Evidence of ventral stream computations in rat lateral extrastriate cortex

Giulio Matteucci [1], Rosilari Bellacosa Marotti [1], Federica B. Rosselli [1, 2, *] and Davide Zoccolan [1, *]

1. *Visual Neuroscience Lab, International School for Advanced Studies (SISSA), Trieste, 34136, Italy*

2. *Department of Behavior and Brain Organization, Center of Advanced European Studies and Research (Caesar), an Institute of the Max Planck Society, Bonn, 53175, Germany*

*Co-corresponding authors

Correspondence and requests for materials should be addressed to F.B.R. (federica.rosselli@caesar.de) or to D.Z. (zoccolan@sissa.it)

Abstract

In rodents, the extrastriate areas located laterally to primary visual cortex (V1) have been assigned to a putative object-processing pathway (homologous to the primate ventral stream), based on anatomical considerations. Yet, functional evidence supporting this attribution is sparse. Here, we probed the specialization for ventral processing of rat laterolateral area (LL), located at the apex of the putative ventral-like hierarchy. To this aim, we recorded neuronal responses in both V1 and LL during presentation of drifting, oriented gratings and noise movies. We found that the degree of non-linearity of the stimulus-response relationship increased from V1 to LL, along with the tendency of neurons to respond to multiple orientations, while the sharpness of orientation tuning decreased. Critically, all these trends matched the predictions derived from a well-established model of ventral computations and from a meta-analysis of the monkey literature, thus reaffirming the specialization of LL for higher-order processing of shape information.

Introduction

In rats and mice, primary visual cortex (V1) is bordered by at least 9 extrastriate visual areas, whose involvement in dorsal and ventral stream processing (i.e., extraction of, respectively, motion and shape information[1,2]) has become the object of intense investigation[3–8]. To date, most studies on this front have been inspired by the segregation of visual information in the specialized channels[9] that, in primates, selectively relay low temporal/high spatial frequency signals (P-channel) to the ventral stream only, and high temporal/low spatial frequency content (M-channel) to both the ventral and dorsal streams[10,11]. Inspired by a possible homology with primates, several investigators have mapped mouse visual areas with drifting gratings of various spatiotemporal frequencies[12–18], finding the signature of dorsal processing

in medial and parietal extrastriate cortex, but only scant evidence of ventral computations in lateral extrastriate regions. On the one hand, the validity of this approach is supported by the existence of distinct neuronal subpopulations within V1, displaying either M- or P-like tuning and making (receiving) target-specific projections with (from) downstream and upstream areas[12,15,19,20]. On the other hand, tracking the routing of spatiotemporal information across visual cortex does not take into account how local computations sculpt the tuning of visual neurons within each area[8] – a process that is expected to play a dominant role especially in the ventral stream, where the specialization for the processing of visual objects can only emerge as the result of local, non-linear integration of presynaptic inputs from upstream areas[21–23].

To find the signature of such computations, we recently tested how visual objects are represented along rat lateral extrastriate areas[24]. Our experiments showed that neuronal populations along this areas' progression become increasingly capable of supporting discrimination of object identity despite transformations in object appearance (e.g., due to translation and scaling), with the largest transformation tolerance achieved in the most ventral region: the laterolateral area (LL). This trend matches very closely the key property of a ventral-like pathway[21], thus strongly supporting the central role of LL in ventral processing. At the same time, studying cortical representations of visual objects has the disadvantage of relying on idiosyncratic choices of the stimulus set (the object conditions) and rather complex interpretive approaches (e.g., information theory and machine learning). This makes the experimental design hard to standardize, and often leads to conclusions that are only partially overlapping across studies. The assignment of rat lateral extrastriate cortex to the ventral stream is no exception, as shown by the much weaker specialization for object processing found, in this region, by an earlier study[25].

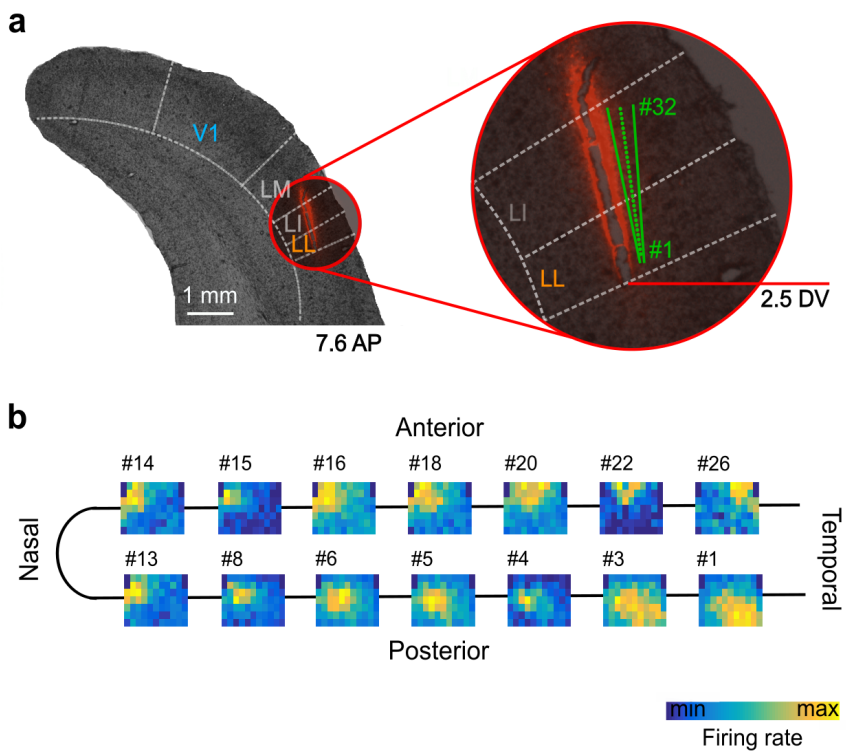
This prompted us to verify the higher rank of LL, relative to V1, in ventral processing, by designing an experiment that exploited the benefits of presenting parametric stimuli (drifting gratings), but, rather than focusing on spatiotemporal tuning, measured the linearity of the stimulus-response relationship – i.e., the extent to which a neuronal response can be explained as a weighted sum of the stimulus intensity values within the neuron's receptive field (RF). This property, which is the most discriminant of higher-order vs. lower-order processing, was assessed by quantifying how locked the neuronal response was to the phase of the gratings, thus distinguishing neurons with a linear, phase-sensitive behavior from neurons with a non-linear, phase-invariant response. We complemented this analysis by estimating the RF structure of the recorded units from their responses to noise movies using the *spike-triggered average* (STA) method[26], and by examining the sharpness and shape of their orientation tuning curves.

Critically, all these properties followed the trends of variations that are expected to occur along the ventral stream, according to the predictions derived from a well-established model of ventral computations and from a meta-analysis of the monkey literature. As such, our results confirm the involvement of LL in higher-order processing of shape information, and support the use of full-field parametric stimuli as a benchmark to screen rodent visual areas for their implication in ventral processing.

Results

We used 32-channel silicon probes to perform extracellular recordings from areas V1 and LL (Fig. 1a) of anaesthetized rats that were presented with drifting gratings of various orientations, spatial (SF) and temporal (TF) frequencies, as well as with 80 30s-long movies of spatially and temporally correlated noise (Methods). A total of 105 and 104 neurons were

91 recorded, respectively, from V1 and LL. With "neurons" we refer here to well-isolated single
92 units that were effectively and reproducibly driven by one or more of the grating stimuli
93 across repeated presentations (Methods). In addition to the main stimulus presentation
94 protocol, we also run a receptive field mapping procedure (Methods) to track the progression
95 of the RFs recorded along a probe (Fig. 1b). Given that, in rodents, the polarity of the
96 retinotopic map reverses at the boundary between each pair of adjacent visual cortical areas[8],
97 tracking the retinotopy of the RFs, combined with the histological reconstruction of the probe
98 insertion track (Fig. 1a), allowed a precise identification of the units that were recorded from
99 LL.



100

101 **Fig. 1** Targeting rat laterolateral (LL) visual cortex for multi-electrode recordings using a
102 linear silicon probe. **a.** Representative Nissl-stained coronal slice showing the insertion track
103 of a single-shank silicon probe aimed at area LL. The probe was inserted obliquely into
104 visual cortex, crossing lateromedial (LM) and lateralintermediate (LI) visual areas before

105 reaching LL. The dashed lines depict the borders of the areas, as reconstructed by tracking
106 the retinotopy of the receptive fields (RFs) recorded along the probe (see panel **b**). The
107 position of the probe during the recordings is marked by the staining with the fluorescent dye
108 (red) that was applied to it before insertion, and by the mechanical lesion produced by the
109 penetration. Based on this information, it was possible to estimate the geometry of the probe
110 (green lines) at the final recording depth (2.5 mm DV), along with the location of the
111 recording sites (green dots), from the tip (site #1) to the base (site #32). **b.** Firing intensity
112 maps displaying the RFs of the units recorded along the probe shown in **a**. The numbers
113 refer the sites each unit was recorded from. The reversal of the retinotopic progression (from
114 leftward to rightward) at sites #13-14 marks the boundary between areas LI and LL.
115

116 **Examples of tuning curves and spatiotemporal receptive fields in V1 and LL**

117 In both V1 and LL, neuronal responses to drifting gratings displayed a variety of tuning
118 profiles that matched well the RF structures inferred through STA. This is illustrated in Fig.
119 2, which reports, for each of two representative V1 (a-b) and LL (c-d) neurons: i) the tuning
120 curve as a function of the direction of the grating, when presented at the most effective SF
121 and TF; ii) the modulation of the firing rate elicited by the grating at the preferred direction;
122 and iii) the sequence of STA images showing the spatiotemporal evolution of the RF
123 structure. Each panel also reports the values of the metrics used to quantify the tuning
124 properties of the neurons (Methods). Orientation and direction tuning were measured using
125 the *orientation* and *direction selectivity indexes* (OSI and DSI), whose values range from 1
126 (maximal selectivity) to 0 (no selectivity), depending on the difference between the responses
127 to the preferred orientation/direction and the orthogonal/opposite one. We also computed a
128 *bimodal selectivity index* (BSI) to assess whether the tuning curve for orientation had
129 multiple peaks[27,28]. This index ranges from 0 (perfectly unimodal curve) to 1 (equal

130 responsiveness to two non-adjacent orientations), and can be taken as a measure of
131 complexity of tuning, with large BSI indicating tuning for multiple oriented features[27,28].

132 Finally, we computed a *modulation index* (MI) to quantify the extent to which the firing rate
133 of a neuron was entrained to the phase of the gratings[29,30], with MI > 3 indicating responses
134 that were strongly phase-locked to the stimulus temporal frequency. As for the
135 spatiotemporal filter estimated via STA, we used it as the linear stage of a Linear-Nonlinear
136 (LN) model of stimulus-response mapping[26] to predict the tuning curve of the neuron over the
137 direction axis and the time course of its response to the preferred grating (in both cases,
138 prediction accuracy was quantified as the fraction of response variance that was explained by
139 the model).

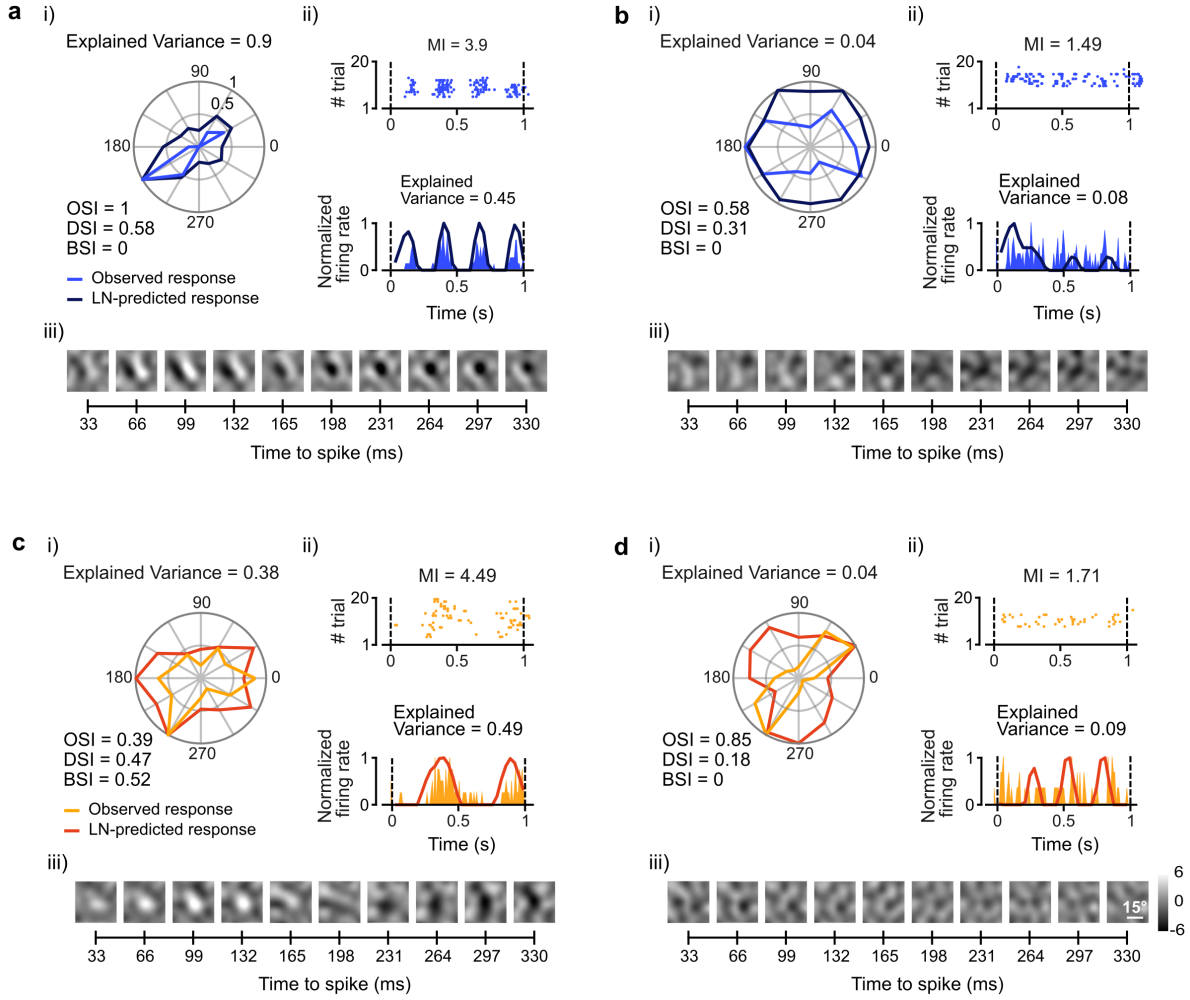
140 Fig. 2a shows an example V1 neuron with sharp orientation tuning, good direction
141 selectivity, and a response that was strongly modulated at the temporal frequency (4 Hz) of
142 the preferred grating. These properties match those of a *simple cell*[31], detecting the presence
143 of an oriented edge at a specific visual field location – hence the modulation of the firing rate,
144 produced by the phasic alternation of light and dark oriented stripes, while the preferred
145 grating swept across the neuron's RF. Such position-dependent neuronal firing suggests that
146 the response of the neuron to a visual pattern can be approximated as the sum of its responses
147 to the individual constituent elements (the "pixels") of the stimulus. Accordingly, the STA-
148 reconstructed RF (Fig. 2a.iii) had a Gabor-like, double-lobed appearance, with an orientation
149 matching the one corresponding to the peak of the tuning curve (Fig. 2a.i). When used as a
150 filter in a LN model, the STA-based RF accurately predicted the tuning of the neuron over
151 the direction and time axes (Fig. 2a.i and ii, respectively; black lines), thus confirming the
152 linearity of the stimulus-response relationship. By contrast, the example unit shown in Fig.
153 2b, while still being selective for a single orientation, displayed a highly non-linear behavior,
154 which is typical of the class of neurons known as *complex cells*[31]. This is shown by the lack

155 of modulation of the response to the preferred grating (Fig. 2b.ii), by the poorly defined

156 spatiotemporal structure of the STA-based RF (Fig. 2b.iii), and by the failure of the latter to

157 account for the tuning and time-modulation of the response (Fig. 2b.i-ii).

158



159 **Fig. 2** Examples of neuronal tuning in V1 and LL. Representative V1 simple (**a**) and complex

160 (**b**) cells are shown in the upper panels (blue/black lines). Representative LL simple-like (**c**)

161 and complex-like (**d**) cells are shown in the lower panels (yellow/orange lines). In each

162 panel, subpanels i-iii illustrate different aspects of neuronal tuning. (i) The polar plot reports

163 the observed (light color) and predicted (dark color) tuning of the neuron for the direction of a

164 drifting grating, when shown at the preferred spatial (SF) and temporal (TF) frequencies of

165 the unit. Predictions were obtained using a linear-nonlinear (LN) model, derived from the

166 spike-triggered average (STA) images shown in iii. For ease of comparison, both the
167 observed and predicted curves were normalized to their maximum. The plot also reports the
168 fraction of variance of the tuning curve that was explained by the LN-model, as well as the
169 values of several metrics that quantifies neuronal tuning: the orientation (OSI), direction
170 (DSI) and bimodal (BSI) selectivity indexes. (ii) Raster plot (top) and peri-stimulus time
171 histogram (PSTH; bottom; light color), showing the time course of the neuronal response to
172 a drifting grating, when presented at the preferred SF, TF and direction. The bottom plot also
173 reports the time course of the response, as predicted by the LN model (dark color), along
174 with the fraction of variance explained by the model. For ease of comparison, both the
175 observed and LN-predicted responses were normalized to their maximum. The top plot
176 reports the value of the modulation index (MI), which measured to what extent the neuronal
177 response was time-locked to the phase of the drifting grating. (iii) Spatial structure and time-
178 evolution of the STA-estimated RF of the neuron. In every STA image, each pixel intensity
179 value was independently z-scored, based on the null distribution of STA values obtained
180 through a permutation test (Methods). As such, all RF intensity values are reported as
181 distances (in units of standard deviation σ; gray scale bar in panel **d**.iii) from what expected
182 in the case of no frame-related, linear information carried by the spikes.

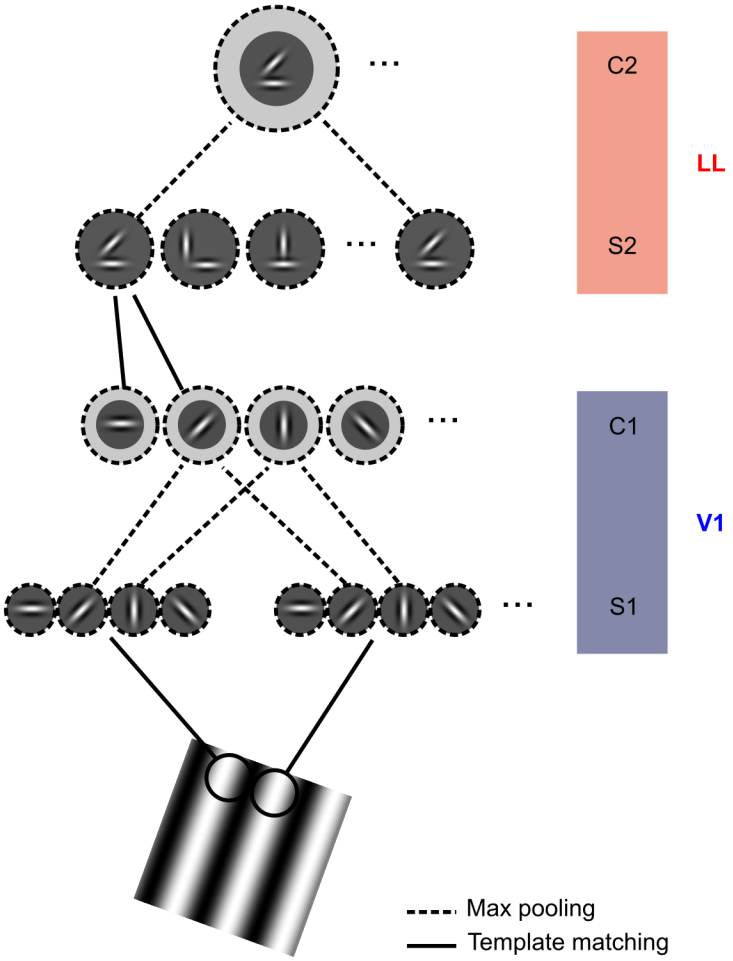
183

184 Interestingly, simple-like and complex-like units were also found in area LL. For
185 instance, Fig. 2c shows an LL neuron with a simple-like behavior, as indicated by the phase
186 modulation of its response (Fig. 2c.ii; yellow dots and line), which was predicted with
187 reasonable accuracy (orange line) by the STA-reconstructed RF. This is consistent with the
188 well-defined spatiotemporal structure of the RF (Fig. 2c.iii), which, however, was not able to
189 explain as much variance of the direction-tuning curve (Fig. 2c.i) as in the case of the
190 example V1 simple cell (Fig. 2a.i). This could be due to the sensitivity of the neuron for
191 multiple stimulus orientations, as shown by the large BSI value (0.52), which reflects the

192 presence of two peaks (at 0° and 240°) in the direction-tuning curve (Fig. 2c.i; yellow line) –
193 a behavior that cannot be captured by a RF with a linear, Gabor-like structure. Fig. 2d shows
194 instead an example LL unit that was sharply tuned for a single orientation, but displayed a
195 typical complex-like behavior: its response to the preferred grating was unmodulated (Fig.
196 2d.ii), and both the direction-tuning curve (Fig. 2d.i) and the response dynamics (Fig. 2d.ii)
197 were poorly explained by the STA-reconstructed RF (yellow vs. orange curves), which, in
198 turn, lacked any clear spatiotemporal structure (Fig. 2d.iii).

199 These examples illustrate how it is possible to compare the level of linearity of the
200 stimulus-response relationship for the neuronal populations recorded in V1 and LL, along
201 with the sharpens and complexity (e.g., bimodality) of their tuning for orientation. To guide
202 such comparison, we derived predictions about the evolution of neuronal tuning along an
203 object processing hierarchy, by assessing how our grating stimuli were represented in HMAX
204 – a neural network model of the ventral stream that is able to account for several qualitative
205 trends in the tuning properties of ventral neurons[23,27,32–35]. In its simplest implementation,
206 HMAX is a 4-layered, feed-forward network, with each layer consisting of either simple-like
207 (S) or complex-like (C) units (Fig. 3). The first layer (S1) is a bank of Gabor filters that
208 simulate V1 simple cells. In the second layer (C1), each unit performs a max pooling over a
209 set of S1 afferents with the same orientation tuning, but slightly different RF positions or
210 sizes. This yields orientation-tuned C1 units that are similar to V1 complex cells – i.e., they
211 have larger RFs (dashed circles in Fig. 3), with increased position- and size-tolerance, as
212 compared to their S1 afferents. In the third layer (S2), each unit performs a non-linear
213 template-matching computation over a set of C1 afferent units with different preferred
214 orientations. As a result, the S2 units are tuned for complex visual patterns, made of multiple
215 oriented features, like the neurons in higher-level stages of the primate ventral stream, such as
216 cortical area V4 and the inferotemporal cortex (IT). Finally, in the forth layer (C2), each unit

performs again a max pooling over S2 afferents that are tuned for the same visual pattern, but with slightly different RF positions and sizes. This results in C2 units with complex shape selectivity and large tolerance to position and size changes. Here, we used this basic 4-layered architecture to predict how the build-up of shape tuning and invariance should affect neuronal sensitivity to the orientation and temporal frequency of the gratings. In our simulations, layers S1 and C1 correspond to the simple and complex cells of V1, while layers S2 and, above all, C2, correspond to a downstream, higher-level area (such as LL).



226 **Fig. 3** Sketch of the HMAX model implementation used in this study. The model is a
227 feedforward, 4-layered neural network that processes arbitrary input images through a
228 cascade of computations. The first layer (S1) consists of a bank of Gabor filters performing
229 local edge detection on small input image patches (dark gray circles), similar to V1 simple
230 cells. In the second layer (C2), each unit pools over many iso-oriented S1 units via a max-
231 like operation (dashed arrows), thus gaining some amount of position and scale tolerance,
232 similar to V1 complex cells (such expansion of the invariance field is represented by the light
233 gray rings). In the third layer (S2), each unit acquires tuning for more complex visual patterns
234 by combining the output of many C1 units with different orientation via a non-linear template-
235 matching operation (solid arrows). In the forth layer, each C2 unit applies again the max-
236 pooling operation to S2 units with the same shape tuning but different RF positions and
237 sizes, in order to further gain tolerance to such transformations. In our application, the input
238 images were drifting gratings, similar to those used in the neurophysiology experiment, so as
239 to obtain predictions about the evolution of the relevant tuning indexes (MI, OSI and BSI; see
240 Fig. 2) across consecutive stages of a ventral-like hierarchy.

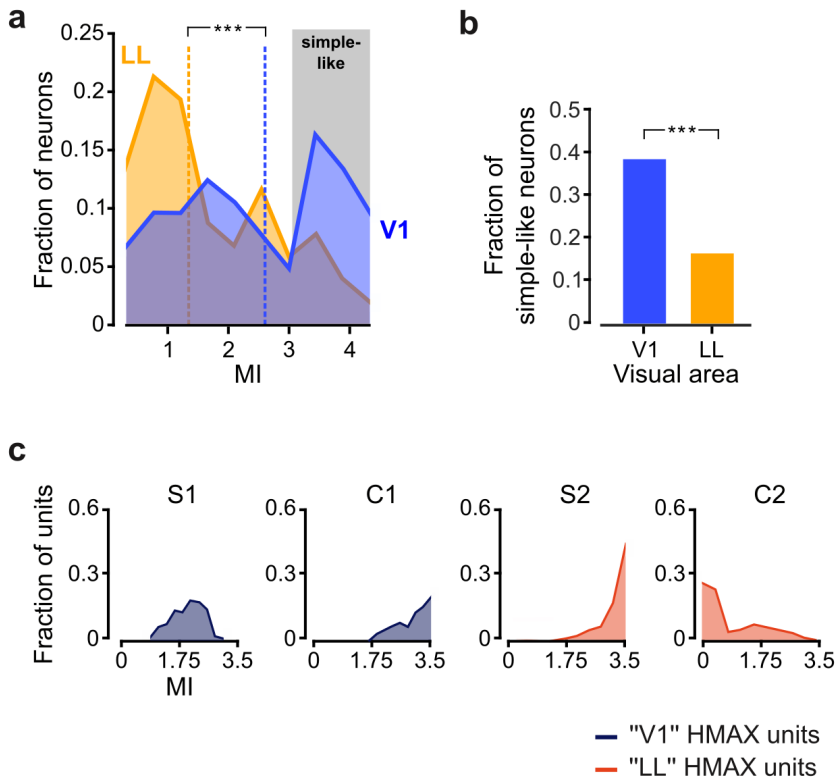
241 **The linearity of the stimulus-response relationship decreases from V1 to LL**

242 To establish whether V1 and LL neurons differed in terms of the linearity of the stimulus-
243 response relationship, we first compared the extent to which neuronal responses in the two
244 areas were modulated by the temporal phase of the drifting gratings. Despite the presence of
245 both strongly phase-modulated and unmodulated units in both areas, the distributions of the
246 modulation index (MI) were strikingly different in V1 and LL (Fig. 4a). In agreement with
247 earlier studies of rat[36] and mouse[16,37] primary visual cortex, the MI distribution obtained in
248 V1 (blue curve) revealed to the existence of two roughly equi-populated classes of neurons,
249 falling, respectively, below and above the MI = 3 threshold that distinguishes complex-like
250 from simple-like cells. By contrast, in LL, most MI values were lower than 3, following a
251 distribution (orange curve) with a prominent peak at MI<1, which is indicative of a strong

252 prevalence of complex-like units. As a result, the median MI was much lower in LL than in
253 V1 (1.28 vs. 2.55; orange vs. blue line) and this difference was highly significant ($p < 0.001$;
254 Mann-Whitney U-test). Similarly, the fraction of simple-like cells in LL was half as large as
255 in V1 (Fig. 4b; $p<0.001$, χ^2 test).

256 As illustrated in Fig. 2d.ii, the poor sensitivity to the phase of the drifting gratings
257 observed in LL indicates that most neurons in this area respond to phasic stimuli with a
258 sustained firing. This suggests that, in LL, most neurons are tolerant to position changes of
259 their preferred visual stimuli. This interpretation is supported by the evolution of the
260 modulation index observed in HMAX (Fig. 4c), whose upper layer (C2), following the max-
261 pooling operation to gain translation invariance, contained units with much lower phase
262 sensitivity, as compared to the units of the previous layers. As a result, the MI distribution
263 obtained in C2 was remarkably similar to the one observed in LL.

264



265 **Fig. 4** The entrainment of the neuronal response to the phase of drifting gratings decreases
266 from V1 to LL. **a.** Distributions of the modulation index (MI) for the populations of V1 (blue; *n*
267 = 105) and LL (orange; *n* = 104) neurons. Medians are shown as dashed lines (***$p < 0.001$;
268 one-tailed Mann-Whitney U-test). The gray-shaded area highlights the region of the MI axis
269 (MI > 3) corresponding to strongly phase-modulated units, i.e., simple-like cells. **b.** Fraction
270 of simple-like cells (i.e., units with MI > 3) in V1 (blue) and LL (orange; ***$p < 0.001$; χ^2 test).
271 **c.** Evolution of the MI distribution across the layers of the HMAX model (Fig. 3).

272

273 To further assess the linearity of the stimulus-response relationship in V1 and LL, we
274 evaluated the quality of the RFs inferred through the STA method. To this aim, each pixel
275 intensity value in a STA image was z-scored, based on the null distribution of STA values
276 obtained for that pixel, after randomly permuting the association between frames of the
277 movie and spike times 50 times. This allowed expressing the intensity values of the STA
278 image in terms of their difference (in units of standard deviation) from what expected in the
279 case of no frame-related information carried by the spikes. We then computed a *contrast*
280 *index* (CI; Methods and Supplementary Fig. 1) to measure the amount of signal contained in
281 the image and establish a conservative threshold (CI > 5.5), beyond which the linear RF
282 structure of a neuron could be considered as well defined.

283 Fig. 5a shows the 10 best STA-based RFs in each area, sorted according to the
284 magnitude of their CI, from largest to smallest. In V1, most of these RFs featured equally
285 prominent excitatory and inhibitory regions, often displaying a sharp, Gabor-like structure,
286 while in LL the best STA images typically presented a single lobe, with much lower contrast.
287 As a result, the median CI was higher for V1 than for LL neurons (blue vs. orange dashed
288 line in Fig. 5b; $p < 0.001$; one-tailed Mann-Whitney U-test), and the fraction of units with
289 well-defined linear RFs (i.e., with CI > 5.5) was 5 times larger in V1 than in LL (Fig. 5c; $p <$

290 0.001, χ^2 test). We also observed a stark difference between the two areas in terms of the
291 number of lobes contained in a STA image. To count the lobes, we binarized each image, by
292 applying a threshold to the modulus of its intensity values (Methods and Supplementary Fig.
293 1; this analysis was restricted to the units with well-defined linear RFs). Regardless of the
294 choice of the binarization threshold, the distribution of lobe counts peaked at 1 in both areas,
295 but had a heavier tail in V1 than in LL (compare matching rows in Fig. 5d), resulting, in most
296 cases, in a statistically significant difference between the two areas (Fisher's exact test; * $p <$
297 0.05; ** $p < 0.01$). Overall, this indicates that, in many cases, the responses of V1 neurons
298 could be well approximated as a linear, weighted sum of the luminance intensity values
299 falling inside their RFs – hence the sharp, multi-lobed STA images. By comparison, the
300 lower-contrast, structurally simpler RFs yielded by STA in LL suggest a dominant role of
301 non-linear computations in establishing the stimulus-response relationship in this area. In
302 fact, in the case of prevalently non-linear units, the STA method would be able to reconstruct
303 at most their linear sub-parts[38–40] – a sort of linear "residue" of the actual non-linear filter.
304 Such linear residue would still show, in some cases, the position and main orientation of the
305 underlying filter, but would fail short at accounting for the full richness and complexity of its
306 structure. This conclusion was supported by the tendency of the STA-based RFs to explain a
307 lower fraction of variance of the orientation tuning curves in LL, as compared to V1
308 (Supplementary Fig. 2). Taken together with the prevalence of complex-like units in LL (Fig.
309 4), these findings are consistent with the non-linear computations that are expected to build
310 invariance along an object-processing hierarchy.

311

312

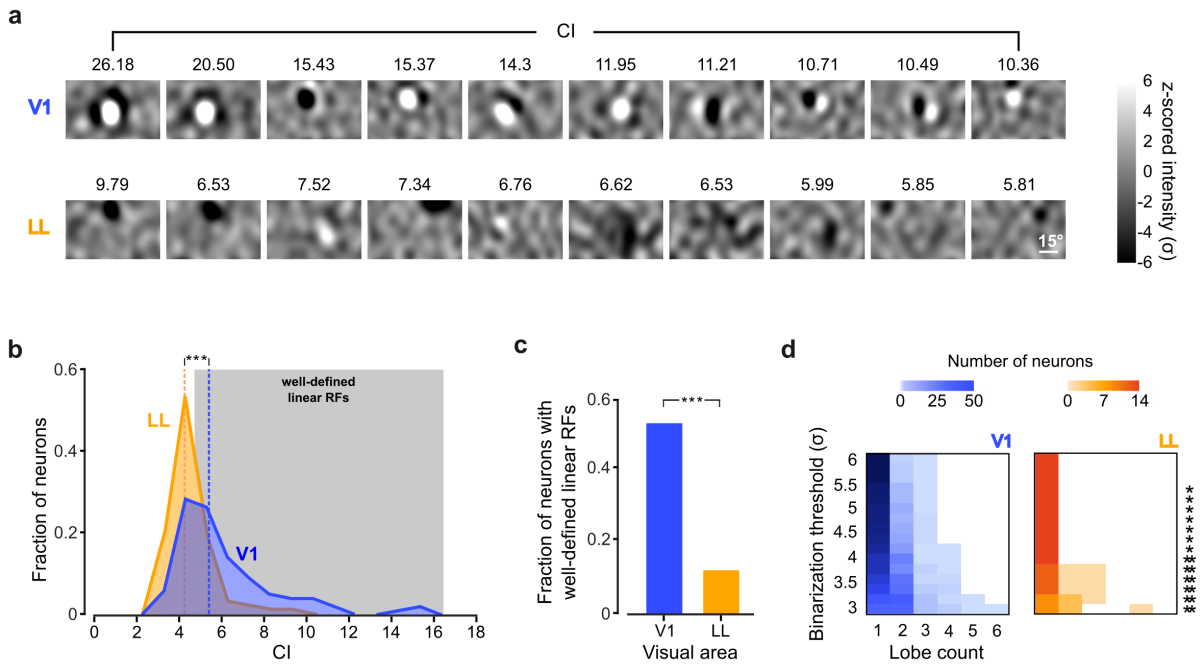


Fig. 5 Spike-triggered average yields better-defined and structurally richer receptive fields in V1 than in LL. **a.** Spatial structure of the ten best linear RFs obtained in V1 (top row) and LL (bottom row) using the STA method. The quality of each RF was assessed by computing a contrast index (CI) that measured the amount of signal contained in a STA image (Supplementary Fig. 1a and c and Methods). For each neuron, the RF shown here is taken at the time before the spike when CI was maximum (the corresponding CI value is reported above each STA image). As explained in Fig. 2, each pixel intensity value in a STA image was independently z-scored according to the null distribution yielded by a permutation test; the resulting z-scored values are reported here in units of standard deviation σ (gray scale bar). **b.** CI distributions for the populations of V1 (blue; n = 105) and LL (orange; n = 104) neurons. Medians are shown as dashed lines (***p < 0.001; one-tailed Mann-Whitney U-test). The gray-shaded area highlights the region of the CI axis (CI > 5.5) corresponding to units with well-defined linear RFs. **c.** Fraction of units with well-defined linear RFs (i.e., with CI > 5.5) in V1 (blue) and LL (orange; ***p < 0.001; χ^2 test). **d.** Heat maps showing the distributions of lobe counts (abscissa) for the units with well-defined linear RFs (i.e., with CI

328 > 5.5) in V1 (blue; n = 13) and LL (red; n = 56) as a function of the binarization threshold
329 (ordinate) used by the lobe-counting algorithm (Supplementary Fig. 1b and d and Methods).
330 Stars indicate threshold values for which the two distributions were significantly different (*p
331 < 0.05, **p < 0.01; Fisher exact test).

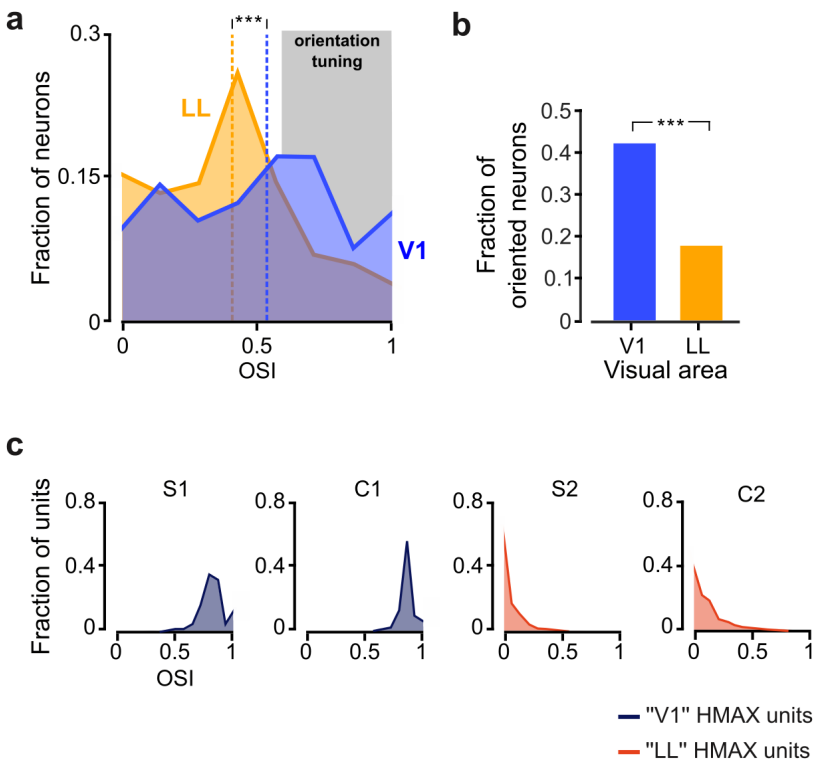
332 **Sharpness and complexity of orientation tuning follow opposite trends from V1 to LL**

333 Next, we compared the V1 and LL neuronal populations in terms of the shape and sharpness
334 of the orientation tuning curves obtained in the two areas. As shown in Fig. 6a, in area LL
335 most units had OSI lower than 0.5 (orange curve), with a median of 0.41 (orange line), while
336 in V1 the peak of the OSI distribution was in the 0.6-0.8 range, with a median of 0.53. The
337 difference between the two OSI distributions was highly significant ($p < 0.01$; Kolmogorov-
338 Smirnov test), as it was the difference between their medians ($p < 0.001$; one-tailed, Mann-
339 Whitney U-test). Following a convention established in previous studies (Supplementary
340 Table 2) we classified as orientation-selective the units with OSI > 0.6, and we compared the
341 fraction of units above such threshold in the two areas (Fig. 6b). In V1, the proportion of
342 selective cells was twice as large as in LL (40% vs. 18%), and such difference was, again,
343 highly significant ($p < 0.001$; χ^2 test).

344 To check the consistency of this finding with ventral-like processing, we looked at the
345 evolution of orientation tuning in HMAX. As shown in Fig. 6c, in the model, the OSI
346 distribution underwent an abrupt shift towards much lower values in the transition from the
347 C1 layer, corresponding to V1 complex cells, to the S2 layer, corresponding to a downstream
348 area, where tuning for combinations of multiple oriented elements is achieved using a non-
349 linear template matching operation (see Fig. 3). The shape of the OSI distribution was then
350 maintained in the C2 layer, performing the max-pooling operation over the S2 afferents. This
351 trend shows that, in HMAX, building tuning for complex visual patterns leads to a strong
352 reduction of orientation tuning. As such, the drop of orientation selectivity observed from V1

353 to LL can be taken as indication that LL neurons are tuned to more complex combinations of

354 visual features, as compared to V1 units.



355 **Fig. 6** Orientation selectivity decreases from V1 to LL. **a.** Distributions of the orientation

356 selectivity index (OSI) for the populations of V1 (blue; n = 105) and LL (orange; n = 104)

357 neurons. Medians are shown as dashed lines (**p < 0.01; one-tailed Mann-Whitney U-test).

358 The gray-shaded area highlights the region of the OSI axis (OSI > 0.6) corresponding to

359 orientation-selective units. **b.** Fraction of orientation-selective cells (i.e., units with OSI > 0.6)

360 in V1 (blue) and LL (orange; ***p < 0.001; χ^2 test). **c.** Evolution of the OSI distribution across

361 the layers of the HMAX model (Fig. 3).

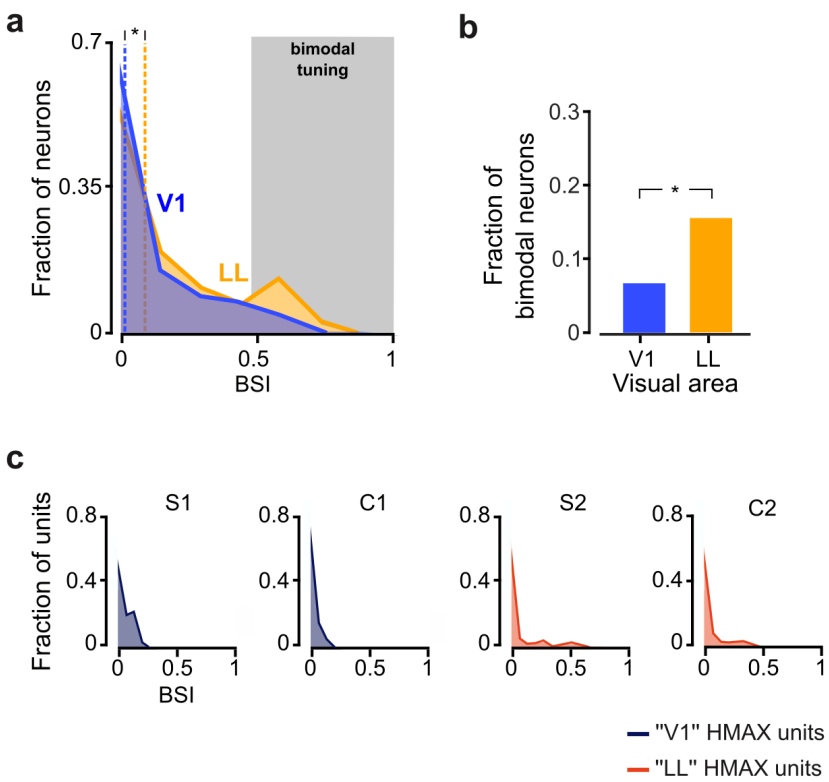
362

363 Support to this conclusion came from comparing the two areas in terms of the

364 tendency of the orientation tuning curves to be bimodal (Fig. 7a). In V1, most units had BSI

365 equal or close to zero (i.e., virtually perfect unimodal tuning), with the distribution dropping

366 sharply at larger BSI values (blue curve). In LL, the BSI distribution also peaked at zero, but
367 featured a heavier tail, with a secondary peak above 0.5 (orange curve). As a result, the two
368 distributions and their medians were significantly different ($p < 0.05$; Kolmogorov-Smirnov
369 and one-tailed, Mann-Whitney U-test, respectively) and the fraction of neurons with bimodal
370 tuning (i.e., with BSI > 0.5) was more than twice as large in LL than in V1 ($p < 0.05$, χ^2 test;
371 Fig. 7b). Once again, a similar trend was observed in HMAX (Fig. 7c), where the BSI
372 distribution displayed a heavier tail in the deeper layers (S2 and C2) than in the early ones
373 (S1 and C1). This supports the interpretation that the decrease of OSI from V1 to LL, along
374 with the concomitant increase of BSI, results from an increment of the complexity of tuning
375 of LL neurons – a property that is consistent with their involvement in ventral processing. By
376 contrast, no significant differences were found in terms of direction selectivity
377 (Supplementary Fig. 3), which suggests a lack of specialization of LL for motion processing,
378 as compared to V1.



379 **Fig. 7** The tendency to be tuned for multiple orientations increases from V1 to LL. **a.**
380 Distributions of the bimodal selectivity index (BSI) for the populations of V1 (blue; n = 105)
381 and LL (orange; n = 104) neurons. Medians are shown as dashed lines (*p < 0.05; one-tailed
382 Mann-Whitney U-test). The gray-shaded area highlights the region of the BSI axis (BSI >
383 0.5) corresponding to units with bimodal tuning. **b.** Fraction of cells with bimodal tuning (i.e.,
384 with BSI > 0.5) in V1 (blue) and LL (orange; *p < 0.05; χ^2 test). **c.** Evolution of the BSI
385 distribution across the layers of the HMAX model (Fig. 3).

386

387 ## Discussion

388 In this study, we investigated whether LL (the most temporal among rat lateral extrastriate
389 areas) displays tuning properties that are consistent with its involvement in ventral-like
390 processing of shape information. Previous anatomical and lesion work has causally
391 implicated rat lateral areas in object recognition[41,42], but only a handful of studies have
392 probed visual processing in these regions at the functional level[24,25,43,44]. Two of them, in
393 particular, have explicitly tested the aptness of LL at supporting transformation-tolerant
394 recognition of visual objects[24,25] – the hallmark property of the ventral pathway[21]. In both
395 cases, linear classifiers were applied to test the ability of LL neurons to support invariant
396 decoding of object identity, but only Tafazoli and colleagues[24] found LL to be superior to V1
397 in such a task. As explained in the Introduction, the conflicting findings of these previous rat
398 studies, as well as the limited functional understanding of ventral processing in mouse visual
399 cortex[3,6,8], motivated our search for further evidence of ventral computations in LL.

400 To be fully complementary with our previous study of rat ventral areas[24], we did not
401 probe LL representations using visual objects. With the goal of devising a more intuitive and
402 standardized approach to the study of ventral processing in rodents, we deployed instead
403 drifting oriented gratings, and we compared the V1 and LL neuronal populations in terms of:

404 1) the phase-sensitivity of their responses; 2) the goodness of the RF structures recovered
405 through STA; and 3) the sharpness and complexity (i.e., bimodality) of their tuning for
406 orientation. Obviously, this approach requires obtaining clear predictions about how these
407 properties should evolve along a ventral-like pathway. In our study, we achieved such
408 predictions based on the simulations of a computational model of the ventral stream (HMAX)
409 and a careful survey of the monkey literature.

410 HMAX is a feedforward neural network that was originally developed to explain the
411 emergence of selective, yet transformation-tolerant representations of visual objects along the
412 primate ventral stream[23]. Over the years, the model has provided useful computational
413 insights about shape tuning in V4 and IT[23,27,32–35], and human perception of visual objects[35,45].
414 The model shares the same basic architecture of more recent artificial vision systems (e.g.,
415 deep convolutional neural networks[46]), with alternating layers of selectivity-increasing and
416 tolerance-increasing pooling, but it is computationally simpler, since, in its basic
417 instantiation, it has only 4 layers and does not involve any supervised learning of feature
418 representations. While this can easily explain why HMAX has been surpassed by deep-
419 learning networks as a state of the art machine vision system[47,48], its simpler and static
420 structure makes it an ideal conceptual model to isolate the role of the selectivity- and
421 invariance-building operations in establishing the tuning properties of ventral neurons. As
422 shown by our simulations, HMAX makes very specific predictions about how the tuning for
423 oriented gratings should evolve from V1 to downstream ventral areas. Because of the
424 tolerance-building operation (the max-pooling), phase-sensitivity should decrease (Fig. 4c).
425 At the same time, because of the selectivity-building computation (the template-matching
426 pooling), orientation tuning should become smaller (Fig. 6c), while neurons acquire
427 preference for multiple orientations (Fig. 7c). Critically, all these trends were matched by the

428 empirical evidence gathered in our study, when comparing V1 to LL (Figs. 4a-b, 6a-b and 7a-
429 b), thus suggesting a specialization of the latter for ventral-like processing.

430 To check if this conclusion was also supported by a comparison with the monkey
431 literature, we performed a simple meta-analysis of the trends of variation of orientation
432 tuning and phase-sensitivity to drifting gratings across areas V1, V4 and IT of the ventral
433 stream. In V1, the fraction of simple cells reported by various authors is quite variable,
434 ranging from ~16% to 63% (Supplementary Table 1), with a mean across studies of ~41%
435 (blue bar in Fig. 8a). According to the few studies that investigated the phase-dependence of
436 neuronal responses to grating stimuli in downstream areas (Supplementary Table 1), the
437 proportion of simple-like cells diminishes to ~14% in V4, to become zero in IT (yellow and
438 orange bars in Fig. 8a), resulting in an overall significant decrement along the ventral stream
439 ($p < 0.01$, $F_{2,13} = 7.25$, one-way ANOVA). With regard to orientation selectivity, most
440 authors report a very large fraction of orientation-tuned cells in monkey V1 (Supplementary
441 Table 2), with a mean across studies close to 89% (blue bar in Fig. 8B). This proportion
442 significantly decreases along downstream areas ($p < 0.001$, $F_{2,14} = 46.17$, one-way ANOVA),
443 with ~65% and ~19% of orientation-tuned neurons reported, respectively, in V4 and IT
444 (yellow and orange bars in Fig. 8b; see also Supplementary Table 2). These trends are
445 confirmed by the few studies that applied the same experimental design to directly compare
446 orientation tuning in V1 and either V4[28] or IT[49,50]. Vogels and Orban[49,50] found ~85% of
447 orientation-tuned neurons in V1, compared to just ~14% in IT. David and colleagues[28]
448 reported a similar figure for V1, while in V4 they found that only 50% of neurons were
449 orientation-tuned. In addition, they found that the fraction of cells with bimodal tuning was
450 more than twice as large in V4 (~28%) than in V1 (~11%), in agreement with the increasing
451 proportion of neurons that, in V4, are tuned for the combinations of oriented elements found
452 in corners and curved boundaries[51–54].

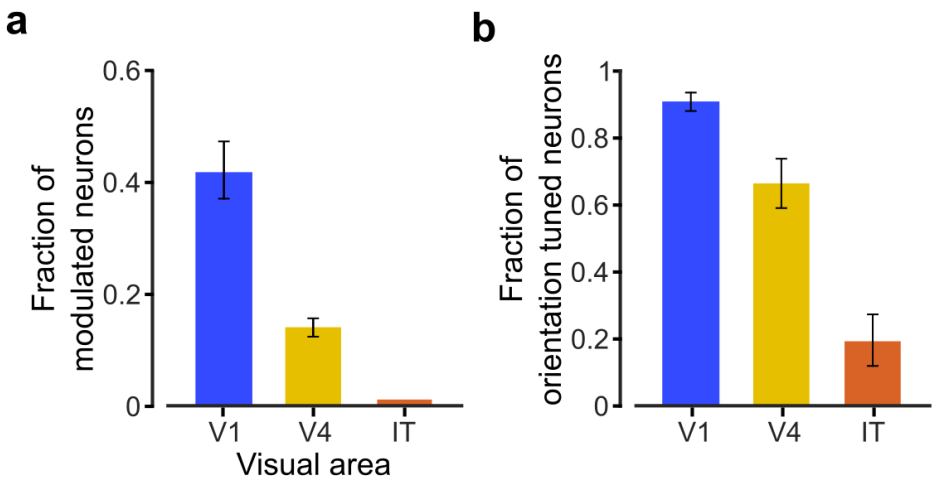


Fig. 8 Meta-analysis of the trends of variation of phase sensitivity and orientation tuning to drifting gratings across the primate ventral stream. **a.** Fraction of cells that have been reported as being strongly modulated by the phase of drifting gratings (i.e., simple-like cells) in areas V1 (red; n = 10), V4 (yellow; n = 3) and IT (red; n = 1) of the monkey ventral stream (mean across n studies in each area ± SE; see Supplementary Table 1). The drop of phase modulation across the three areas was statistically significant ($p < 0.01$, one-way ANOVA). **b.** Fraction of cells that have been classified as being tuned to the orientation of drifting gratings in areas V1 (red; n = 7), V4 (yellow; n = 4) and IT (red; n = 4) of the monkey ventral stream (mean across n studies in each area ± SE; see Supplementary Table 2). The drop of orientation tuning across the three areas was statistically significant ($p < 0.001$, one-way ANOVA).

Among rodent studies of visual extrastriate areas, our results are the first that are in complete agreement with the monkey literature reviewed above and are fully consistent with the computational principles thought to be at the root of ventral processing (as implemented in the HMAX simulations). In mice, area LL has not been identified, but areas LM (lateromedial) and LI (laterointermediate), which, as in rats, are located progressively more

470 lateral to V1 (see Fig. 1a), have been attributed to the ventral stream based on anatomical
471 considerations[55,56] and have been tested with drifting gratings in several electrophysiological
472 and imaging studies, often along with other medial and parietal visual areas[13,14,16,17,57].
473 Consistently with our findings (Fig. 4), the only study that measured the phase dependence of
474 neuronal responses in V1 and lateral visual areas of the mouse found that the latter had a
475 lower proportion of simple cells (15% vs. 38% in V1)[16]. Conversely, the authors reported no
476 difference between the areas in terms of OSI. This result contrasts with the general increase
477 of orientation tuning that was found in nearly all extrastriate areas (compared to V1) by
478 Marshell and colleagues[14], although the increment of OSI was much larger in putative dorsal-
479 stream areas than in LM and LI. A significant increase of OSI in LM, relative to V1, was also
480 observed by Tohmi and colleagues[18], while other authors found no differences between V1
481 and some of the putative dorsal-stream areas in terms of orientation tuning[13,17]. As for rat
482 cortex, Vermaercke and colleagues[25] found the opposite trend of that observed in our study –
483 i.e., a progressive increase of OSI from V1 to LL and beyond[25].

484 Different causes can be at the root of such discrepancies among rodent studies. In
485 mice, many experiments were based on calcium imaging, performed using different
486 indicators (e.g., OGB-1[14] vs. GCaMP3[13]), which diverge from a perfectly linear relationship
487 between fluorescence changes and underlying spiking activity in an-indicator specific way[58].
488 As pointed out by Niell[3], this could potentially distort the orientation tuning curves and bias
489 the OSI measurements, as compared to approaches where well-isolated single-units are
490 recorded with micro-electrodes, as in our study. Other important differences could concern
491 the specific stimuli that were used by different authors. For instance, while we presented very
492 large (60ºx110º), full-field gratings, Vermaercke and colleagues[25] used smaller (33º) circular
493 gratings, shown at the center of each recorded receptive field. Our presentation protocol
494 assured that, for most neurons in both areas, the stimuli covered the entire *classical* receptive

field and, very likely, a large fraction of the *extra-classical* receptive field, where surround-suppression contributes to shape tuning both in monkey[1,2] and rodent[59–63] visual cortex. By contrast, the smaller and more localized gratings used in [25] likely covered an increasingly smaller fraction of both the classical and extra-classical receptive fields along the V1-LM-LI-LL progression, given that RF size gradually increases along rat lateral extrastriate areas, being twice as large in LL as in V1[24]. This, in turn, may have prevented the tuning for multiple oriented elements, located at different RF locations, to emerge in neurons with larger RFs, thus artificially boosting OSI in higher-order areas. Finally, another factor that is worth considering is whether neuronal recordings were performed in anesthetized or awake animals. In general, both volatile (e.g. isofluorane) and injectable (e.g., urethane and ketamine) anesthetics can depress cortical activity in high-order areas[64], put the cortex in a highly synchronized state[65], and alter the excitation-inhibition balance that is typical of the wakeful state[59,66]. However, most of these shortcomings can be prevented by applying anesthesia protocols that, as the one used in our study, are based on a combination of dissociative analgesics (such as fentanyl) with α2 receptor agonists (such as medetomidine or xylazine). This form of anesthesia maximizes retinal function[67], allows for desynchronized cortical states, and results in stimulus-driven responses that are similar to those of the wakeful state[68,69]. In addition, the authors that compared neuronal activity in mouse visual cortex under anesthesia and during active wakefulness found that several key tuning properties, both in V1 and extrastriate areas, were very similar in the two conditions [37,70–73]. As such, it is very unlikely that the choice of using or not anesthesia determined the differences found in the rodent literature, with regard to orientation selectivity and other tuning properties.

Ultimately, four considerations support the strength and reliability of our conclusions, as compared to previous studies of rodent visual cortex based on gratings. First, the full agreement with the trends found across the monkey ventral stream. Second, the full

agreement with the predictions derived by simulating the basic computational architecture of the ventral stream. Third, the self-consistency of our findings – for instance, the reduction of orientation tuning from V1 to LL was matched by a concomitant increase of bimodal tuning, with both trends pointing to the development of selectivity for complex visual patterns in LL. Forth, the agreement with our previous investigation of rat lateral visual areas, where we found the signature of specialized processing of visual objects, and the consistency with the advanced processing of shape information reported, for rats, by several behavioral studies[74–76]. In the light of these arguments, we believe that our study provides strong evidence of the specialization of LL for ventral-like computations, thus adding crucial support to a still unsettled and debated question. At the methodological level, our study offers a standardized and easily interpretable way to carry out a large-scale screening of ventral processing across rodent visual cortical areas. This could help optical imaging studies finding the yet missing functional signature of ventral processing in mouse visual cortex (e.g., by analyzing existing large datasets of stimulus-response associations, as the one made available by the Allen Brain Observatory[77]). In this regard, having provided clear predictions about the trends of variations of several neuronal tuning properties, based on computational modeling and an exhaustive survey of the monkey literature, will likely serve as a reference to guide future explorations of ventral functions in both mice ad rats.

Methods

Animal preparation and surgery. All animal procedures were in agreement with international and institutional standards for the care and use of animals in research and were approved by the Italian Ministry of Health: project N. DGSAF 22791-A, submitted on Sep. 7, 2015 and approved on Dec. 10, 2015 (approval N. 1254/ 2015-PR). 18 naïve Long-Evans

544 male rats (Charles River Laboratories), with age 3-12 months and weight 300-700 grams,
545 underwent extracellular recordings in either primary visual cortex (V1) or laterolateral (LL)
546 extrastriate visual cortex. Each rat was anesthetized with an intraperitoneal (IP) injection of a
547 solution of 0.3 mg/kg of fentanyl (Fentanest®, Pfizer) and 0.3 mg/kg of medetomidin
548 (Domitor®, Orion Pharma). The level of anesthesia was monitored by checking the absence
549 of tail, ear and hind paw reflex, as well as monitoring blood oxygenation, heart and
550 respiratory rate through a pulse oximeter (Pulsesense-VET, Nonin). A constant flow of
551 oxygen was delivered to the rat throughout the experiment to prevent hypoxia. A constant
552 level of anesthesia was maintained through continuous IP infusion of the same anesthetic
553 solution used for induction, but at a lower concentration (0.1 mg/kg/h Fentanyl and 0.1 g/kg/h
554 Medetomidin), by means of a syringe pump (NE-500; New Era Pump Systems). Temperature
555 was thermostatically kept at 37°C using a heating pad to prevent anesthesia-induced
556 hypothermia.

557 After induction, the rat was secured to a stereotaxic apparatus (Narishige, SR-5R)
558 and, following a scalp incision, a craniotomy was performed over the target area in the left
559 hemisphere (typically, a 2x2 mm^2 window) and the dura was removed to allow the insertion
560 of the electrode array. When targeting V1, the coordinates of penetration were ~6.5 mm
561 posterior from bregma and ~4.5 mm left to the sagittal suture (i.e., AP 6.5, ML 4.5). When
562 targeting LL, the coordinates were ~1 mm anterior from lambda and ~1 mm medial from the
563 cranial ridge (i.e., AP 8, ML 5). Throughout the procedure, the rat eyes were protected from
564 direct light and kept hydrated by repeated application of an ophthalmic ointment (Epigel®,
565 Ceva Vetem).

566 Once the surgical procedure was completed, and prior to probe insertion, the stereotax
567 was placed on a rotating platform and the rat's left eye was covered with black, opaque tape,
568 while the right eye (placed at 30 cm distance from the monitor) was immobilized using a

569 metal eye-ring anchored to the stereotax. The platform was then rotated, so as to align the
570 right eye with the center of the stimulus display and bring the binocular portion of its visual
571 field to cover the left side of the display. During the recordings, the eye and cortex were
572 periodically irrigated using saline solution to keep them properly hydrated.

573 **Neuronal recordings.** Extracellular recordings were performed using single-shank 32-
574 channel silicon probes (NeuroNexus Technologies) with site recording area of 775 μm^2 and
575 25 μm of inter-site spacing. After grounding (by wiring the probe to the animal's head skin),
576 the electrode was manually lowered into the cortical tissue using an oil hydraulic
577 micromanipulator (Narishige, MO-10; typical insertion speed: ~ 5 μm/s), up to the chosen
578 insertion depth (~800-1,000 μm from the cortical surface when targeting V1, and ~2,500 μm
579 when targeting LL). To reach LL, the probe was tilted of 30° relative to the cortical surface
580 (Fig. 1a), whereas it was inserted perpendicularly into V1. Extracellular signals were
581 acquired using a system three workstation (Tucker-Davis Technologies) with a sampling rate
582 of 25 kHz. Before insertion, the probe was coated with Vybrant® DiI cell-labeling solution
583 (Invitrogen, Oregon, USA), to allow visualizing the probe insertion track *post-mortem*,
584 through histological procedures. To this aim, at the end of the recording session, an
585 electrolytic lesion was also performed by delivering current (5 μA for 2 seconds) through the
586 4 deepest channels at the tip of the shank.

587 **Visual stimuli.** During a recording session, two kinds of visual stimulation protocols were
588 administered to the rat.

589 A 15 min-long receptive field (RF) mapping procedure was used to precisely identify
590 the visual area each neuron was recorded from (see details below) and optimize the location
591 of the RFs for the main presentation protocol (i.e., ensure that most RFs fell inside the
592 monitor, by rotating the platform or repositioning the eye through adjustments of the eye-

593 ring). During the RF mapping protocol, the animal was presented with 10°-long drifting bars
594 with four different orientations (0º, 45º, 90º and 135º), tiling a grid of 66 different visual field
595 locations (i.e., 6 rows, spanning vertically 50°, and 11 columns, spanning horizontally 100°).
596 The bars were white over a black background. During presentation of these stimuli, multiunit
597 spiking responses (MUA) were plotted as a function of screen position in real-time, so as to
598 estimate the positions of the RF centers across the recording sites of the probe. This allowed
599 identifying in real-time when area LL was reached during an oblique probe insertion.
600 Specifically, area LL was identified by the third "rightwards" (i.e., from nasal to temporal)
601 reversal of the retinotopy at a recording depth close to 2.5 mm DV[24] (Fig. 1). The same
602 analysis was applied off-line to the well-isolated single units obtained by spike sorting (see
603 below) to precisely identify the neurons that were recorded from LL.

604 Once the probe was positioned in the final recording location, the main presentation
605 protocol was administered. The rat was presented with: 1) 1s-long drifting gratings, made of
606 all possible combinations of 3 spatial frequencies (SF; 0.02, 0.04 and 0.08 cpd), 3 temporal
607 frequencies (TF; 2, 4 and 8 Hz), and 12 directions (from 0° to 330°, in 30° increments); and
608 2) 30s-long movies of spatially and temporally correlated noise. To allow for a precise
609 estimate of neuronal tuning properties, each grating stimulus was presented in 20 repeated
610 trials, while 80 distinct noise movies were shown for an overall duration of 40 minutes. All
611 stimulus conditions were randomly interleaved, with a 1s-long inter stimulus interval (ISI),
612 during which the display was set to a uniform, middle-gray luminance level (later used to
613 estimate spontaneous firing rates; see below). To generate the movies, random white noise
614 movies were spatially correlated by convolving them with a Gaussian kernel. The kernel full
615 width at half maximum (FWHM) was chosen to match a 0.02 cpd SF, which was found to
616 elicit a robust spiking response in a series of pilot experiments. For the same reason (i.e., to

617 increase the chance of evoking spikes), temporal correlations were also introduced, by
618 convolving the movies with a causal exponential kernel with a 33 ms decay time-constant.

619 Stimuli were generated and controlled in MATLAB (The MathWorks) using the
620 Psychophysics Toolbox package and displayed with gamma correction on a 47-inch LCD
621 monitor (SHARP PNE471R) with 1920x1080 pixel resolution, 220 cd/m^2 maximum
622 brightness and spanning a visual angle of 110° azimuth and 60° elevation. Grating stimuli
623 were presented at 60 Hz refresh rate, whereas noise movies were played at 30 Hz.

624 **Histology.** At the end of the recording session, each animal was deeply anesthetized with an
625 overdose of urethane (1.5 gr/kg) and perfused transcardially with phosphate buffer saline
626 (PBS) 0.1 M, followed by 4% paraformaldehyde (PFA) in PBS 0.1 M, pH 7.4. The brain was
627 then removed from the skull, post-fixed in 4% PFA for 24 h at 4°C, and then immersed in
628 cryoprotectant solution (15% w/v sucrose in PBS 0.1 M, then 30% w/v sucrose in PBS 0.1
629 M) for 48 h at 4 °C. The brain was finally sectioned into 30µm-thick coronal slices using a
630 freezing microtome (Leica SM2000R, Nussloch, Germany). Sections were mounted
631 immediately on Superfrost Plus slides and let dry at room temperature overnight. A brief
632 wash in distilled water was performed, to remove the excess of crystal salt sedimented on the
633 slices, before inspecting them at the microscope. Each slice was then photographed with a
634 digital camera adapted to a Leica microscope (Leica DM6000B-CTR6000, Nussloch,
635 Germany), acquiring both a DiI fluorescence image (700 nm DiI filter) and a bright-field
636 image at 2.5X and 10X magnification. Following the acquisition of this set of images, the
637 sections displaying the electrode fluorescent track were further stained for Nissl substance
638 using a Cresyl Violet Acetate solution, and new pictures were taken at 2.5X and 10X
639 magnification. By superimposing the fluorescence, bright-field and Nissl-stained images, it

640 was possible to reconstruct the tilt and the anteroposterior (AP) position of the probe during
641 the recording session.

642 **Selection of the single units included in the analyses.** Responses of single units were
643 isolated offline by applying the spike sorting package KlustaKwik-Phy[78]. Automated spike
644 detection, feature extraction and expectation maximization (EM) clustering were followed by
645 manual refinement of the sorting using a customized version of the "Kwik-GUI" interface.
646 Specifically, the manual curation of the automatic output was performed by taking into
647 consideration many features of the candidate clusters: 1) the distance between their centroids
648 and their compactness in the space of the principal components of the waveforms (a key
649 measure of goodness of spike isolation); 2) the shape of the auto- and cross-correlograms
650 (important to decide whether to merge or not two clusters); 3) the variation, over time, of the
651 principal component coefficients of the waveform (important to detect and take into account
652 the effects of possible electrode drifts); and 4) the shape of the average waveform (to exclude,
653 as artifacts, clearly non-physiological signals). Clusters suspected to contain a mixture of one
654 or more single units were separated using the "reclustering" feature of the GUI (i.e., by
655 rerunning the EM clustering algorithm on the spikes of these clusters only). At the end of the
656 manual refinement step, only responsive, well-isolated single units, with reproducible firing
657 across repeated stimulus presentations, were included in the dataset used to perform all
658 subsequent analyses. Specifically, such units were defined by having: i) less than 0.5% of
659 "rogue" spikes within 2 ms in their autocorrelogram (i.e., units displaying a clear refractory
660 period); ii) a mean stimulus response firing rate of 2 spikes/s above baseline (i.e., responsive
661 units); and iii) responsiveness in at least 30% of the trials in one or more stimulus conditions
662 (i.e., stable isolation during the recording and reproducible stimulus-evoked response). We
663 also applied an additional, very loose screening on the reproducibility of the firing rate across
664 repeated stimulus presentations, by including only neurons with a mean correlation

665 coefficient of the direction-tuning curve across trials not lower than 0.03. The average
666 baseline (spontaneous) firing-rate of each well-isolated unit was computed by averaging its
667 spiking activity over every ISI condition. These criteria led to the selection of 105 units in V1
668 and 104 units in LL.

669 **Analysis of the responses to drifting gratings.** The response of a neuron to a drifting
670 grating was computed by counting how many spikes the neuron fired (on average, across
671 repeated presentations of the stimulus) during the whole duration of stimulus presentation,
672 and then subtracting the spontaneous firing rate (see above). To quantify the tuning of a
673 neuron for the orientation and direction of drifting gratings, we computed two standard
674 metrics, the orientation and direction selectivity indexes (OSI and DSI), which are defined as:
675 $\mathrm{OSI} = (R_{\mathrm{pref}} - R_{\mathrm{ortho}})/(R_{\mathrm{pref}} + R_{\mathrm{ortho}})$, and $\mathrm{DSI} = (R_{\mathrm{pref}} - R_{\mathrm{opposite}})/(R_{\mathrm{pref}} + R_{\mathrm{opposite}})$,
676 where R_{pref} is the response of the neuron to the preferred direction, R_{ortho} is the response to
677 the orthogonal direction, relative to the preferred one (i.e., $R_{\mathrm{ortho}} = R_{\mathrm{pref}} + \pi/2$), and
678 R_{opposite} is the response to the opposite direction, relative to the preferred one (i.e.,
679 $R_{\mathrm{opposite}} = R_{\mathrm{pref}} + \pi$). Values close to one indicate very sharp tuning, whereas values close
680 to zero are typical of untuned units.

681 The modulation of the spiking response to drifting gratings at the temporal frequency
682 f_1 of the stimulus was quantified by a modulation index (MI) adapted from [29] and defined as:

$$\mathrm{MI} = \left| \frac{PS(f_1) - \langle PS \rangle_f}{\sqrt{\langle PS^2 \rangle_f - \langle PS \rangle_f^2}} \right|,$$

683

684 where PS indicates the power spectral density of the stimulus-evoked response, i.e., of the
685 peri-stimulus time histogram (PSTH), and $\langle\ \rangle_f$ denotes the average over frequencies. This

686 metric measures the difference between the power of the response at the stimulus frequency
687 and the average value of the power spectrum in units of its standard deviation. The power
688 spectrum was computed by applying the Blackman-Tukey estimation method to the baseline-
689 subtracted, 10ms-binned PSTH. Being MI a standardized measure, values greater than 3 can
690 be interpreted as signaling a strong modulation of the firing rate at the stimulus frequency
691 (typical of simple-like cells), whereas values smaller than 3 indicate poor modulation (typical
692 of complex-like cells).

693 Bimodal tuning (i.e., the presence of two distinct peaks in the orientation-tuning
694 curve) was quantified by the Bimodal Selectivity Index[27,28], defined as
695 $\mathrm{BSI} = \left(R_{\text{2nd peak}} - R_{\text{2nd trough}}\right)/\left(R_{\text{1st peak}} - R_{\text{1st trough}}\right)$, where $R_{\text{1st peak}}$ and $R_{\text{2nd peak}}$ indicate
696 the two largest peaks in the orientation-tuning curve (with $R_{\text{1st peak}} > R_{\text{2nd peak}}$), and
697 $R_{\text{1st trough}}$ and $R_{\text{2nd trough}}$ indicate the deepest troughs (with $R_{\text{1st trough}} < R_{\text{2nd trough}}$). This
698 index takes values close to one for tuning curves displaying two equally high peaks and close
699 to 0 for tuning curves with a single peak.

700 **Analysis of the responses to the noise movies.** We used the spike-triggered average (STA)
701 analysis[26] to estimate the linear RF structure of each recorded neuron. The method was
702 applied to the spike trains fired by the neuron in response to the spatiotemporally correlated
703 noise (see above). The method yields an ordered sequence of images, each showing the linear
704 approximation of the spatial RF at different times before spike generation (see examples in
705 Fig. 2). To take into account the correlation structure of our stimulus ensemble and prevent
706 artifactual blurring of the reconstructed filters, we "decorrelated" the resulting STA images
707 by dividing them by the covariance matrix of the whole stimulus ensemble[26], using Tikhonov
708 regularization to handle covariance matrix inversion. Statistical significance of the STA
709 images was then assessed pixelwise via a permutation test. After randomly reshuffling the

710 spike times, the STA analysis was repeated multiple times (n=50) to derive a null distribution
711 of intensity values for the case of no linear stimulus-spike relationship. This allowed z-
712 scoring the actual STA intensity values using the mean and standard deviation of this null
713 distribution (Fig. 2 and 5). The temporal span of the spatiotemporal linear kernel
714 reconstructed via STA extended till to 330 ms before spike generation, corresponding to a
715 duration of 10 frames of the noise movie played at 30 Hz (Fig. 2).

716 After estimating the linear spatiotemporal structure of a RF with STA, we used it as a
717 filter in the input stage of a classical Linear-Nonlinear (LN) model of stimulus-response
718 relationship[26]. To obtain a prediction of the tuning of the neuron over the direction axis, the
719 sequence of frames of each drifting grating was fed as an input to the STA-estimated linear
720 filter. The output of the filter was then passed through a rectifying nonlinearity with unit gain
721 to obtain the final response of the model to each stimulus frame. We finally integrated the
722 response of the LN model through time to predict the activity of the neuron for each direction
723 of the tuning curve. The agreement between the LN-predicted and observed tuning curves
724 was quantified by computing the fraction of explained variance (dark vs. light tuning curves
725 in Fig. 2.i). In addition, we also quantified how well the time course of the neuronal response
726 to the preferred grating was explained by the time course of the LN-model prediction (dark
727 vs. light dynamics in Fig. 2.ii), again, in terms of explained variance.

728 We also estimated the amount of signal contained in a given STA image by defining a
729 quality metric based on a measure of maximal local contrast. Specifically, we used MATLAB
730 "rangefilt" function (MATLAB Image Processing toolbox) to compute the local contrast in a
731 STA image over convolutional patches with size equal to one fifth of the total image size and
732 stride 1 (the local contrast in every patch was measured as the difference between the
733 maximum and the minimum image intensity values within the patch itself). From the
734 resulting local-contrast map (Supplementary Fig. 1a and c, middle plot), we obtained a

735 distribution of contrast intensity values (Supplementary Fig. 1a and c, rightmost plot), and we
736 defined a contrast index (CI) by taking the 0.9 quantile of this distribution (dashed line; the
737 0.9 quantile was chosen, instead of simply using the maximum local-contrast value, to make
738 this measure more noise-robust). This CI metric was then used to quantify the amount of
739 signal contained in the STA image. A STA-based RF was considered as well defined if,
740 among the sequence of STA images corresponding to each pre-spike-generation time lag (see
741 examples in Fig. 2), the one with the highest contrast had CI > 5.5 (usually, such highest-
742 contrast image was the third in the sequence, corresponding to 66-99 ms before the spiking
743 event; see examples in Fig. 5a). It should be noted that, since the intensity values of the
744 original STA images (Supplementary Fig. 1a and c, leftmost plot) were expressed as z-scores
745 (see above), the 5.5 threshold can be interpreted in terms of peak-to-peak (i.e. white-to-black)
746 distance in sigma units of the z-scored STA values.

747 To further analyze the spatial structure of each well-defined, STA-based RF we
748 automatically counted how many distinct lobes the highest-contrast STA image obtained for a
749 unit contained. To this aim, the selected image was background-suppressed to enhance the
750 high-intensity regions within the RF and reduce noise-driven false lobe detections. We then
751 took the modulus of the resulting image and we set a threshold over its z-score values
752 (ranging from 3 to 6 units of standard deviations; Supplementary Fig. 1b and d). Based on
753 such binarized image, we computed: i) the centroid position of each simply connected region
754 (blue dots); and ii) the area-weighted center of mass (CM) of the image (green dot). Finally,
755 we defined a "region of acceptability" for the centroids of the simply connected regions to be
756 included in the lobe count. This region was a circle, centered on CM, with a radius that was
757 twice as large as the diameter of the largest simply connected region. Setting this
758 acceptability circle allowed excluding spurious, simply connected (often very small) regions
759 that were far away from the CM of the STA image (yellow dots in Supplementary Fig. 1b and

760 d), especially in the case of poorly defined RFs. The total number of accepted centroids at the
761 end of this selection procedure is the lobe count reported in Fig. 5d for the different
762 binarization thresholds.

763 **HMAX simulations.** The structure of the HMAX model of the ventral stream has been
764 already described in the Results (Fig. 3) and motivated in the Discussion. For our application,
765 we have chosen the version of the model described in [32] and downloaded from
766 http://maxlab.neuro.georgetown.edu/docs/hmax/hmaxMatlab.tar. Briefly, the model is a
767 feedforward neural network with alternating layers of units performing: 1) either a max-
768 pooling (invariance-building) operation over input units with the same feature-selectivity, but
769 RFs having different positions/scales (C1 and C2 units; dashed lines in Fig. 3); or 2) a
770 template-matching (selectivity-building) operation over input units with different feature
771 selectivity (S1 and S2 units; solid lines). As pointed out in the Discussion, the feature
772 representations built by the model in the S-type layers are not learned, and there is no attempt
773 to maximize classification accuracy in a given discrimination task. The S1 layer is simply a
774 bank of Gabor filters with various orientations, spatial frequencies, positions and scales
775 (similar to V1 simple cells), while the S2 units are tuned to random patches of images taken
776 from various databases (e.g., Caltech 101), with different units having as a template the same
777 image patch, but at different positions and scales. Such static, hardwired architecture makes
778 the model very suitable to isolate the role of the max-pooling and template-matching
779 computations in determining the tuning for oriented gratings along a ventral-like processing
780 hierarchy.

781 To this aim, we fed as an input to the network (i.e., to the S1 layer) a set of drifting
782 gratings spanning the same range of directions used in the neurophysiology experiment. Each
783 drifting grating was decomposed as a sequence of consecutive frames, and the activations of

784 the units of the network in response to such frames were computed to simulate the evolution
785 over time of the neuronal responses. Given a unit, its preferred stimulus direction was chosen
786 and the power spectrum of its time-dependent (i.e., frame-dependent) response was computed
787 to estimate the modulation index MI. We also integrated the activation of each unit over the
788 whole stimulus presentation to compute its overall mean response for a given grating
789 direction. By iterating this procedure for each stimulus direction, we built orientation and
790 direction tuning curves and computed OSI and BSI values. All these indexes were computed
791 for a randomly selected subset of units (n = 250) in each layer. The resulting distributions are
792 those shown in Figs. 4c, 6c and 7c.

793 **Statistical analysis.** Differences between median values of two distributions were quantified
794 using a Mann-Whitney U-test. Differences in the faction of units above or below a certain
795 threshold index value (for MI, OSI, BSI or CI) were quantified using the χ^2 test for
796 homogeneity with one degree of freedom. When needed (i.e., in case of small sample size),
797 Fisher exact test was used to compare two distributions (e.g., the lobe count distributions in
798 Fig. 5d). Finally, in the meta-analysis shown in Fig. 8, we compared the mean fractions of
799 modulated (Fig. 8a) and orientation-tuned (Fig. 8b) units across primate visual cortical areas
800 V1, V4 and IT using a one-way ANOVA.

801

802 ## References

803

804 1. Maunsell, J. H. R. & Newsome, and W. T. Visual Processing in Monkey Extrastriate
805 Cortex. *Annu. Rev. Neurosci.* **10,** 363–401 (1987).
806 2. Orban, G. A. Higher Order Visual Processing in Macaque Extrastriate Cortex. *Physiol.*
807 *Rev.* **88,** 59 (2008).

3. Niell, C. M. Exploring the Next Frontier of Mouse Vision. *Neuron* **72,** 889–892 (2011).
4. Niell, C. M. Cell Types, Circuits, and Receptive Fields in the Mouse Visual Cortex. *Annu. Rev. Neurosci.* **38,** 413–431 (2015).
5. Huberman, A. D. & Niell, C. M. What can mice tell us about how vision works? *Trends Neurosci.* **34,** 464–473 (2011).
6. Glickfeld, L. L., Reid, R. C. & Andermann, M. L. A mouse model of higher visual cortical function. *Curr. Opin. Neurobiol.* **24,** 28–33 (2014).
7. Katzner, S. & Weigelt, S. Visual cortical networks: of mice and men. *Curr. Opin. Neurobiol.* **23,** 202–206 (2013).
8. Glickfeld, L. L. & Olsen, S. R. Higher-Order Areas of the Mouse Visual Cortex. *Annu. Rev. Vis. Sci.* **3,** 251–273 (2017).
9. Livingstone, M. S. & Hubel, D. H. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J. Neurosci.* **7,** 3416–3468 (1987).
10. Maunsell, J. H., Nealey, T. A. & DePriest, D. D. Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *J. Neurosci.* **10,** 3323–3334 (1990).
11. Ferrera, V. P., Nealey, T. A. & Maunsell, J. H. Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. *J. Neurosci.* **14,** 2080–2088 (1994).
12. Gao, E., DeAngelis, G. C. & Burkhalter, A. Parallel Input Channels to Mouse Primary Visual Cortex. *J. Neurosci.* **30,** 5912–5926 (2010).
13. Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L. & Reid, R. C. Functional Specialization of Mouse Higher Visual Cortical Areas. *Neuron* **72,** 1025–1039 (2011).

14. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional Specialization of Seven Mouse Visual Cortical Areas. *Neuron* **72,** 1040–1054 (2011).
15. Glickfeld, L. L., Andermann, M. L., Bonin, V. & Reid, R. C. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nat. Neurosci.* **16,** 219–226 (2013).
16. Van den Bergh, G., Zhang, B., Arckens, L. & Chino, Y. M. Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys. *J. Comp. Neurol.* **518,** 2051–2070 (2010).
17. Roth, M. M., Helmchen, F. & Kampa, B. M. Distinct Functional Properties of Primary and Posteromedial Visual Area of Mouse Neocortex. *J. Neurosci.* **32,** 9716–9726 (2012).
18. Tohmi, M., Meguro, R., Tsukano, H., Hishida, R. & Shibuki, K. The Extrageniculate Visual Pathway Generates Distinct Response Properties in the Higher Visual Areas of Mice. *Curr. Biol.* **24,** 587–597 (2014).
19. Ji, W. *et al.* Modularity in the Organization of Mouse Primary Visual Cortex. *Neuron* **87,** 632–643 (2015).
20. Matsui, T. & Ohki, K. Target dependence of orientation and direction selectivity of corticocortical projection neurons in the mouse V1. *Front. Neural Circuits* **7,** (2013).
21. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How Does the Brain Solve Visual Object Recognition? *Neuron* **73,** 415–434 (2012).
22. Kourtzi, Z. & Connor, C. E. Neural Representations for Object Perception: Structure, Category, and Adaptive Coding. *Annu. Rev. Neurosci.* **34,** 45–67 (2011).
23. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat Neurosci* **2,** 1019–25 (1999).
24. Tafazoli, S. *et al.* Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife* **6,** e22794 (2017).

857 25. Vermaercke, B. *et al.* Functional specialization in rat occipital and temporal visual cortex.
858 *J. Neurophysiol.* jn.00737.2013 (2014). doi:10.1152/jn.00737.2013
859 26. Schwartz, O., Pillow, J. W., Rust, N. C. & Simoncelli, E. P. Spike-triggered neural
860 characterization. *J. Vis.* **6,** (2006).
861 27. Cadieu, C. *et al.* A model of V4 shape selectivity and invariance. *J Neurophysiol* **98,**
862 1733–50 (2007).
863 28. David, S. V., Hayden, B. Y. & Gallant, J. L. Spectral Receptive Field Properties Explain
864 Shape Selectivity in Area V4. *J. Neurophysiol.* **96,** 3492–3505 (2006).
865 29. Wypych, M. *et al.* Standardized F1 – A consistent measure of strength of modulation of
866 visual responses to sine-wave drifting gratings. *Vision Res.* **72,** 14–33 (2012).
867 30. Skottun, B. C. *et al.* Classifying simple and complex cells on the basis of response
868 modulation. *Vision Res.* **31,** 1078–1086 (1991).
869 31. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey
870 striate cortex. *J Physiol* **195,** 215–43 (1968).
871 32. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. & Poggio, T. Robust object recognition
872 with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell* **29,** 411–26 (2007).
873 33. Zoccolan, D., Kouh, M., Poggio, T. & Dicarlo, J. Trade-off between object selectivity
874 and tolerance in monkey inferotemporal cortex. *J Neurosci* **27,** 12292–307 (2007).
875 34. Leibo, J. Z., Liao, Q., Anselmi, F., Freiwald, W. A. & Poggio, T. View-Tolerant Face
876 Recognition and Hebbian Learning Imply Mirror-Symmetric Neural Tuning to Head
877 Orientation. *Curr. Biol.* **27,** 62–67 (2017).
878 35. Cox, P. H. & Riesenhuber, M. There Is a "U" in Clutter: Evidence for Robust Sparse
879 Codes Underlying Clutter Tolerance in Human Vision. *J. Neurosci.* **35,** 14148–14159
880 (2015).

881 36. Girman, S. V., Sauvé, Y. & Lund, R. D. Receptive Field Properties of Single Neurons in
882 Rat Primary Visual Cortex. *J. Neurophysiol.* **82,** 301–311 (1999).
883 37. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J*
884 *Neurosci* **28,** 7520–36 (2008).
885 38. Sedigh-Sarvestani, M., Fernandez-Lamo, I., Jaegle, A. & Taylor, M. M. Second Order
886 Receptive Field Properties of Simple and Complex Cells Support a New Standard Model
887 of Thalamocortical Circuitry in V1. *J. Neurosci.* **34,** 11177–11179 (2014).
888 39. Fournier, J., Monier, C., Pananceau, M. & Frégnac, Y. Adaptation of the simple or
889 complex nature of V1 receptive fields to visual statistics. *Nat. Neurosci.* **14,** 1053–1060
890 (2011).
891 40. Fournier, J. *et al.* Hidden Complexity of Synaptic Receptive Fields in Cat V1. *J.*
892 *Neurosci.* **34,** 5515–5528 (2014).
893 41. Aggleton, J. P., Keen, S., Warburton, E. C. & Bussey, T. J. Extensive cytotoxic lesions
894 involving both the rhinal cortices and area TE impair recognition but spare spatial
895 alternation in the rat. *Brain Res Bull* **43,** 279–87 (1997).
896 42. Wörtwein, G., Mogensen, J., Williams, G., Carlos, J. & Divac, I. Cortical area in the rat
897 that mediates visual pattern discrimination. *Acta Neurobiol. Exp. (Warsz.)* **54,** 365–376
898 (1993).
899 43. Vinken, K., Van den Bergh, G., Vermaercke, B. & Op de Beeck, H. P. Neural
900 Representations of Natural and Scrambled Movies Progressively Change from Rat Striate
901 to Temporal Cortex. *Cereb. Cortex N. Y. NY* **26,** 3310–3322 (2016).
902 44. Vinken, K., Vogels, R. & Op de Beeck, H. Recent Visual Experience Shapes Visual
903 Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement.
904 *Curr. Biol.* **27,** 914–919 (2017).

45. Serre, T., Oliva, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci U A* (2007).

46. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444 (2015).

47. Cadieu, C. F. *et al.* Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLOS Comput. Biol.* **10,** e1003963 (2014).

48. Khaligh-Razavi, S.-M. & Kriegeskorte, N. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLOS Comput. Biol.* **10,** e1003915 (2014).

49. Vogels, R. & Orban, G. A. Quantitative study of striate single unit responses in monkeys performing an orientation discrimination task. *Exp. Brain Res.* **84,** 1–11 (1991).

50. Vogels, R. & Orban, G. A. Responses of monkey infero-temporal units in an orientation discrimination task. in *Progress in Brain Research* (eds. Hicks, T. P., Molotchnikoff, S. & Ono, T.) **95,** 427–444 (Elsevier, 1993).

51. Gallant, J. L., Braun, J. & Essen, D. V. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* **259,** 100–103 (1993).

52. Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W. & Van Essen, D. C. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J. Neurophysiol.* **76,** 2718–2739 (1996).

53. Pasupathy, A. & Connor, C. E. Responses to Contour Features in Macaque Area V4. *J. Neurophysiol.* **82,** 2490–2502 (1999).

54. Pasupathy, A. & Connor, C. E. Population coding of shape in area V4. *Nat. Neurosci.* **5,** 1332–1338 (2002).

55. Wang, Q., Gao, E. & Burkhalter, A. Gateways of Ventral and Dorsal Streams in Mouse Visual Cortex. *J. Neurosci.* **31,** 1905–1918 (2011).

929 56. Wang, Q., Sporns, O. & Burkhalter, A. Network Analysis of Corticocortical Connections
930 Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. *J. Neurosci.* **32,**
931 4386–4399 (2012).
932 57. Smith, I. T., Townsend, L. B., Huh, R., Zhu, H. & Smith, S. L. Stream-dependent
933 development of higher visual cortical areas. *Nat. Neurosci.* **20,** 200 (2017).
934 58. Akerboom, J. *et al.* Optimization of a GCaMP Calcium Indicator for Neural Activity
935 Imaging. *J. Neurosci.* **32,** 13819–13840 (2012).
936 59. Vaiceliunaite, A., Erisken, S., Franzen, F., Katzner, S. & Busse, L. Spatial integration in
937 mouse primary visual cortex. *J. Neurophysiol.* **110,** 964–972 (2013).
938 60. Self, M. W. *et al.* Orientation-Tuned Surround Suppression in Mouse Visual Cortex. *J.*
939 *Neurosci.* **34,** 9290–9304 (2014).
940 61. Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J. & Scanziani, M. A neural circuit for
941 spatial summation in visual cortex. *Nature* **490,** 226–231 (2012).
942 62. Pecka, M., Han, Y., Sader, E. & Mrsic-Flogel, T. D. Experience-Dependent
943 Specialization of Receptive Field Surround for Selective Coding of Natural Scenes.
944 *Neuron* **84,** 457–469 (2014).
945 63. Alwis, D. S., Richards, K. L. & Price, N. S. C. Masking reduces orientation selectivity in
946 rat visual cortex. *J. Neurophysiol.* **116,** 2331–2341 (2016).
947 64. Heinke, W. & Schwarzbauer, C. Subanesthetic isoflurane affects task-induced brain
948 activation in a highly specific manner: a functional magnetic resonance imaging study.
949 *Anesthesiology* **94,** 973–981 (2001).
950 65. Steriade, M., Nunez, A. & Amzica, F. A novel slow (< 1 Hz) oscillation of neocortical
951 neurons in vivo: depolarizing and hyperpolarizing components. *J. Neurosci.* **13,** 3252–
952 3265 (1993).

953 66. Haider, B., Häusser, M. & Carandini, M. Inhibition dominates sensory responses in the
954 awake cortex. *Nature* **493,** 97–100 (2013).
955 67. Nair, G. *et al.* Effects of common anesthetics on eye movement and electroretinogram.
956 *Doc. Ophthalmol. Adv. Ophthalmol.* **122,** 163–176 (2011).
957 68. Altwegg-Boussac, T., Chavez, M., Mahon, S. & Charpier, S. Excitability and
958 responsiveness of rat barrel cortex neurons in the presence and absence of spontaneous
959 synaptic activity in vivo. *J. Physiol.* **592,** 3577–3595 (2014).
960 69. Constantinople, C. M. & Bruno, R. M. Effects and Mechanisms of Wakefulness on Local
961 Cortical Networks. *Neuron* **69,** 1061–1068 (2011).
962 70. Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in
963 mouse visual cortex. *Neuron* **65,** 472–9 (2010).
964 71. Froudarakis, E. *et al.* Population code in mouse V1 facilitates readout of natural scenes
965 through increased sparseness. *Nat. Neurosci.* **17,** 851–857 (2014).
966 72. Juavinett, A. L. & Callaway, E. M. Pattern and Component Motion Responses in Mouse
967 Visual Cortical Areas. *Curr. Biol.* **25,** 1759–1764 (2015).
968 73. Durand, S. *et al.* A Comparison of Visual Response Properties in the Lateral Geniculate
969 Nucleus and Primary Visual Cortex of Awake and Anesthetized Mice. *J. Neurosci.* **36,**
970 12144–12156 (2016).
971 74. Rosselli, F. B., Alemi, A., Ansuini, A. & Zoccolan, D. Object similarity affects the
972 perceptual strategy underlying invariant visual object recognition in rats. *Front. Neural*
973 *Circuits* **9,** 10 (2015).
974 75. Alemi-Neissi, A., Rosselli, F. B. & Zoccolan, D. Multifeatural Shape Processing in Rats
975 Engaged in Invariant Visual Object Recognition. *J. Neurosci.* **33,** 5939–5956 (2013).

976 76. Djurdjevic, V., Ansuini, A., Bertolini, D., Macke, J. H. & Zoccolan, D. Accuracy of Rats
977 in Discriminating Visual Objects Is Explained by the Complexity of Their Perceptual
978 Strategy. *Curr. Biol.* **28,** 1005-1015.e5 (2018).
979 77. Hawrylycz, M. *et al.* Inferring cortical function in the mouse visual system through large-
980 scale systems neuroscience. *Proc. Natl. Acad. Sci.* **113,** 7337–7344 (2016).
981 78. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19,** 634–
982 641 (2016).
983

984 ## Acknowledgments

985 This work was supported by a Human Frontier Science Program Grant (DZ, contract n.
986 RGP0015/2013) and a European Research Council Consolidator Grant (DZ, project n.
987 616803-LEARN2SEE). We thank Margherita Riggi for her technical support with the
988 histological procedures.

989

990 ## Author contributions

991 F.B.R., G.M. and D.Z. conceived the study. G.M. designed the stimuli and implemented the
992 presentation protocol. F.B.R., G.M. and R.B.M. performed the experiments. F.B.R., G.M.,
993 R.B.M. and D.Z. analyzed the data. G.M. performed the simulations of the computational
994 model. F.B.R. and G.M. carried out the meta-analysis of the ventral stream literature. F.B.R.,
995 G.M. and R.B.M. did the artwork. F.B.R., G.M., R.B.M. and D.Z. wrote the manuscript.

996

997 **Competing interests.** The authors declare no competing financial interests.

Materials & Correspondence. Correspondence and material requests should be addressed to Davide Zoccolan (zoccolan@sissa.it)