www.sciencemag.org/cgi/content/full/330/6005/845/DC1

Supporting Online Material for

Functional Compartmentalization and Viewpoint Generalization within the Macaque Face Processing System

Winrich A. Freiwald* and Doris Y. Tsao*

*To whom correspondence should be addressed.
E-mail: wfreiwald@rockefeller.edu (W.A.F.), dortsao@caltech.edu (D.Y.T.)

Published 5 November 2010, *Science* **330**, 845 (2010)
DOI: 10.1126/science.1194908

This PDF file includes:

Methods

All procedures conformed to local and US National Institutes of Health guidelines, including the US National Institutes of Health Guide for Care and Use of Laboratory Animals, regulations for the welfare of experimental animals issued by the German federal government, and stipulations of local authorities.

Detailed descriptions of surgical, scanning and training procedures are given in (*1*). In brief, two male rhesus macaques (*Macaca mulatta*) were implanted with Ultem headposts that were fixed to the skull via dental acrylic and ceramic screws. The animals were then trained via standard operant conditioning techniques to maintain fixation on a spot of light on a CRT screen at 57 cm distance for a juice reward, and then scanned in a 3 T horizontal bore magnet to identify face-selective regions. Both animals had six prominent face-selective regions at stereotyped locations in the temporal lobe. In each animal, three of these regions were targeted for single-unit recordings (M1: MF, AL, AM; M2: ML, AL, AM). Because we only designed the FV stimulus set after recordings from ML in monkey M2 had been completed, a third monkey (M3) was used to provide ML data for the FV set. See (*2-3*) for further details on fMRI and single-unit recording procedures. All three animals used in this study grew up in breeding colonies, where they had experience with the faces of several dozen other monkeys, and with the faces of human caretakers. M1 arrived in the laboratory vivarium at age 3, M2 at age 2, and M3 at age 2. From their arrival on, the animals had experience with the faces of four to six other monkeys they were housed with, and daily and extensive experience with the faces of human caretakers, experimenters, and more infrequently, other members of the laboratory. At the beginning of electrophysiological recording experiments, monkeys M1 and M2 were 4 years old, M3 6 years, thus all had ample experience with real-life human faces of various identities and changing orientations.

Alert monkey fMRI. All fMRI procedures for identifying the face patches were identical to those described in (*2*). We used a multi-echo sequence (EPI, TR 3 or 4 s, TE 30 or 25 ms, 64 x 64 matrix, 1.25 mm isotropic resolution). In combination with a concomitantly acquired fieldmap, this allowed high fidelity reconstruction by undistorting most of the B0-field inhomogeneities (*4-5*), MION contrast agent was used to improve signal/noise ratio. Face patches were determined

by identifying regions responding significantly more to faces than to bodies, fruits, gadgets, hands, and scrambled patterns, and were confirmed across multiple independent scan sessions.

Targeting a face patch for single-unit recording. Targeting of face patches required a two-step procedure: 1) computing skull position and orientation of a recording chamber to allow access to the face patch, and 2) computing which grid angle and grid hole to use within the chamber. To accomplish step 1, we used a custom MATLAB program to position a virtual chamber such that it would reach the desired face patch and the rim of the chamber would lie over the skull. This yielded coordinates specifying the center of the chamber axis on the skull as well as two angles for this axis (in the coronal and sagittal planes; in most instances the angle in the sagittal plane was 0°). To accomplish step 2, we acquired an anatomical scan of the monkey's brain with a recording grid positioned inside the cylinder whose holes were filled with MR-visible silicone. Registering the face patch localizer data to this anatomical scan allowed us to determine which grid hole to use to reach the desired face patch. In most cases we recorded from only 1-3 grid holes in each animal; the pattern of overlying vessels precluded exhaustive sampling of the fMRI-identified face-selective region. The use of angled grids allowed us to reach face patches lying outside the direct projection of the chamber. Thus, for example, we were able to target three different face patches (ML, AL, and AM) from the same recording chamber in monkey M1. Although both animals M1 and M2 had both an ML and an MF, the pattern of overlying blood vessels allowed only one to be reached in each animal.

Single-unit recording. We recorded extracellularly with fine electropolished Tungsten electrodes coated with vinyl lacquer. Extracellular signals were amplified and fed into a dual-window discriminator. The spike train was recorded at 1 ms resolution. Only well-isolated single units that showed a refractory period were studied. The LFP was bandpass filtered between 3 Hz and 90 Hz or 0.7 Hz and 170 Hz. Data was acquired using a PC running RF3, a program for visual stimulus delivery and data acquisition developed by the late David Freeman and by Margaret Livingstone. Eye position was monitored with an ISCAN primate infrared eye tracking system. The monkey sat in a dark box with its head rigidly fixed and was given a juice reward for keeping fixation for 3 s in a 2.5° window.

Visual stimuli. The face patch localizer for fMRI experiments has been previously described (*1, 6*). Stimuli were generated in MATLAB using the Psychophysics Toolbox extensions, and were presented with an LCD projector (1024 x 768 pixels, 75 Hz refresh rate), onto a screen 53 cm in front of the monkey's eyes. The display spanned 28º laterally and 21º vertically. The size of the images for the fMRI experiments was 11° x 11°. For single-unit recordings, two image sets were used, a set of 128 images consisting of 16 faces, 16 human bodies, 16 fruits and vegetables, 16 gadgets, 16 hands, 16 scrambled patterns, 16 monkey body parts, and 16 monkey whole bodies (one example from each of these categories is shown in Fig. 2), and a set of 200 images consisting of 25 different individuals at left full profile, left half profile, straight, right half profile, right full profile, up, down, and back. Thus pictures of the same 25 individuals were presented at each of these 8 head orientations (one example at each view angle is shown in Fig. 2). Of the 25 individuals in the FV set, 24 were unknown and one was well known to the animals. Pictures were presented for 200 ms, separated by 200 ms blank interval. For each of the two stimulus sets, images were presented in random order, between 3-10 times each. Stimuli were presented at a 60 Hz monitor refresh rate and 640 x 480 resolution on a CRT monitor. The monitor was positioned 53 cm in front of the monkey's eyes. Stimuli were 7° x 7° and presented in the center of the screen. To evaluate size and position invariance (fig. S10), a set of 40 images was used comprising 8 frontal faces, 8 left full profiles (of the same identities), 8 right full profiles, and 16 non-face objects; example images are shown in fig. S9A. For the size tuning experiment, the 40 images were presented at 5 different sizes (1.12°, 2.24°, 4.48°, 8.96°, and 17.92°), with different images and sizes interleaved in random order. For the position tuning experiment, the 40 images were presented at 13 different positions and four eccentricities (0°, 1.75°, 5.95°, and 12.67°) along the vertical and horizontal meridians. For each cell, we performed two versions of the position tuning experiment, with the sizes of images either scaled for eccentricity (2.38° 3.22° 5.18° 8.33°; based on the formula for V4 RF size as a function of eccentricity given in (*7*)), equivalent to 34, 46, 74, and 119 pixels) or held fixed at 8.33° (all sizes refer to the size of the square bitmap, as shown in fig. S10A). The overall pattern of results was the same for both sets of position stimuli, and the results presented in the paper are for images at fixed size.

To generate the rotating head for Fig. 3, a custom C program, utilizing a commercial face modelling software package, was used to render the head. The head was refreshed at 6 Hz. In each frame, a random set of values for up down angle (-90° to 90°), left right angle (-180° to 180°), and picture plane angle (-180° to 180°) were chosen. A standard frontal head (the same head was used for all cells) was first rotated by picture plane angle, then by up down angle, and finally by left right angle. All rotations occurred in object-centered rather than absolute coordinates.

Data analysis. Single-unit data were analyzed using custom programs written in MATLAB. **Fig. 1C-E:** the Face Selectivity Index (FSI) was defined by FSI = $(R_{faces} - R_{nonface\ objects}) / (R_{face} + R_{nonface\ objects})$, where R_{faces} was the mean response above baseline to faces and $R_{nonface\ objects}$ the mean response above baseline to non-face objects. An FSI of 0 indicates equal responses to face and non-face objects. An FSI of 0.33 indicated twice as strong response to faces as to non-face objects. For cases where ($R_{face} > 0$) and ($R_{nonface\ objects} < 0$), FSI was set to 1; for cases where ($R_{face} < 0$) and ($R_{nonface\ objects} > 0$), FSI was set to -1. Cells were considered visually responsive if they gave a response above one standard deviation of the baseline for 20 consecutive bins, starting at a time point less than 300 ms post-stimulus onset. Baseline was defined as the mean response across all images over the first 50 ms, i.e. prior to any observable firing rate increases. Also, cells were required to fire at least a total of 300 spikes to be considered visually responsive. For each visually responsive cell, we determined the response latency by first computing the average response to each of the 8 categories; we then took the latency as the time at which the response of the best category exceeded 3 standard deviations above baseline. Throughout the paper, to compute the mean normalized response to an image, for each cell, the mean baseline response was subtracted, and the responses to all images normalized by the maximum response. **Fig. 1F, fig. S8:** For each cell, sparseness was defined as $\frac{(\sum_{i=1}^{N} R_i / N)^2}{(\sum_{i=1}^{N} R_i^2 / N)}$, where N = # stimuli (128 or 175) (*8*), with responses R_i not baseline subtracted. The view invariant identity correlation coefficient for a cell was defined as the mean correlation between the 25-element response vector at the preferred head orientation and the 6 response vectors at the remaining non-backwards head orientations. **Figs. 4A-C, fig. S4:** Classical multi-dimensional scaling (*9*) was performed on the population responses to the 200-image face views set of all

visually-responsive cells in each patch, using a Euclidean distance metric and the MATLAB command cmdscale. **Figs. 4D-F:** The analyses for this figure were based on average activity over the stimulus duration (200 ms), beginning at the response latency of the cell; we did not consider the temporal structure of the response time course (which could provide additional information for identification). A 200 x 200 similarity matrix of correlation coefficients was computed between the population response vector (across all visually responsive cells, averaged over stimulus repeats) to each of the 200 stimuli in the FV data set. **Fig. 4G:** Identity tuning half-widths were computed from responses of cells to the 25 identities in the FV set at the preferred head orientation. Width at half maximum was computed for each cell by sorting responses to the 25 identities and finding the index with the smallest of all responses greater than half the maximum response. **Fig. 4H:** Head orientation tuning depth was computed using the mean response to frontal faces ($R_{frontal}$), and the mean response to full profile faces in the preferred direction ($R_{profile}$) as follows: Tuning Depth = ($R_{frontal}$ - $R_{profile}$) / ($R_{frontal}$ + $R_{profile}$); responses not baseline subtracted. Tuning depths close to 0 indicate broad tuning. **Fig. 4I:** To compute view-invariant identity selectivity as a function of time, at each time point *t* between 0 and 600 ms following stimuli onset in increments of 10 ms, a similarity matrix was computed from mean responses between t and t+200 ms in visually-responsive cells. We then computed a "view-invariant identity selectivity index" as follows: For each 200 x 200 similarity matrix, we extracted the 175 x 175 submatrix ("X") of population distances to non-backwards views. We then computed the mean correlation along the off-center diagonals {y=x+25, y=x+50, y=x+75, y=x+100, y=x+125, y=x+150} of X, with wrap-around: view-invariant identity selectivity index

$$= \frac{\sum_{i=}^{\{25,50,75,100,125,150\}} \sum_{j=1}^{175} X(j, \mathrm{mod}(j+i-1,175)+1)/(6*175)}{\sum_{i=}^{\{1,2,...,174\}/\{25,50,75,100,125,150\}} \sum_{j=1}^{175} X(j, \mathrm{mod}(j+i-1,175)+1)/(168*175)}.$$

fig. S5B: For each cell, the mean response to the 25 individuals at each of the 8 head orientations was computed, and the head orientation yielding the maximum of these 8 values was considered the preferred head orientation. The responses to the 25 individuals at the preferred head orientation were sorted from minimum to maximum. Then the 25 responses at each of the remaining 7 head orientations (ranked according to mean response) were sorted according to the same ordering as that at the best head orientation. The averages across the population are shown

as 7 traces. Thus an increasing trace indicates presence of view-invariant identity tuning. **fig. S8A:** To compute the view-invariant identity selectivity index as a function of sparseness, for every value of sparseness between 0 and 1 in steps of 0.05, we computed a similarity matrix using cells with sparseness between *s* and *s*+0.3. We then computed the view-invariant identity selectivity index as in Fig. 4I. **fig. S10D:** A set of 40 images was used to assess size and position invariance. For each cell, the position-invariant identity selectivity index was defined as the mean correlation coefficient between the response vector to the 40 images at each of the 12 eccentric positions and the response vector to the images presented at the fovea. Similarly, the size-invariant identity selectivity index was defined as the mean correlation coefficient between the response vector to the 40 images at size {2.24°, 4.48°, 8.96°} with the response vector to the images presented at size 17.92°. **fig. S11A:** For each recording site, a stimulus-dependent LFP map was computed by averaging the evoked LFP response to repeated presentations of each stimulus; the map was normalized by subtracting the mean and dividing by the maximum absolute value. For each animal, maps were averaged across recording sites, normalized, and then resulting maps from both animals were averaged together and normalized.

References

1. S. Moeller, W. A. Freiwald, D. Y. Tsao, *Science* **320**, 1355 (2008).
2. D. Y. Tsao, W. A. Freiwald, R. B. H. Tootell, M. S. Livingstone, *Science* **311**, 670 (2006).
3. W. A. Freiwald, D. Y. Tsao, M. S. Livingstone, *Nature Neuroscience* **12**, 1187 (2009).
4. H. Zeng, R. T. Constable, *Magn Reson Med* **48**, 137 (Jul, 2002).
5. R. Cusack, M. Brett, K. Osswald, *Neuroimage* **18**, 127 (Jan, 2003).
6. D. Y. Tsao, S. Moeller, W. A. Freiwald, *Proceedings of the National Academy of Sciences* **105**, 19513 (2008).
7. A. T. Smith, K. D. Singh, A. L. Williams, M. W. Greenlee, *Cereb Cortex* **11**, 1182 (Dec, 2001).
8. E. T. Rolls, H. D. Critchley, A. S. Browning, K. Inoue, *Experimental Brain Research* **170**, 74 (2006).
9. J. B. Kruskal, M. Wish, in *Sage University Paper series on Quantitative Application in the Social Sciences*. (Sage Publications, Beverly Hills and London, 1978), pp. 7-11.

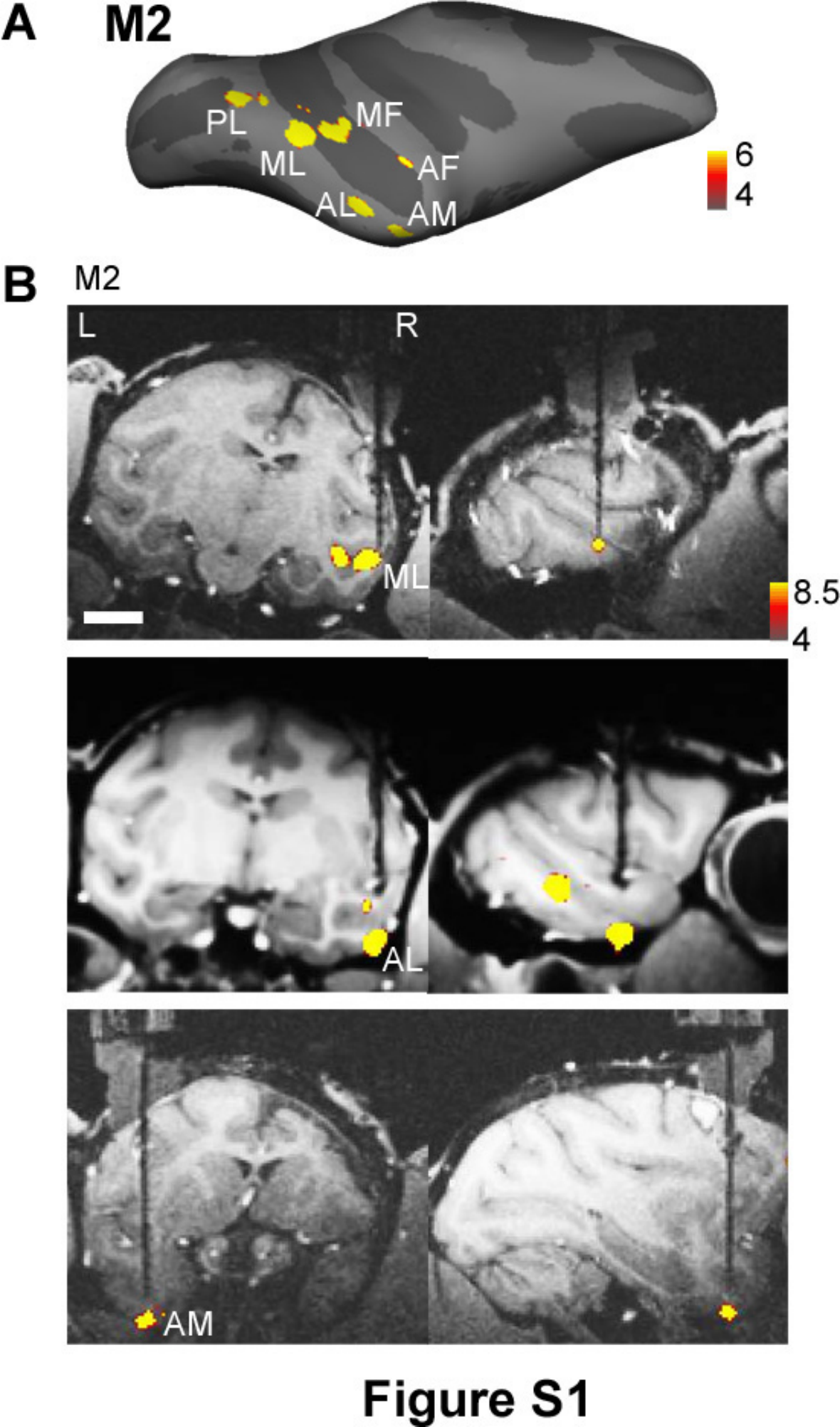


Figure S1

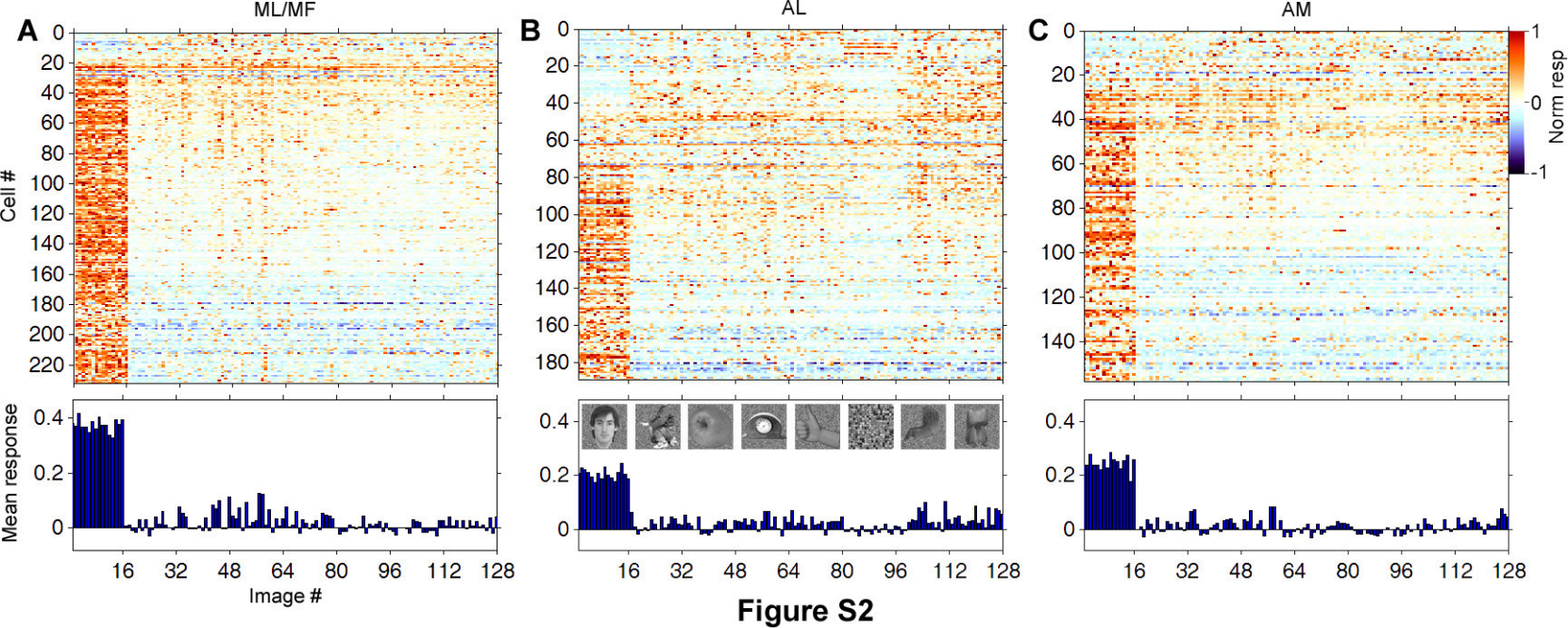

A
ML/MF
B
AL
C
AM
Cell #
Mean response
Image #
Norm resp
1
0
-1
0
20
40
60
80
100
120
140
160
180
200
220
0.4
0.2
16
32
48
64
80
96
112
128

Figure S2

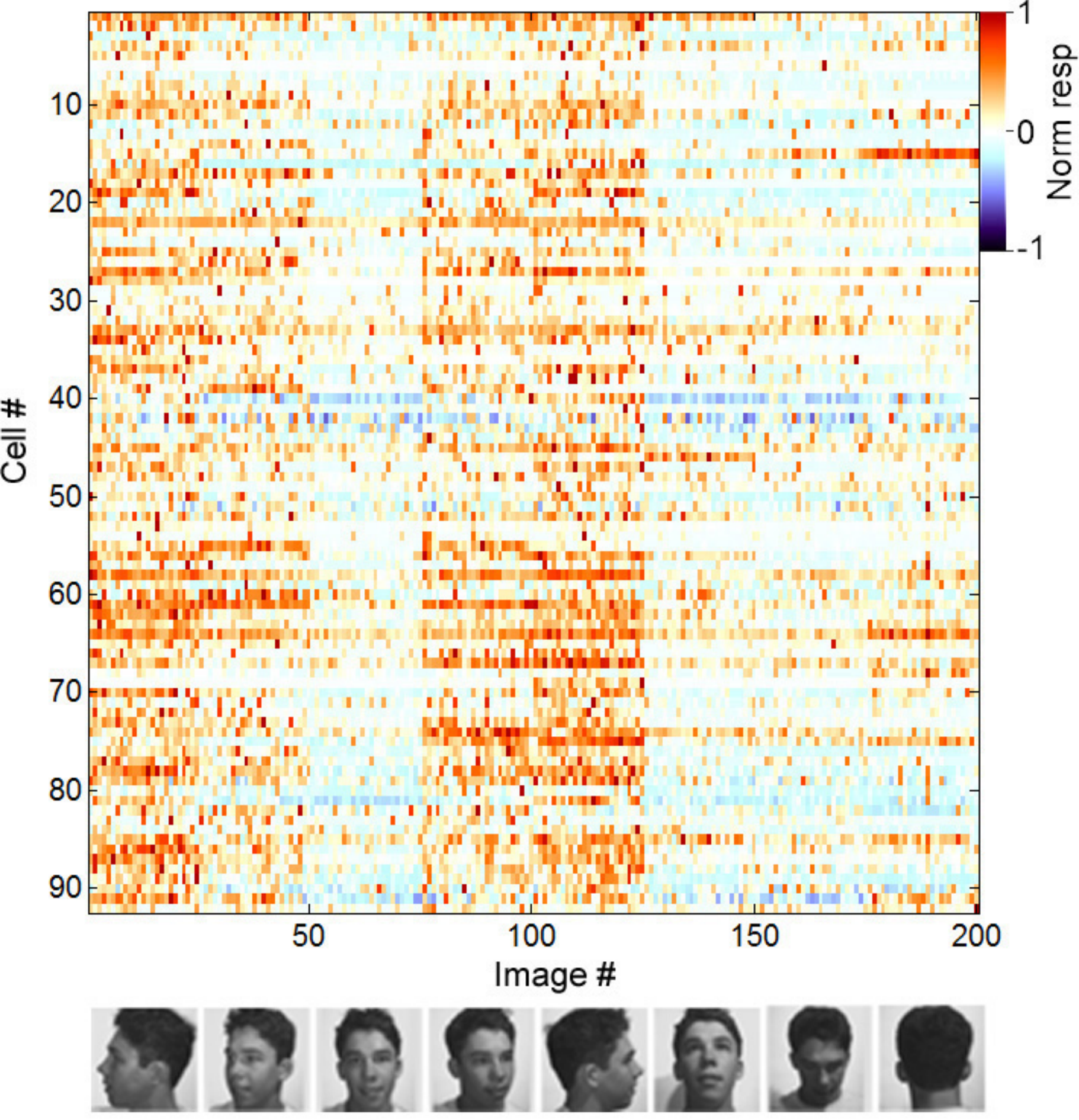


Figure S3

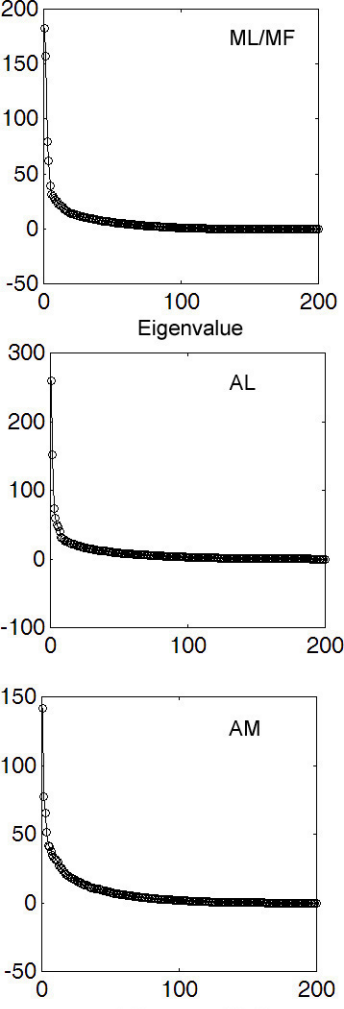
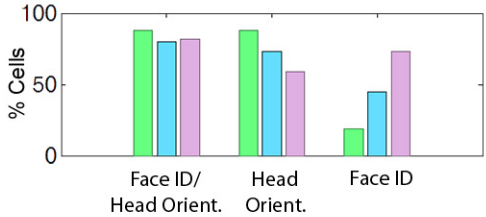

ML/MF
200
150
100
50
0
-50
0
100
200
Eigenvalue
AL
300
200
100
0
-100
0
100
200
AM
150
100
50
0
-50
0
100
200

Figure S4

A

Face Patch	ML/MF	AL	AM
Modulation by Face Identity or Direction ($p<0.005$)	88% (107/121)	80% (151/189)	82% (129/158)
Modulation by Face Direction ($p<0.005$)	88% (107/121) $F_{mean}=24.1$	73% (151/189) $F_{mean}=17.1$	59% (93/158) $F_{mean}=8.4$
Modulation by Face Identity ($p<0.005$)	19% (23/121) $F_{mean}=1.7$	45% (85/189) $F_{mean}=2.7$	73% (118/158) $F_{mean}=6.4$

*Responses to "back of head" not included in this analysis



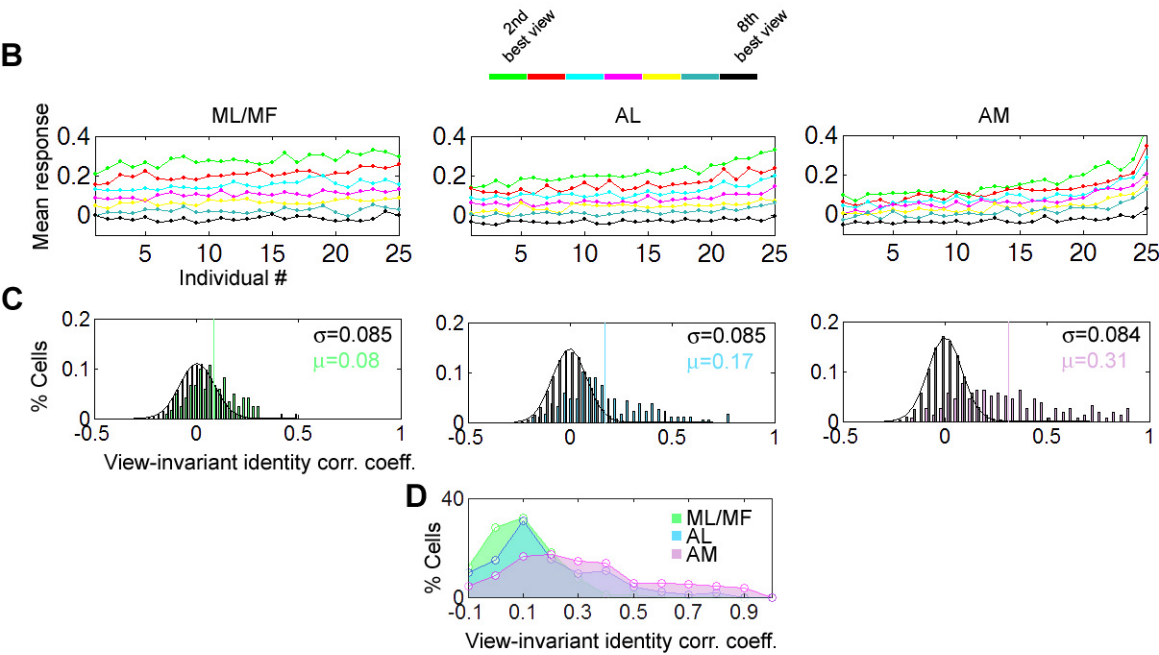


Figure S5

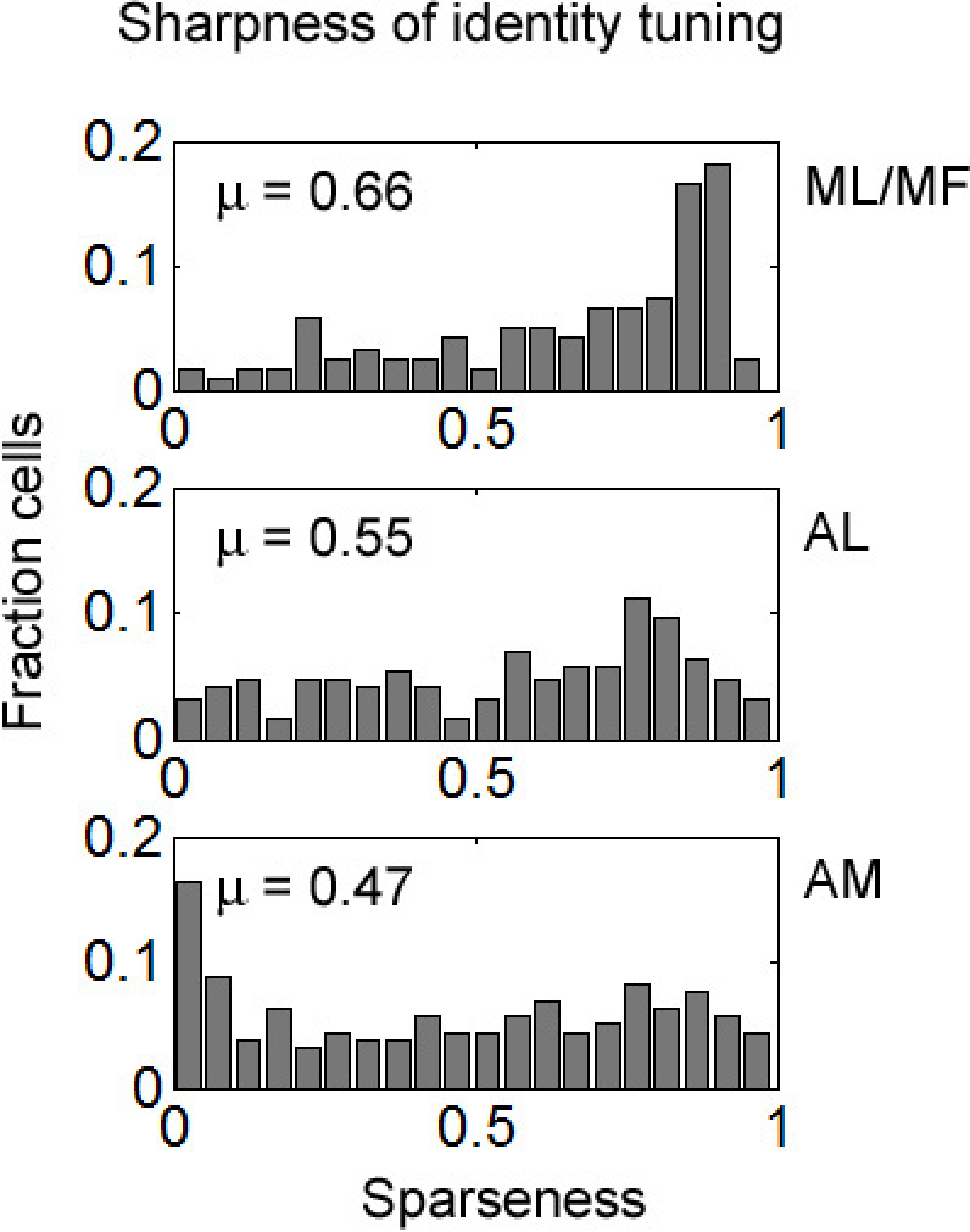

Sharpness of identity tuning
0.2
0.1
0
μ = 0.66
ML/MF
0
0.5
1
0.2
0.1
0
μ = 0.55
AL
0
0.5
1
0.2
0.1
0
μ = 0.47
AM
0
0.5
1
Fraction cells
Sparseness

Figure S6

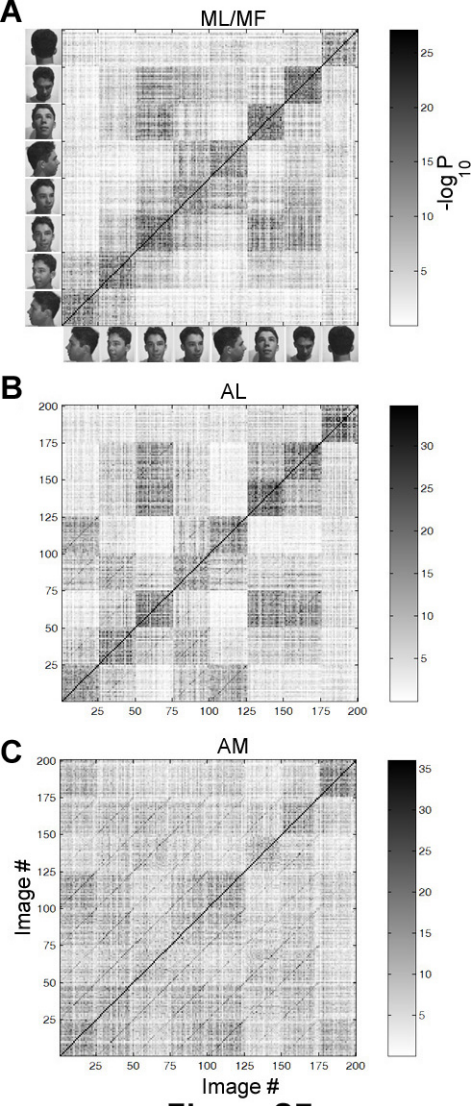


Figure S7

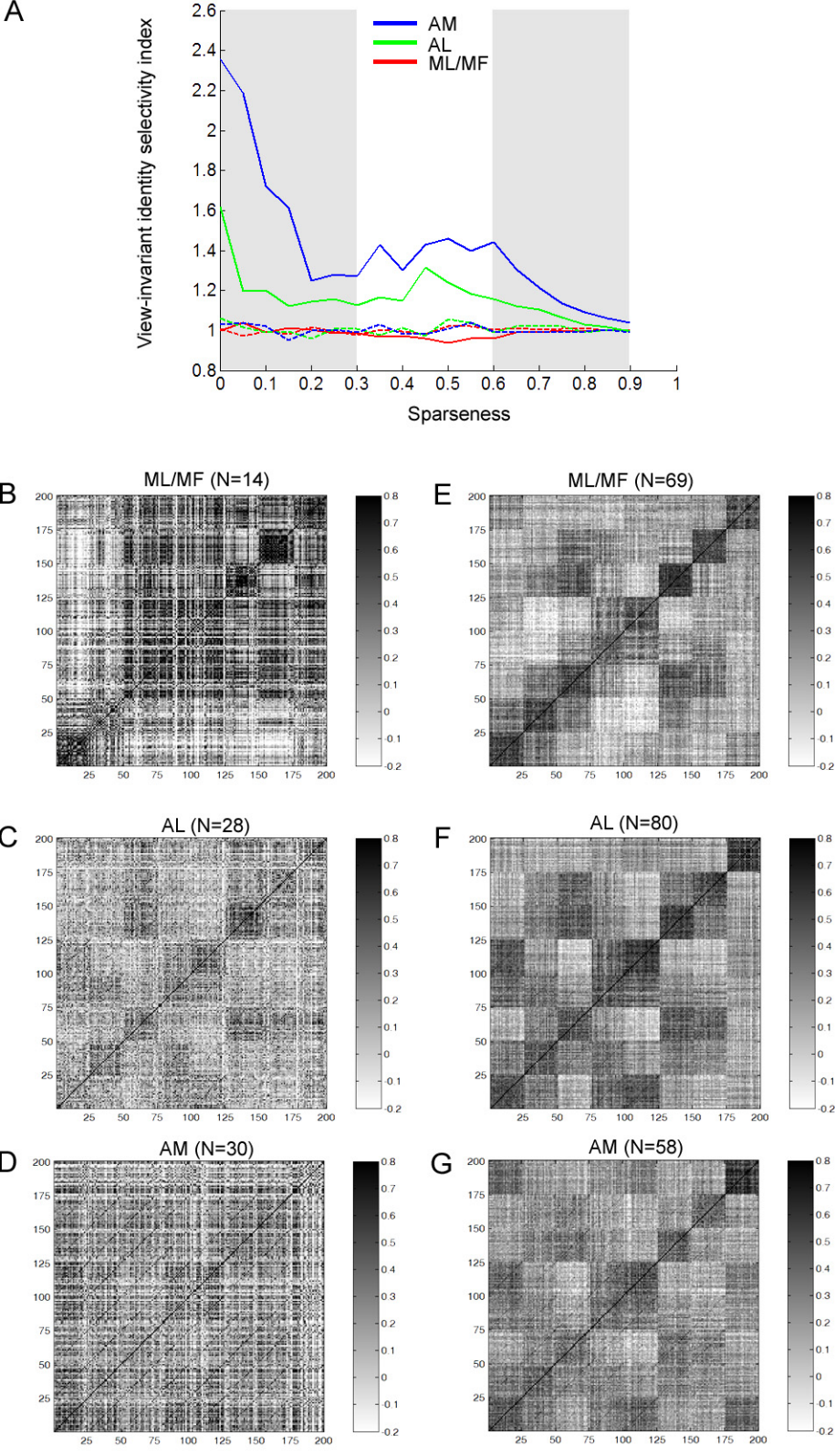

A
View-invariant identity selectivity index
Sparseness
AM
AL
ML/MF
B
ML/MF (N=14)
C
AL (N=28)
D
AM (N=30)
E
ML/MF (N=69)
F
AL (N=80)
G
AM (N=58)
0 < Sparseness < 0.3
0.6 < Sparseness < 0.9

Figure S8

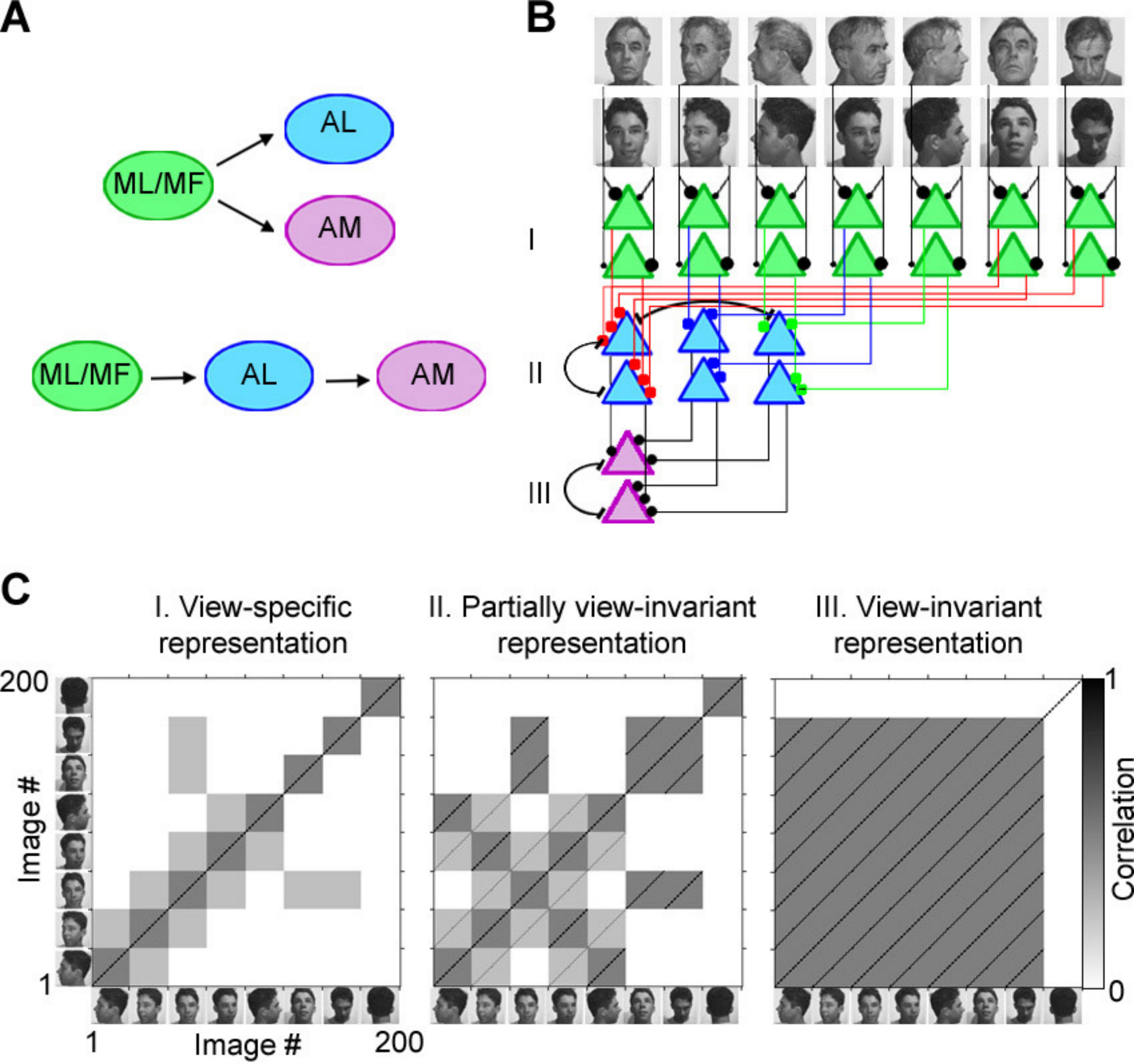

A
AL
ML/MF
AM
ML/MF
AL
AM
B
I
II
III
C
I. View-specific representation
II. Partially view-invariant representation
III. View-invariant representation
200
1
Image #
1
Image #
200
1
0
Correlation

Figure S9

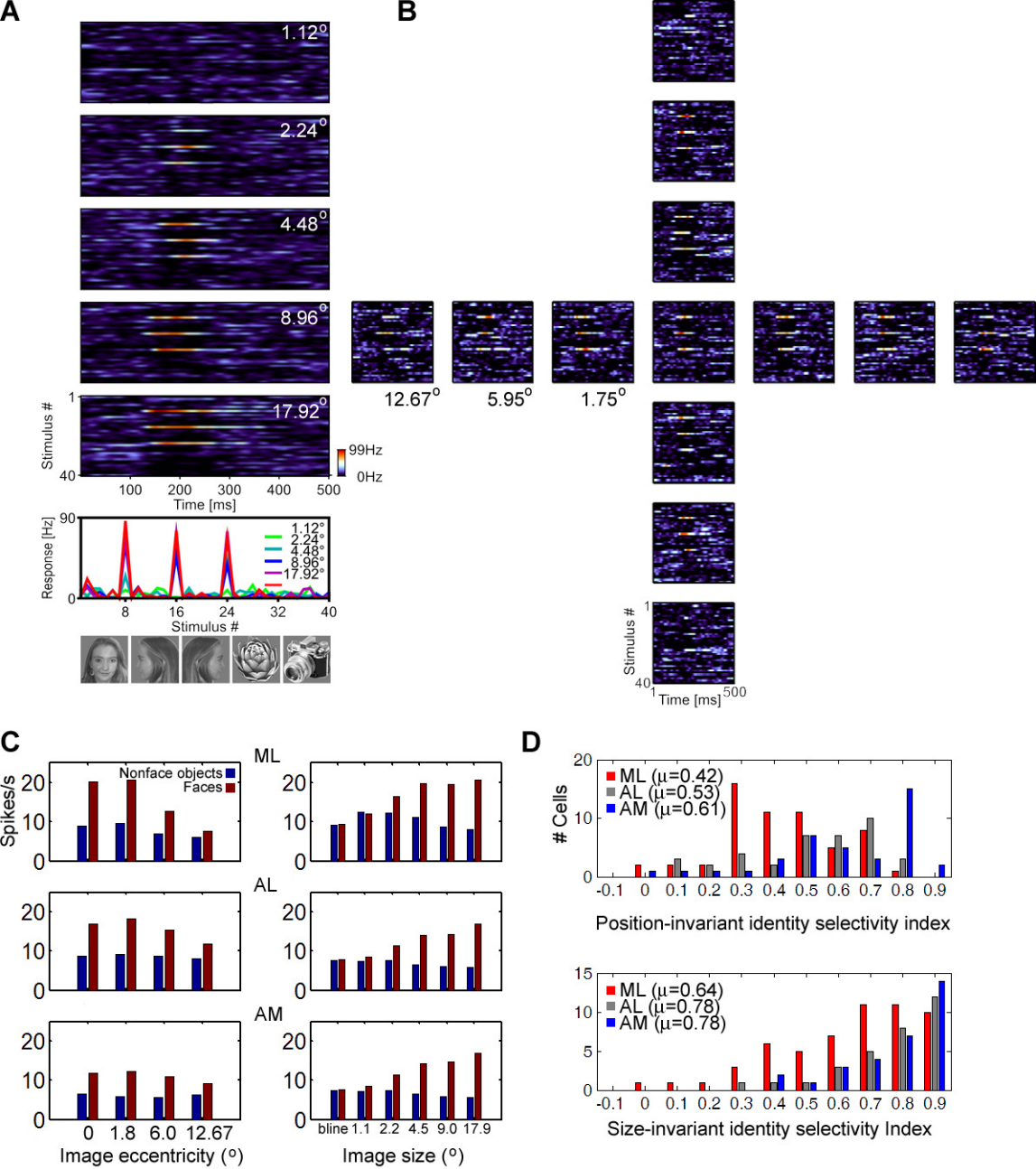

A
1.12°
2.24°
4.48°
8.96°
17.92°
Stimulus #
1
40
99Hz
0Hz
100
200
300
400
500
Time [ms]
Response [Hz]
90
0
1.12°
2.24°
4.48°
8.96°
17.92°
8
16
24
32
40
Stimulus #
B
12.67°
5.95°
1.75°
Stimulus #
1
40
1
Time [ms]
500
C
Spikes/s
Nonface objects
Faces
ML
AL
AM
20
10
0
0
1.8
6.0
12.67
Image eccentricity (°)
bline
1.1
2.2
4.5
9.0
17.9
Image size (°)
D
Cells
ML (μ=0.42)
AL (μ=0.53)
AM (μ=0.61)
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Position-invariant identity selectivity index
ML (μ=0.64)
AL (μ=0.78)
AM (μ=0.78)
15
10
5
Size-invariant identity selectivity Index

Figure S10

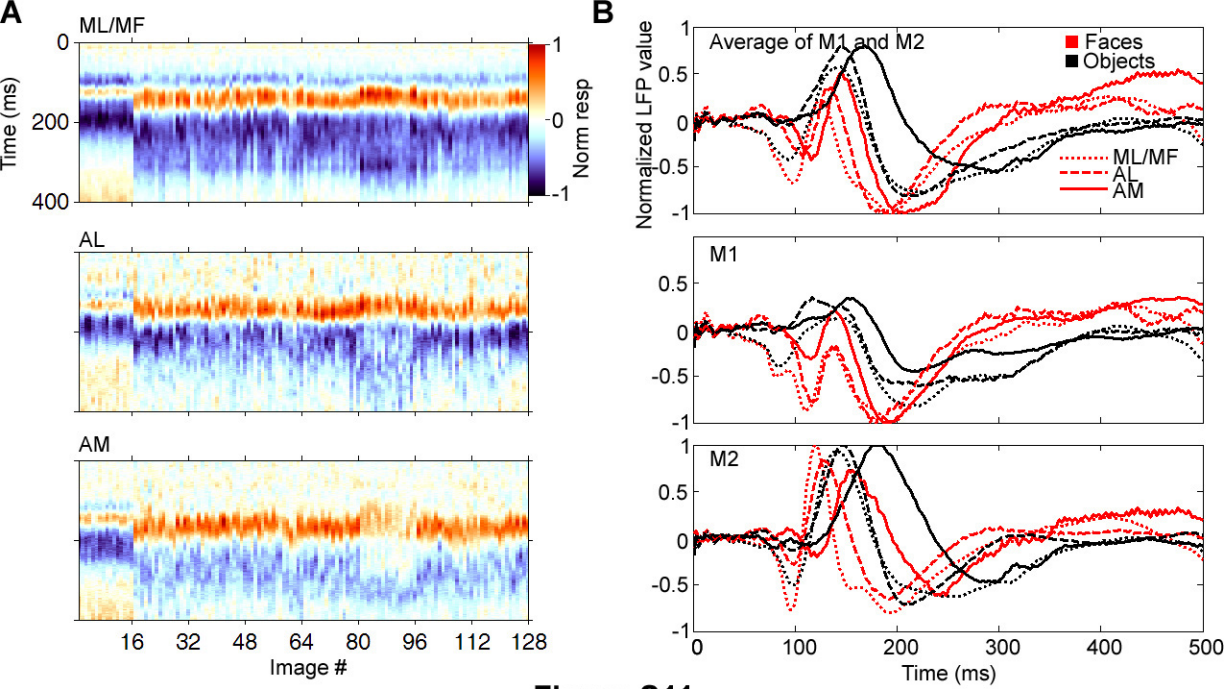

A
ML/MF
AL
AM
Time (ms)
0
200
400
1
0
-1
Norm resp
16
32
48
64
80
96
112
128
Image #
B
Average of M1 and M2
Faces
Objects
ML/MF
AL
AM
Normalized LFP value
M1
M2
0
100
200
300
400
500
Time (ms)

Figure S11

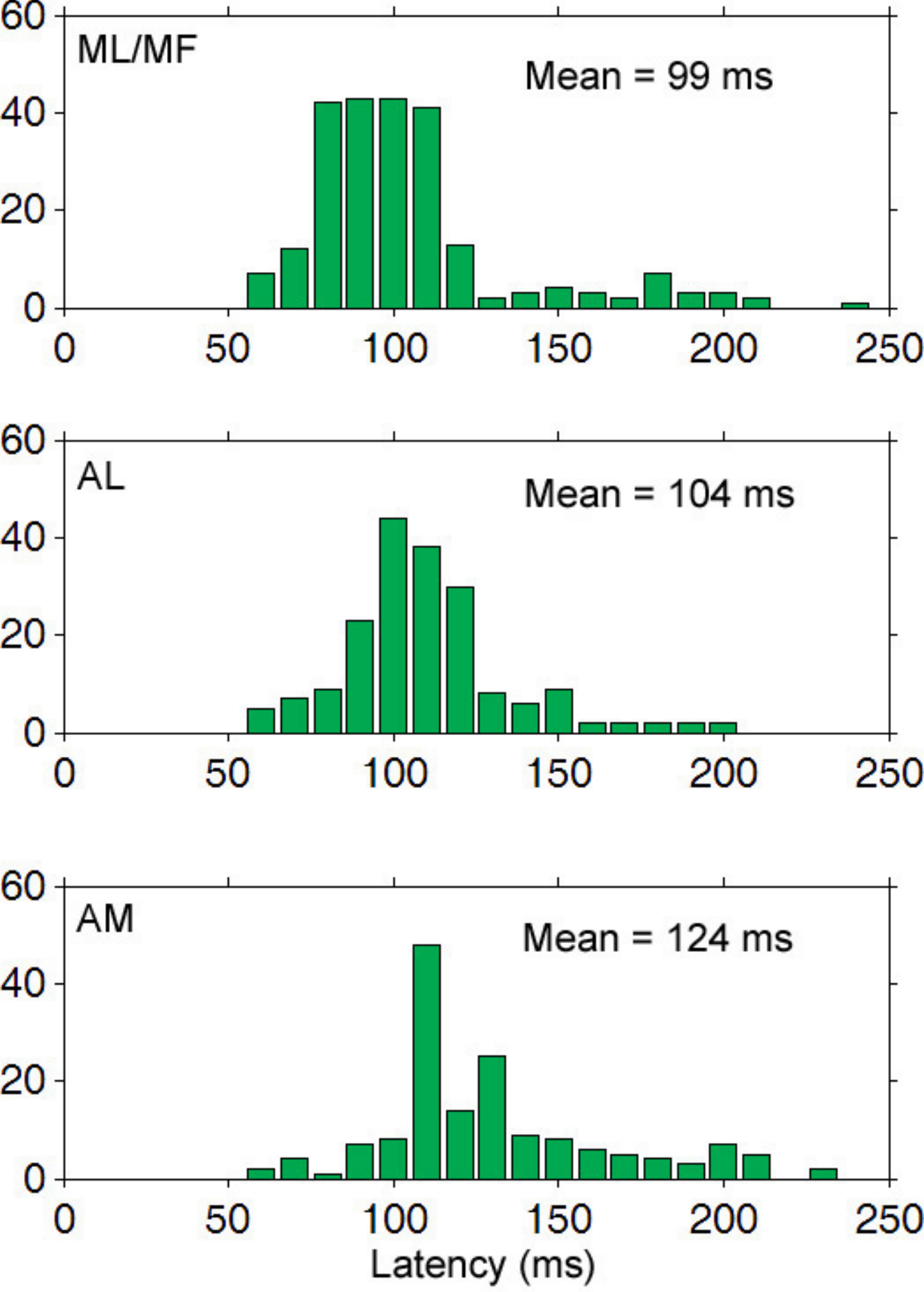

ML/MF
Mean = 99 ms
AL
Mean = 104 ms
AM
Mean = 124 ms
60
40
20
0
0
50
100
150
200
250
Latency (ms)

Figure S12

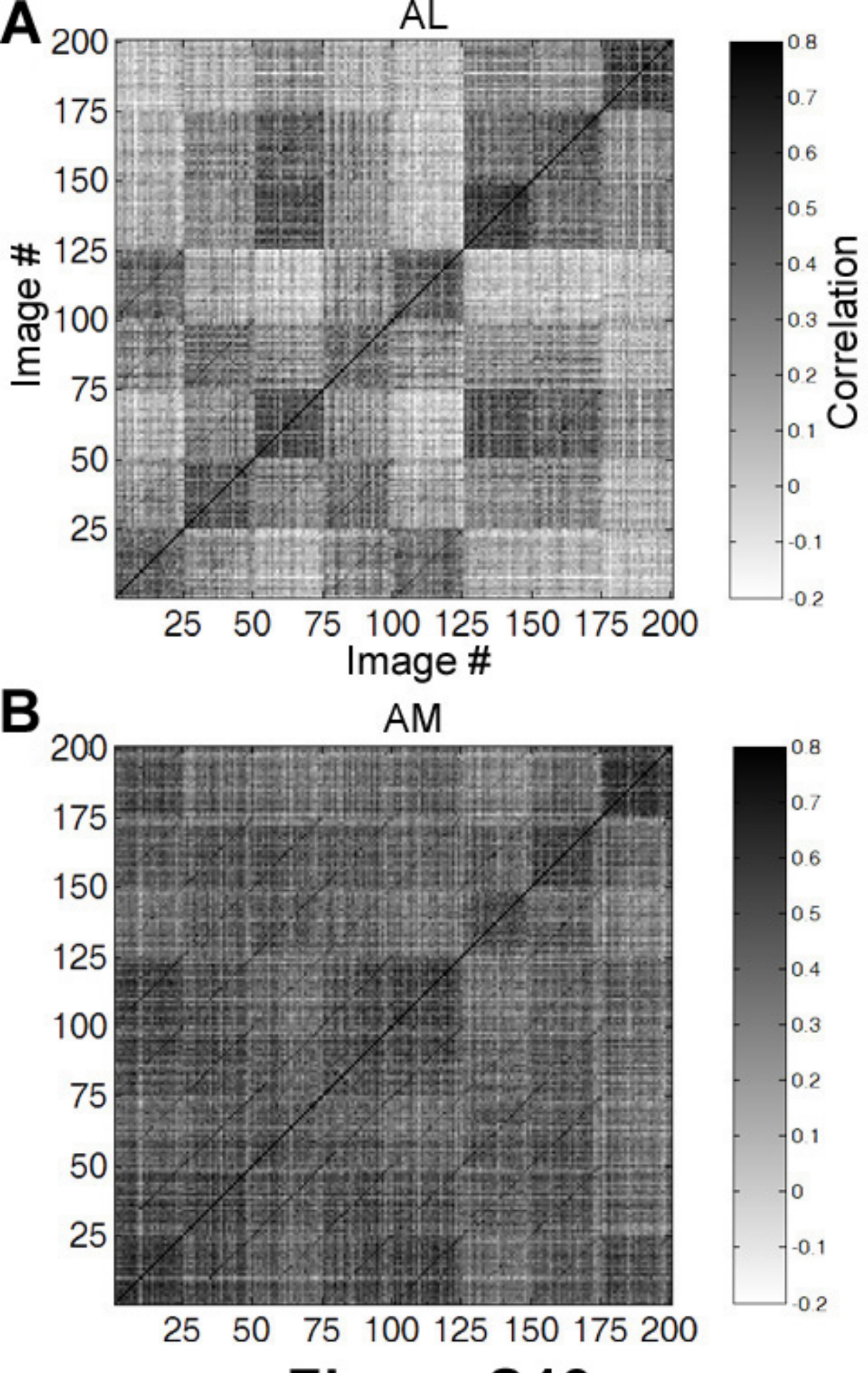
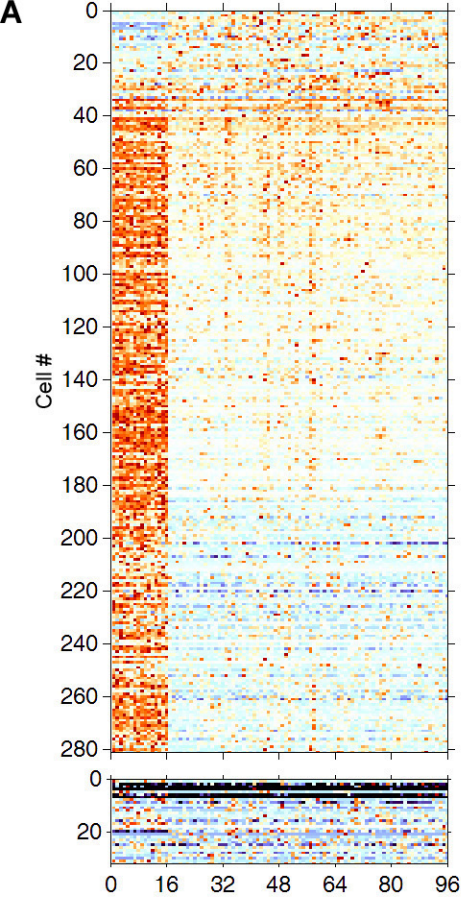

A
AL
Image #
200
175
150
125
100
75
50
25
25 50 75 100 125 150 175 200
Image #
Correlation
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
B
AM
200
175
150
125
100
75
50
25
25 50 75 100 125 150 175 200
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2

Figure S13


A
Cell #
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
0
20
0
16
32
48
64
80
96

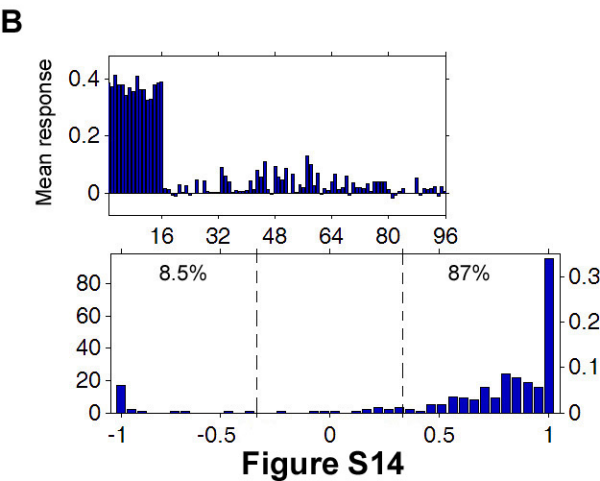

B
Mean response
0.4
0.2
0
16
32
48
64
80
96
8.5%
87%
80
60
40
20
0
0.3
0.2
0.1
0
-1
-0.5
0
0.5
1

Figure S14

Fig. S1. Targeting ML, AL, and AM in monkey M2. **(A)** Inflated right hemisphere showing six regions in the temporal lobe of monkey M2 that responded significantly more to faces than to objects. Color scale indicates negative common logarithm of P-value. **(B)** Coronal (left) and sagittal (right) anatomical MR images show the electrode descending into ML (top, located at AP +5mm), AL (middle, at +13mm), and AM (bottom, at +17 mm) in monkey M2. Co-registered face-selective functional activation is overlaid on the MR images.

Fig. S2. Response matrices of all visually responsive cells in ML/MF (**A**), AL (**B**), and AM (**C**), respectively (combined from monkeys M1 and M2) to the 128 FOB stimuli consisting of 16 faces, 16 human bodies, 16 fruits and vegetables, 16 gadgets, 16 human hands, 16 scrambled patterns, 16 monkey body parts, and 16 monkey whole bodies. Each row indicates one cell, and each column one image; cells sorted from top to bottom by their face selectivity index (FSI, see Methods). (Bottom) Average population response of visually responsive cells to each image.

Fig. S3. Responses of 92/215 AL cells that responded at least twice as strongly to one of the two full profiles as to frontal faces. Predominance of mirror symmetric responsiveness is evident.

Fig. S4. Eigenvalues corresponding to each MDS dimension in the three face patches.

Fig. S5. (A) (left) Table summarizing ANOVA analysis of influence of face head orientation and face identity in FV stimulus set on firing of visually responsive cells in ML/MF, AL, and AM. (right) Bar graph representation of the fractions of cells in each patch significantly modulated by face identity, head orientation, or both. **(B)** Normalized population response magnitudes to FV stimuli sorted by increasingly preferred individual (determined at preferred head orientation for each cell independently) and head orientation preference (color coded, from 2^{nd} best to 8^{th} best). An increasing trace indicates presence of view-invariant identity tuning. (**C**) Distributions of view-invariant identity correlation coefficients (see Methods) of cell populations in ML/MF, AL, and AM (colored histogram bars), means of these distributions (colored vertical line in each plot) together with corresponding shuffle control distributions (black bars and Gaussian fit). Shuffle controls, 100 per cell, were obtained by computing the view-invariant identity correlation coefficient from responses in which the 25 identities at each head orientation were randomly and

independently shuffled. The ratios between the mean view-invariant identity correlation coefficients and the standard deviations of the shuffle distributions were: 1.0 (ML/MF), 2.0 (AL), and 3.7 (AM). (**D**). Distributions of view-invariant identity correlation coefficients across the population in ML/MF, AL, and AM. The means of these distributions were significantly different from each other ($P << 0.001$ for all three comparisons, Mann-Whitney U Test), and from their respective shuffle distributions ($P << 0.001$ for all three comparisons, Mann-Whitney U Test).

Fig. S6. Sharpness of identity tuning in ML/MF, AL, and AM, quantified by the distributions of identity sparseness indices computed from responses of cells to the 25 identities in the FV set at the preferred head orientation. A lower sparseness index indicates sharper identity tuning. Central tendencies of distributions were significantly different from each other ($p < 0.01$ for comparison of sparseness distribution of AL vs. AM; $p << 0.001$ for all other comparisons, Mann-Whitney U Test).

Fig. S7. Significance matrices corresponding to population similarity matrices in ML/MF, AL, and AM. The matrices plot the significance ($-\log_{10} P$) of individual entries of the correlation coefficient matrices in Figs. 4D-F.

Fig. S8. Relationship between view-invariant identity selectivity and sparseness. **(A)** View-invariant identity selectivity index (see Methods) as a function of sparseness, plotted for AM, AL, and ML/MF (solid traces). A low index on the abscissa indicates a high degree of sparseness. Dotted traces show the mean view-invariant individual correlations computed from shuffled similarity matrices. **(B-D)** Similarity matrices computed using only sparse cells ($0 <$ sparseness < 0.3, light grey region on left in (A)). The number of cells used to compute each matrix is indicated at the top of each plot. **(E-H)** Similarity matrices computed using only non-sparse cells ($0.6 <$ sparseness < 0.9, light grey region on right in (A)).

Fig. S9. Hypothetical arrangements of face patches. **(A)** Two possible wiring schemes linking ML/MF, AL, and AM: parallel (top), and serial (bottom). **(B)** A serial model for establishing a view-invariant representation of faces. Pictures on top denote preferred stimuli, triangles

symbolize neurons (color conventions as in (A), circles indicate excitatory synapses, short lines inhibitory synapses, and lines connections between cells). **(C)** Predicted population similarity matrices for cells at each of the three levels in (B). The light gray squares result from weaker inputs from 45° adjacent views, not shown in the wiring diagram (B).

Fig. S10. Size and position invariance of response selectivity in ML, AL, and AM. **(A)** (Top five plots) Mean response time courses from an example AM cell to 40 faces and objects (8 frontal faces, 8 left full profile faces, 8 right full profile faces, and 16 non-face objects), each at five different sizes (stimulus size indicated in upper right). Response magnitude is color coded according to bottom right scale bar. (Middle) Mean responses to the 40 images, with each color indicating one size. (Bottom) Examples images illustrating each of the five image categories from the 40 image set. This cell responded to only 3 of the 40 images, corresponding to the face of one individual at straight, left profile, and right profile views – across more than an order of magnitude of image size. **(B)** Mean response time course from the same cell to the 40 images presented at 13 positions corresponding to four different eccentricities. The cell retained its selectivity for these 3 images out to 13° eccentricity. **(C)** Population response average to faces (red) and objects (blue) at four different eccentricities (left) and five different sizes (right). Spontaneous activity (baseline, "bline", indicated in right plot) not subtracted. While the response to objects remained approximately constant across eccentricities, the response to faces decreased at the two most peripheral eccentricities in ML/MF, but less so in AL and AM. With increasing image size, response magnitudes to faces increased in all three face patches, and responses to objects decreased. **(D)** Distributions of position (top) and size (bottom) invariant identity selectivity indices (see Methods) across visually-responsive cells in ML (N=58 cells for position, N=56 cells for size), AL (N=38 cells for position, N = 31 cells for size), and AM (N=39 cells for position, N = 31 cells for size). These distributions are significantly different from each other, with the exception of AM versus AL for size invariance (Position: $P << 0.001$ for ML/MF versus AL, $P < 0.003$ for AL versus AM; Size: $P < 0.002$ for ML/MF versus AL, not significant for AL versus AM; Mann-Whitney U Test).

Fig. S11. Evoked local field potential responses in three face patches. **(A)** Mean evoked local field potentials in ML/MF, AL, and AM, respectively, to the 128 FOB stimuli, averaged across

multiple recording sites in monkeys M1 and M2. **(B)** Traces of average LFPs in ML/MF, AL, and AM evoked by faces (red) and objects (black). (Top) LFPs averaged across monkeys M1 and M2. (Middle) LFP from monkey M1. (Bottom) LFP from monkey M2. Latencies to first positive peak were as follows: Combined (faces: ML/MF 126 ms, AL 133 ms, AM 145 ms, objects: ML/MF: 143 ms, AL 145 ms, AM 168 ms). M1 (faces: ML/MF 134 ms, AL 138 ms, AM 139 ms, objects: ML/MF: 145 ms, AL 115 ms, AM 156 ms). M2 (faces: ML/MF 120 ms, AL 130 ms, AM 156 ms, objects: ML/MF: 142 ms, AL 147 ms, AM 180 ms).

Fig. S12. Distributions of response latencies of all visually-responsive cells to the FOB image set within ML/MF (M1 and M2, N = 232), AL (N = 189), and AM (N = 158). These distributions are significantly different from each other ($P << 0.001$ for all three comparisons, Mann-Whitney U Test).

Fig. S13. Population similarity matrices for AL and AM computed using the first 121 cells (same as the number of cells used to compute the similarity matrix for ML/MF in Figs. 4E-F).

Fig. S14. Selectivity in ML/MF to pictures of faces and non-face objects across three animals. **(A)** (Top) Responses of visually-responsive cells in ML/MF (combined from monkeys M1, M2, and M3) to the first 96 FOB stimuli. (Bottom) Responses of visually-unresponsive cells. Conventions as in fig. **S2A**. **(B)** (Top) Mean response to each of the 96 images across the population of visually responsive cells. (Bottom) Distribution of face selectivity indices for visually responsive cells.

Supplementary Table 1. The number of visually-responsive and non-responsive cells and the number of LFP sites recorded from each monkey, for the FOB and FV stimulus sets. Note that for monkey M2, no FV stimuli were tested in ML.

Face Patch	**ML/MF**			**AL**			**AM**		
Stimulus	FOB (#cells)	FV (#cells)	LFP (#sites)	F0B (#cells)	FV (#cells)	LFP (#sites)	FOB (#cells)	FV (#cells)	LFP (#sites)
M1	71	71	43	31	31	13	116	116	68
# cells non-responsive to FOB	13			6			18		
M2	160	not tested	101	158	158	77	42	42	26
# cells not responsive to FOB	12			20			17		
M3	50	50							
# cells not responsive to FOB	7								
#visually responsive/ #non-visually responsive cells	**281/32**	**121/20**		**189/26**			**158/35**		

Supplementary Table 1. The number of visually-responsive and non-responsive cells and the number of LFP sites recorded from each monkey, for the FOB and FV stimulus sets. Recordings in M1 were taken from the left hemisphere, in M2 from the right hemisphere, and in M3 from the right hemisphere.