4C	0.0002
	Deep _U	0.7331	Deep _U	0.2251	Deep _U	0.0013
	Deep _L	0.2635	Deep _L	0.2468	Deep _L	0.0916
Deep_L	1	0.1218	1	0.2857	1	0.3963
	2/3	0.0452	2/3	0.0635	2/3	0.4921
	Upp-4	0.1355	Upp-4	0.0873	Upp-4	0.1249
	4C	0.19	4C	0.4127	4C	0.0136
	Deep _U	0.0213	Deep _U	0.3968	Deep _U	0.3996
	Deep _L	0.2544	Deep _L	0.4206	Deep _L	0.1683

Gray shading indicates a statistically significant difference

Table S4

Pairwise Ranksum Test across layers and conditions: near vs. far surround

(Related to data in Fig. 5 and Table S2)

near_{square}	far	
Layer	Layer	p
1	1	0.0997
	2/3	0.0018
	Upp-4	0.0023
	4C	0.00003
	Deep _U	0.0004
	Deep _L	0.4961
2/3	1	0.0093
	2/3	0.0003
	Upp-4	0.0011
	4C	0.00003
	Deep _U	0.0001
	Deep _L	0.0315
Upper 4	1	0.0089
	2/3	0.2464
	Upp-4	0.7717
	4C	0.4982
	Deep _U	0.3689
	Deep _L	0.0037
4C	1	0.0006
	2/3	0.1643
	Upp-4	0.9077
	4C	0.2071
	Deep _U	0.2836
	Deep _L	0.00004
Deep_U	1	0.0072
	2/3	0.0003
	Upp-4	0.0004
	4C	0.00003
	Deep _U	0.00007
	Deep _L	0.007
Deep_L	1	0.5309
	2/3	0.0323
	Upp-4	0.0057
	4C	0.0003
	Deep _U	0.0236
	Deep _L	0.9505

near_{annulus}	far	
Layer	Layer	p
1	1	0.4659
	2/3	0.0437
	Upp-4	0.0151
	4C	0.0011
	Deep _U	0.0101
	Deep _L	0.6
2/3	1	0.0866
	2/3	0.0044
	Upp-4	0.0028
	4C	0.0003
	Deep _U	0.0016
	Deep _L	0.166
Upper 4	1	0.1516
	2/3	0.0257
	Upp-4	0.0042
	4C	0.0006
	Deep _U	0.0026
	Deep _L	0.2771
4C	1	0.0297
	2/3	0.5311
	Upp-4	0.6899
	4C	0.2275
	Deep _U	0.895
	Deep _L	0.0017
Deep_U	1	0.9
	2/3	0.0628
	Upp-4	0.0128
	4C	0.0003
	Deep _U	0.0131
	Deep _L	0.4573
Deep_L	1	0.818
	2/3	0.0486
	Upp-4	0.0194
	4C	0.0003
	Deep _U	0.0133
	Deep _L	0.7934

Gray shading indicates a statistically significant difference

Table S5

Pairwise Ranksum Test across layers and conditions:

near_{annulus} vs. near_{square}

(Related to data in Fig. 5 and Table S2)

near_{annulus}	near_{square}	
Layer	Layer	p
1	1	0.9297
	2/3	0.6637
	Upp-4	0.0161
	4C	0.0015
	Deep _U	0.3011
	Deep _L	0.733
2/3	1	0.2838
	2/3	0.7392
	Upp-4	0.005
	4C	0.0004
	Deep _U	0.98
	Deep _L	0.4396
Upper 4	1	0.3672
	2/3	0.7023
	Upp-4	0.0046
	4C	0.0009
	Deep _U	0.8623
	Deep _L	0.4189
4C	1	0.0011
	2/3	0.0011
	Upp-4	0.7267
	4C	0.6608
	Deep_U	0.0011
	Deep _L	0.1059
Deep_U	1	0.2227
	2/3	0.076
	Upp-4	0.0252
	4C	0.00045
	Deep _U	0.0763
	Deep _L	1
Deep_L	1	0.4266
	2/3	0.086
	Upp-4	0.0252
	4C	0.00045
	Deep_U	0.0284
	Deep _L	0.953

Gray shading indicates a statistically significant difference

Table S6

Mean and median absolute onset latency of spiking response and surround suppression (SS) across layers
(Related to data in Fig. 7)

RF				full-SS				near-SS				tuned near-SS				far-SS			
Layer	n	Latency		Layer	n	Latency		Layer	n	Latency		Layer	n	Latency		Layer	n	Latency	
		Mean±sem (ms)	Median (ms)			Mean±sem (ms)	Median (ms)			Mean±sem (ms)	Median (ms)			Mean±sem (ms)	Median (ms)			Mean±sem (ms)	Median (ms)
2/3	25	44.8 ± 1.9	44.8	2/3	26	44.03 ± 1.97	40.8	2/3	25	53.12 ± 3.33	48.8	2/3	25	70.34 ± 6.87	56.8	2/3	23	53.9 ± 3.6	50.8
Upp-4	12	31.45 ± 2.3	30.8	Upp-4	9	33.9 ± 2.2	36.8	Upp-4	12	35.95 ± 3.0	34.8	Upp-4	12	45.46 ± 4.82	46.8	Upp-4	12	56.64 ± 3.7	54.8
4C	12	23.11 ± 1.9	20.7	4C	12	26.11 ± 1.6	26.8	4C	11	61.72 ± 7.4	56.8	4C	11	67.18 ± 5.49	66.8	4C	9	78.6 ± 12.1	64.8
Deep _U	13	31.1 ± 2.6	30.7	Deep _U	15	38.12 ± 3.6	34.8	Deep _U	13	45.57 ± 5.8	36.8	Deep _U	13	60 ± 9	58.8	Deep _U	9	41 ± 5.2	36.8
Deep _L	10	34 ± 3.98	35.8	Deep _L	14	54.52 ± 3.97	57.8	Deep _L	10	53.2 ± 8.8	52.8	Deep _L	9	85.5 ± 17.83	78.8	Deep _L	8	48.5 ± 7.8	39.8

Gray shading indicates the shortest latency across layers for a given stimulus condition
n= number of contacts

Pairwise Ranksum Test across conditions within each layer (Related to Fig. 7 and Table above)

Layer	RF vs full-SS	RF vs near-SS	RF vs tuned near-SS	RF vs far-SS	full-SS vs near-SS	full-SS vs tuned near-SS	full-SS vs far-SS	near-SS vs tuned near-SS	near-SS vs far-SS	tuned near-SS vs far-SS
2/3	p=0.62	p=0.08	p=0.0013	p=0.055	p=0.0367	p=0.0013	p=0.0435	p=0.0735	p=0.909	p=0.09
Upper 4	p=0.431	p=0.234	p=0.0256	p=0.0001	p=0.7751	p=0.1089	p=0.0002	p=0.13	p=0.0009	p=0.12
4C	p=0.242	p<0.0001	p<0.0001	p=0.0001	p<0.0001	p<0.0001	p=0.0001	p=0.29	p=0.128	p=0.73
Deep _U	p=0.239	p=0.071	p=0.0146	p=0.14	p=0.4181	p=0.0684	p=0.7422	p=0.28	p=0.66	p=0.19
Deep _L	p=0.004	p=0.1195	p=0.0082	P=0.151	p=0.5376	p=0.1956	p=0.3553	p=0.17	p=0.946	p=0.13

Gray shading indicates statistically significant difference

Table S7

Pairwise Ranksum Test across layers and conditions: RF vs. surround suppression

(Related to data in Fig. 7 and Table S6)

RF	full-SS		near-SS		tuned near-SS		far-SS	
Layer	Layer	p	Layer	p	Layer	p	Layer	p
L2/3	2/3	0.6293	2/3	0.0898	2/3	0.0013	2/3	0.055
	Upp-4	0.0049	Upp-4	0.0151	Upp-4	0.8705	Upp-4	0.0157
	4C	0.000005	4C	0.0401	4C	0.0002	4C	0.0001
	Deep _U	0.096	Deep _U	0.3854	Deep _U	0.2464	Deep _U	0.2313
	Deep _L	0.005	Deep _L	0.3493	Deep _L	0.0091	Deep _L	0.9495
Upper4	2/3	0.0012	2/3	0.00005	2/3	0.00002	2/3	0.0001
	Upp-4	0.4317	Upp-4	0.2342	Upp-4	0.0256	Upp-4	0.0001
	4C	0.1448	4C	0.0002	4C	0.00006	4C	0.0001
	Deep _U	0.1707	Deep _U	0.0528	Deep _U	0.0206	Deep _U	0.1248
	Deep _L	0.0004	Deep _L	0.0514	Deep _L	0.0062	Deep _L	0.0635
4C	2/3	0.000002	2/3	0.000001	2/3	0.000001	2/3	0.000003
	Upp-4	0.0047	Upp-4	0.004	Upp-4	0.0012	Upp-4	0.00003
	4C	0.242	4C	0.00005	4C	0.00005	4C	0.0001
	Deep _U	0.003	Deep _U	0.0007	Deep _U	0.0006	Deep _U	0.003
	Deep _L	0.0002	Deep _L	0.0023	Deep _L	0.0008	Deep _L	0.0015
Deep_U	2/3	0.0009	2/3	0.00006	2/3	0.00002	2/3	0.0001
	Upp-4	0.48	Upp-4	0.2404	Upp-4	0.0191	Upp-4	0.00006
	4C	0.19	4C	0.0002	4C	0.00004	4C	0.0001
	Deep _U	0.239	Deep _U	0.0715	Deep _U	0.0146	Deep _U	0.141
	Deep _L	0.0004	Deep _L	0.0466	Deep _L	0.004	Deep _L	0.059
Deep_L	2/3	0.0241	2/3	0.0017	2/3	0.0005	2/3	0.0028
	Upp-4	0.8433	Upp-4	0.6191	Upp-4	0.086	Upp-4	0.0009
	4C	0.112	4C	0.0021	4C	0.0004	4C	0.00008
	Deep _U	0.6357	Deep _U	0.2003	Deep _U	0.0402	Deep _U	0.3654
	Deep _L	0.004	Deep _L	0.1195	Deep _L	0.0082	Deep _L	0.159

Gray shading indicates statistically significant difference

Table S8

Pairwise Ranksum Test across layers and conditions: full-SS vs. near- and far-SS
(Related to data in Fig. 7 and Table S6)

full-SS	near-SS		tuned near-SS		far-SS	
Layer	Layer	p	Layer	p	Layer	p
L2/3	2/3	0.0367	2/3	0.0013	2/3	0.0435
	Upp-4	0.0776	Upp-4	0.67	Upp-4	0.0028
	4C	0.0089	4C	0.0001	4C	0.0001
	Deep _U	0.6111	Deep _U	0.2262	Deep _U	0.334
	Deep _L	0.4676	Deep _L	0.0138	Deep _L	0.9837
Upper4	2/3	0.0005	2/3	0.0002	2/3	0.0012
	Upp-4	0.7751	Upp-4	0.109	Upp-4	0.0002
	4C	0.0006	4C	0.0002	4C	0.00004
	Deep _U	0.2509	Deep _U	0.0605	Deep _U	0.4451
	Deep _L	0.1364	Deep _L	0.0059	Deep _L	0.1287
4C	2/3	0.000002	2/3	0.000002	2/3	0.000005
	Upp-4	0.0234	Upp-4	0.0048	Upp-4	0.00003
	4C	0.00005	4C	0.00005	4C	0.0001
	Deep _U	0.0029	Deep _U	0.0027	Deep _U	0.01
	Deep _L	0.0109	Deep _L	0.0022	Deep _L	0.0042
Deep_U	2/3	0.0049	2/3	0.0003	2/3	0.007
	Upp-4	0.695	Upp-4	0.3525	Upp-4	0.0053
	4C	0.0046	4C	0.0005	4C	0.0005
	Deep _U	0.4181	Deep _U	0.0684	Deep _U	0.742
	Deep _L	0.1331	Deep _L	0.0072	Deep _L	0.1737
Deep_L	2/3	0.297	2/3	0.428	2/3	0.3711
	Upp-4	0.0011	Upp-4	0.103	Upp-4	0.8972
	4C	1.0	4C	0.1385	4C	0.0578
	Deep _U	0.131	Deep _U	0.8457	Deep _U	0.0368
	Deep _L	0.5376	Deep _L	0.1956	Deep _L	0.3553

Gray shading indicates statistically significant difference

Table S9

Pairwise Ranksum Test across layers and conditions: near- vs. far-SS

(Related to data in Fig. 7 and Table S6)

near-SS	far-SS	
Layer	Layer	p
L2/3	2/3	0.909
	Upp-4	0.28
	4C	0.0074
	Deep_U	0.0418
	Deep _L	0.332
L4B	2/3	0.0014
	Upp-4	0.0009
	4C	0.0001
	Deep _U	0.5436
	Deep _L	0.2611
L4C	2/3	0.4722
	Upp-4	0.9754
	4C	0.1279
	Deep_U	0.0301
	Deep _L	0.1233
Deep_U	2/3	0.0918
	Upp-4	0.06
	4C	0.0109
	Deep _U	0.6627
	Deep _L	0.771
Deep_L	2/3	0.9062
	Upp-4	0.5746
	4C	0.067
	Deep _U	0.38
	Deep _L	0.9466

tuned near-SS	far-SS	
Layer	Layer	p
L2/3	2/3	0.0919
	Upp-4	0.495
	4C	0.3286
	Deep_U	0.0066
	Deep _L	0.088
L4B	2/3	0.17
	Upp-4	0.118
	4C	0.0076
	Deep _U	0.5442
	Deep _L	0.9692
L4C	2/3	0.0527
	Upp-4	0.0959
	4C	0.7321
	Deep_U	0.0062
	Deep _L	0.058
Deep_U	2/3	0.7919
	Upp-4	0.9348
	4C	0.2549
	Deep _U	0.1924
	Deep _L	0.514
Deep_L	2/3	0.089
	Upp-4	0.255
	4C	1.0
	Deep_U	0.0315
	Deep _L	0.1321

Gray shading indicates statistically significant difference

Table S10

**Pairwise Ranksum Test across layers and conditions:
near-SS vs. tuned near-SS**

(Related to data in Fig. 7 and Table S6)

Near-SS	tuned near-SS	
Layer	Layer	p
L2/3	2/3	0.0735
	Upp-4	0.306
	4C	0.0211
	Deep _U	0.7
	Deep _L	0.0785
Upper4	2/3	0.0001
	Upp-4	0.1386
	4C	0.0001
	Deep_U	0.049
	Deep_L	0.0094
4C	2/3	0.642
	Upp-4	0.206
	4C	0.292
	Deep _U	0.728
	Deep _L	0.361
Deep_U	2/3	0.01
	Upp-4	0.87
	4C	0.0136
	Deep _U	0.28
	Deep_L	0.0351
Deep_L	2/3	0.194
	Upp-4	0.6
	4C	0.168
	Deep _U	0.534
	Deep _L	0.174

Gray shading indicates statistically significant difference