

Functional parcellation of visual cortex in humans and monkeys

Reza Rajimehr^{1*}, Simon Kornblith^{2,3}, Doris Y. Tsao⁴, Robert Desimone¹

1 McGovern Institute for Brain Research, Massachusetts Institute of Technology (MIT), Cambridge, MA

2 Picower Institute for Learning and Memory, Massachusetts Institute of Technology (MIT), Cambridge, MA

3 Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA

4 Department of Computation & Neural Systems, California Institute of Technology (Caltech), Pasadena, CA

*** Corresponding author:**

Reza Rajimehr

McGovern Institute for Brain Research

Massachusetts Institute of Technology (MIT)

43 Vassar St., Building 46, Room 5127

Cambridge, MA 02139

Phone: 617-324-5530

Fax: 617-324-6875

Email: rajimehr@mit.edu , rajimehr@gmail.com

Running title: Functional parcellation of visual cortex

Number of main figures: 8

Number of supplementary figures: 18

Number of tables: 1

Number of equations: 1

Abstract:

Characterizing the functional organization of visual cortex is a fundamental step in understanding how visual information is processed in the brain. Despite monumental efforts over the past decades to map visual cortical areas in primates, there has been no comprehensive parcellation scheme for the entire visual cortex. Here we used functional MRI (fMRI) in humans and macaque monkeys to map the occipito-temporal cortex based on responses to a 96-minute natural movie stimulus. Using a hierarchical clustering algorithm, we found a remarkable macro-organization of visual cortex at multiple spatial scales. In the human data, the clustering revealed category-selective areas, including a previously unknown region in the lateral temporal cortex that was selectively activated by images of human-object interactions. The existence of such a region was further confirmed with a block-design fMRI experiment. In the monkey data, the clustering revealed subdivisions of inferior temporal (IT) cortex, plus a previously uncharacterized 'place patch' in the anterior-medial temporal cortex that responded selectively to specific types of outdoor scenes in the movie. The cortical parcellation schemes presented here could guide subsequent studies in defining the functional properties of many new cortical areas.

Keywords: visual cortex, human fMRI, monkey fMRI, cortical map, parcellation, clustering

Introduction:

Visual cortical areas can be identified by four main criteria: histology (cytoarchitecture and myeloarchitecture), retinotopy, global functional properties, and connectivity patterns (Felleman and Van Essen, 1991). Based on these criteria, more than 30 visual areas have been identified in non-human primates, spanning about half of the cerebral cortex (Van Essen, 2004). In early visual cortex, the borders separating areas can be relatively unambiguously defined. For example, primary visual cortex (V1) contains a histological landmark (*stria of Gennari*) and a distinct topographic organization, which can be used to distinguish it from its neighboring area V2 (Tootell et al., 1988). Another example is area MT (middle temporal), whose borders with adjacent cortex can be defined either by an abrupt change in motion selectivity or a change in the pattern of anatomical connections to other cortical areas (Born and Bradley, 2005). At progressively higher-tier stages of visual cortical hierarchy (e.g. in the IT cortex, which is the final stage of the ventral visual pathway), the borders between areas become less clear, and

consequently, there has been multiple, non-conclusive parcellation schemes for IT cortex in non-human primates (Van Essen, 2004).

In humans, our knowledge about the spatial organization of high-level visual areas is further limited due to restrictions in doing invasive neurobiological studies. Using fMRI and other neuroimaging techniques, parts of human visual cortex have been mapped. These maps include retinotopic areas in the occipital cortex and category-selective areas in the temporal cortex (Rajimehr and Tootell, 2008). For example, in a series of fMRI localizer experiments, previous studies have reported that discrete patches in the occipito-temporal cortex respond categorically and selectively to images of faces, scenes/places, and body parts (see Grill-Spector and Weiner, 2014 for review). However, a large extent of temporal cortex in humans is still uncharted. Moreover, there has been no comprehensive map showing the spatial layout and borders of visual areas throughout the entire occipito-temporal cortex.

Here we used a totally data-driven approach to functionally parcellate the entire occipito-temporal cortex in humans and macaque monkeys. Since the anatomical and topographic borders are not well-defined in higher-tier regions of visual cortex, it is conceivable that visual areas in those regions are identified mainly based on functional properties. For functional parcellation, we first activated the visual cortex effectively using a natural movie stimulus that contained a rich collection of visual stimuli. Next, we used a two-step clustering analysis to group together the occipito-temporal vertices of the cortical surface based on similarity/commonality in the pattern of their responses to the movie. In this analysis, the vertices were first represented in a lower-dimensional activity space using principal component analysis (PCA), then hierarchical clustering was applied in the PCA space. By visualizing the functionally defined clusters on the cortical surface, we evaluated their large-scale spatial organization. In addition, the frames/segments of the movie that produced the highest and most consistent responses in certain clusters suggested possible functional specializations.

This data-driven approach could have several advantages over conventional localizer experiments for finding new cortical areas. The entire set of visual stimuli contains a vast number of possible stimulus categories. Designing localizer experiments to test hypotheses about cortical selectivity for all these categories would be inefficient and perhaps practically impossible. Furthermore, a cortical area may have counterintuitive selectivity to a combination of visual stimuli or to complex visual actions. Such selectivity can be easily overlooked when designing the localizer experiments.

Results:

Functional parcellation of human visual cortex

We first analyzed anatomical/structural data and computationally reconstructed the cortical surfaces in each human subject. The inflated surface in a representative subject was then processed to generate a 2-D flattened patch of the entire cerebral cortex (**Figure 1a**). This flat patch was used for the visualization of functional data. Using an automated algorithm for anatomical parcellation, the occipito-temporal vertices/points of the flat patch were selected (**Figure 1b**). Only these vertices were included in our functional parcellation since visual areas are typically located within the occipito-temporal cortex.

To collect functional data, we scanned the occipito-temporal cortex of human subjects (**Supplementary Figure 1**) while they were viewing a naturalistic movie (**Figure 1c**). The movie was *Baraka*, and the entire 96-minute movie was played silently to the subjects in 10 functional scans. The movie contained a variety of visual stimuli (e.g. pictures of people, indoor scenes, and outdoor scenes) and visual actions. Subjects viewed the movie with free eye movements. It has been shown that visual representations in high-level visual cortex are tolerant to eye movements when watching a natural movie ([Nishimoto et al., 2017](#)). Functional data in individual subjects were preprocessed and registered to a common anatomical space before they were averaged (see Methods). Thus, we obtained an averaged time-course of response to the movie, in each occipito-temporal vertex. Since the movie was presented only once to the subjects, the inter-subject averaging of functional data provided more reliable activation patterns.

Next, we constructed an activity space in which each axis corresponded to the functional activity at a given time-point (**Figure 1d**). Our data acquisition included 2810 time-points for the entire movie based on the sampling rate of every two seconds (TR = 2 sec); thus, the activity space contained 2810 orthogonal axes. Vertices were data-points in this space (~ 40,000 occipito-temporal vertices in a given hemisphere). Our primary goal was to find distinct clusters of vertices based on the geometric distance between data-points in the activity space. However, the clustering algorithms may fail in high-dimensional spaces because the data-points get too sparse (i.e. the distribution of distances becomes flat) ([Steinbach et al., 2004](#)). To achieve an optimal clustering, we used PCA to reduce the dimensionality of the original activity space. **Figure 1e** shows the projected/transformed vertices in a new (low-

dimensional) space constructed based on the first three principal components. We performed qualitative and quantitative analyses on this space before proceeding with the clustering analysis.

To qualitatively evaluate the PCA space, we assigned a color (an RGB value) to each vertex based on its position in the 3-D PCA space, then visualized the colored vertices on the flat patch. **Figure 1f** shows the PCA map of occipito-temporal vertices in the right hemisphere. The map revealed a large-scale spatial organization of the PCA values on the cortical surface, despite the fact that no information about the spatial location of vertices had been included in PCA. This result suggests that vertices within a certain patch of cortex could have a distinct functional property, compared to vertices in the other parts of cortex; this spatial variation in function would be demonstrated more explicitly in the clustering analysis. These patches often spanned a large extent (few centimeters) of the cortical sheet, ruling out the possibility that they were artifactual structures (random PCA values that were spatially blurred in fMRI).

The PCA map clearly showed two major gradients of cortical representation: 1) a dorsal-ventral gradient in the occipital cortex, presumably reflecting the representation of visual field eccentricity, and 2) a posterior-anterior gradient, which may be related to the transition from retinotopic to category selectivity along the occipito-temporal axis. Although the PCA space was essentially a continuous space, parts of its cortical map showed abrupt changes in the PCA values (e.g. transition from green to blue at/near the border of fusiform cortex with inferior temporal cortex).

To test the reproducibility of the PCA map, we conducted PCA on vertices in the left hemisphere. As shown in **Supplementary Figure 2a,b**, the PCA map in the left hemisphere was virtually identical to the PCA map in the right hemisphere. As a control test, the time-course of response in each vertex was circularly shifted by a random amount, and the shuffled time-courses were used in PCA. As expected, we obtained noisy maps in this analysis (**Supplementary Figure 2c,d**). Thus, the patchy organization in the PCA map disappears when the time-courses of adjacent vertices are decorrelated.

The first principal component should be, in principle, related to the most fundamental visual feature in the movie. We hypothesized that the first principal component reflects the luminance variation over time. Temporal variation in luminance, which would be ‘stimulus on and off’ in an extreme case, is perhaps the most important factor for driving the visual cortex. To quantitatively test this hypothesis, we first calculated all the pixel-level luminance variations from one frame to the next frame of the

movie. These values were used to compute the total luminance variation (luminance variation index, LVI) within a two-second segment of the movie (**Figure 2a**). Thus, for every two-second segment of the movie (corresponding to our TR in fMRI), we had LVI and a weight/loading value for every principal component. The weights were eigenvectors of PCA. Next, we looked at the correlation of LVI with the weights of the first three principal components (**Figure 2b**). To have a robust estimate for correlation, adjacent data-points in the scatter plots were averaged. LVI showed a strong positive correlation with the weights of the first principal component, no correlation with the weights of the second principal component, and a negative correlation with the weights of the third principal component (**Figure 2b**; stats are shown in the figure). This finding is consistent with previous studies that have found a correlation between the first principal component in fMRI data and ‘motion energy’ in movie stimuli ([Huth et al., 2012](#)). **Figure 2c** shows the map of the first principal component on the flat patch. This map suggests greater sensitivity for luminance variation in posterior/occipital cortex compared to anterior regions in lateral temporal cortex. A control analysis confirmed that this map (and other PCA maps) could not be simply explained by variation of MR signal intensity across the cortical surface (**Supplementary Figure 3**). In the SNR (signal-to-noise ratio) map, the spatial variation was predominantly along the dorsoventral axis (ventral regions had generally lower SNR, compared to dorsal regions) (**Supplementary Figure 3**, [Ojemann et al., 1997](#)).

For the purpose of clustering, we used the first ten principal components that captured most of the variance in the fMRI data (**Figure 3a**). The value of ten was also greater than the number of shared principal components across subjects; this number is typically around five in fMRI datasets ([Huth et al., 2012](#)). Thus, the clustering was applied on vertices in a 10-dimensional PCA space. A ‘scree analysis’ confirmed that additional (> 10) principal components did not substantially contribute to the explained variance (**Supplementary Figure 4a**). In this analysis, the PCA eigenvalues, which were directly related to the values of explained variance, were plotted as a function of the principal component number, then the second derivative of the plot was obtained. For principal components greater than ten, the second derivative was close to zero, meaning that those principal components contributed minimally and linearly to the explained variance.

For the clustering analysis, a hierarchical clustering algorithm was used. Unlike other clustering algorithms (such as *k*-means clustering) in which the number of clusters is fixed and arbitrarily predefined, the hierarchical clustering would group data-points at various levels/scales. Such multi-scale

approach can be particularly useful for testing hierarchical ('coarse-to-fine') partitioning of the spatially organized maps. **Figure 3b** shows top levels (levels = 2-25) of the hierarchical clustering tree that was obtained by clustering vertices in a 10-dimensional PCA space. A specific color was assigned to vertices within each branch of the tree structure, then the colored vertices were visualized on the flat patch (**Figure 3c**). The maps in **Figure 3c** demonstrate a remarkable spatial organization of functionally defined clusters on the cortical surface. Note that we did not include any information about the location of vertices in the analysis; nonetheless, the clusters were spatially contiguous. At the level of the top two clusters, the entire occipito-temporal cortex was divided into two large patches (red and blue patches). The red patch was spatially localized at the foveal/parafoveal regions of the occipital cortex. The map with two clusters could be considered as a map of visual cortex at a very coarse scale. By progressively increasing the number of clusters, the patches were recursively subdivided into smaller patches, and the resulting maps showed the macro-organization of visual cortex at finer scales. The map with 25 clusters ('map25') could be considered as a tentative parcellation scheme for the occipito-temporal cortex in humans; such a scheme could constrain the possible layout of visual areas. The map25 included patches in early visual cortex that were arranged along the dorsoventral axis. These posterior patches may well correspond to the representation of different eccentricity bands (see below). More anteriorly, the map25 included another set of distinct patches. Some of the anterior patches may actually be the category-selective areas (see below). As shown in **Supplementary Figure 4b**, the clusters became very small in size by continuing the hierarchical clustering procedure. We chose the cutoff point of 25 clusters to visualize the most robust clusters that could be attributed to known visual areas.

Figure 4a shows selected clusters from the map25 that are located in early visual cortex. These clusters did not appear to correspond to known early visual areas (e.g. V1, V2, V3), but rather seemed to be coarsely related to the representation of visual field eccentricity or possibly a combination of areal boundaries and eccentricity bands. Such relationship was supported by a visual comparison of these clusters and an averaged eccentricity map of 53 subjects (**Figure 4b**). Thus, the functionally defined clusters in early visual cortex were indeed the visual field map clusters (Wandell et al., 2007). In the foveal/parafoveal cortex, which has a magnified representation (Tootell et al., 1982), more than one cluster was found in a given eccentricity band. In particular, the most anterior part of foveal/parafoveal cortex ('extended foveal cortex' – XFC) matched topographically to a cluster that could possibly be the putative human homologue of macaque TEO (Kolster et al., 2010).

In **Figure 4c**, we have highlighted some of the map25 clusters (yellow, cyan, and magenta patches) that seem to correspond to the category-selective areas including face-selective areas (FFA, fusiform face area; pSTS, posterior superior temporal sulcus) (Kanwisher et al., 1997; Pitcher et al., 2011), scene-selective areas (PPA, parahippocampal place area; OPA, occipital place area) (Epstein and Kanwisher, 1998; Dilks et al., 2013), and body-selective areas (EBA, extrastriate body area; FBA, fusiform body area) (Downing et al., 2001; Schwarzlose et al., 2005). To make such correspondences, we considered the anatomical/topographic location of these clusters relative to cortical gyri/sulci, relative to retinotopic areas, and relative to each other. For example, the FFA cluster was located lateral to mid-fusiform sulcus and ~ 1 cm anterior to the retinotopic area hv4, and it was bordered by the PPA cluster medially and the FBA cluster laterally (**Figure 4c**). All these relationships are consistent with where the classically defined FFA has been shown to lie with respect to its neighboring areas (Halgren et al., 1999). A region-of-interest analysis confirmed that the putative face, scene, and body patches in the map25 were functionally more active/responsive for frames of the movie that included faces, scenes, and bodies, respectively (**Supplementary Figure 5a-c**). For instance, the majority of preferred movie frames for the body patches included pictures of body parts, hands, and even internal body organs like a skeleton (**Supplementary Figure 5c**). For each stimulus category, we found a pair of patches separated by a relatively large distance on the cortical surface. Therefore, these patches have not resulted from artifactual correlations between the fMRI hemodynamic responses of neighboring voxels (a.k.a. spatial autocorrelations (Kriegeskorte et al., 2008)).

Tracing of clusters in the hierarchical clustering tree showed that the face and body clusters had a common node at one level up in the hierarchy (**Figure 3b**). This suggests a link in the representations of face and body categories that are physically and semantically related to each other. Interestingly, the scene clusters were derived from a different node in the tree structure, and they were part of a larger cluster associated with the peripheral representations in the occipito-temporal cortex (**Figure 3b**). Such link in the representations of scenes and visual periphery is consistent with the current models for the origin of scene selectivity in visual cortex (Levy et al., 2001; Hasson et al., 2003).

For FFA, PPA, and EBA, which are considered the most robust and perhaps the most replicable category-selective areas in humans, we compared the location of FFA, PPA and EBA clusters with FFA, PPA and EBA labels/areas obtained through conventional localizer experiments (i.e. by comparing the blocks of one object category versus other object categories). The FFA cluster almost perfectly matched the

classically localized FFA – cFFA (**Figure 4c**). The PPA cluster had an overlap with the posterior-lateral part of the classically localized PPA – cPPA (**Figure 4c**). The anterior-medial part of the classically localized PPA had an overlap with another cluster which also showed a preference for scene frames in the movie (**Supplementary Figure 6a**) – though we could not find a systematic/parametric difference between the preferred scene images in the anterior PPA cluster and the preferred scene images in the other scene clusters. Recent fMRI studies have also reported two (posterior and anterior) subdivisions within PPA, with the posterior subdivision having functional connectivity with OPA located further posteriorly and dorsally in visual cortex ([Baldassano et al., 2013](#); [Nasr et al., 2013](#)). The EBA cluster also corresponded topographically to the classically localized EBA (**Supplementary Figure 6b,c**). Interestingly, the subparts of this cluster matched the ‘hot spots’ of body-selective activity in the EBA localizer map (**Supplementary Figure 6b,c**). As also shown previously ([Weiner and Grill-Spector, 2011](#)), these subparts (named here as EBA1, EBA2 and EBA3) formed a crescent-shaped region surrounding area MT/V5.

Our data-driven approach revealed clusters with properties and locations that were similar to those of the previously reported category-selective areas. Thus, these clusters were not just a geographical feature of the parcellation map, but rather they appeared to be genuine cortical areas, each having a specific functional property. One obvious question is whether this approach could reveal clusters with a novel functional property. To address this question, we chose a cluster in a relatively uncharted region of temporal cortex, and performed qualitative and quantitative analyses to characterize its function. This cluster (the green patch in the map25) was named LT – lateral temporal, since it was anatomically located in the lateral temporal cortex (more specifically, at the caudal end of middle temporal cortex) (**Figure 4c**). The LT cluster was located immediately anterior and lateral to the EBA/FBA cluster. The proximity of the LT cluster to the body and face clusters could be an indication that this cluster might respond selectively to a specific class of stimuli. **Figure 4d** shows example frames from six segments of the movie that produced the highest response in the LT cluster. In some (but not all) of these movie segments, one could see/perceive a specific human action involving a tool use (e.g. applying face makeup). Thus, one possible scenario for the function of LT cluster would entail the processing of human-object interactions. In a quantitative test, we computed the fMRI response in the LT cluster for different movie segments that were sorted based on an averaged motion energy within those segments. As shown in **Figure 4e**, there was a strong negative correlation between the response in the LT cluster and motion energy in the movie (stats are shown in the figure). Thus, in contrast to motion-selective areas (such as area MT), the LT cluster showed a counterintuitive response bias for static and slowly

moving visual stimuli. Taken together, the LT cluster may respond selectively to static images of human-object interactions.

To further explore the selectivity in this region, we scanned the same human subjects in a more controlled, block-design experiment in which two types of stimulus blocks were used: 1) dynamic stimuli: short (5-second) movie clips of human-object interactions from 20 action categories, 2) static stimuli: snapshots of an individual frame from those movie clips (**Figure 5a**). In a direct comparison between dynamic and static stimuli, dynamic stimuli activated a broad swath of cortex including MT and other motion-responsive regions in occipital and posterior superior temporal cortex (**Figure 5b**, [Bartels et al., 2008](#)). Conversely, static stimuli selectively activated a highly localized cortical region, exactly where the LT cluster was located (**Figure 5b**). The magnitude and topography of these activations varied slightly across subjects – though the activation produced by static stimuli was always located within the LT cluster (**Figure 5c**). A vertex-wise analysis confirmed a ‘dynamic bias’ in MT and a ‘static bias’ in LT, in almost all the vertices within these two regions (**Figure 5d,e**). It is important to note that in this analysis, both MT and LT labels/regions-of-interest were obtained from an independent dataset (MT from a motion localizer map, LT from the map25).

In the process of taking snapshots from videos, some images became inevitably blurry due to motion in the video. Does the activity in the LT cluster reflect the ‘motion blur’ *per se*? To investigate this possibility, we applied spectral (Fourier) transform on each static image to calculate its normalized power for low spatial frequencies (< 1 cycle/degree), then classified it as either blurry image (containing more power at low spatial frequencies) or non-blurry image (containing less power at low spatial frequencies). **Supplementary Figure 7a,b** show examples of blurry and non-blurry images in our dataset. In a direct comparison between blurry and non-blurry images, blurry images did not activate the LT cluster (**Supplementary Figure 7c**). However, they activated a cortical territory outside of the LT cluster. This territory, which was located immediately posterior and ventral to the LT cluster, may provide low-frequency signals to the LT cluster. In pilot fMRI experiments with independent sets of stimuli, we also tested the effects of other visual factors (such as 2-D motion of simple geometrical shapes vs. formless random-dot motion, moving objects on a static scene background vs. static objects on a moving scene background) on the activity of the LT cluster. None of these factors activated the LT cluster (data not shown).

In addition to testing the functional specificity of some of the clusters, it is important to assess the reliability of the parcellation maps. For this, we did several cross-validation analyses by comparing the map25 in right vs. left hemisphere (**Supplementary Figure 8a,b**), in odd-labeled vs. even-labeled subjects (**Supplementary Figure 8c,d**), and in odd vs. even functional runs (**Supplementary Figure 8e,f**). In general, there was a qualitative agreement between the maps – though some notable differences could be also observed. For example, the pSTS face cluster was robustly found only in the right hemisphere (**Supplementary Figure 8a**), consistent with the hemispheric lateralization reported for this face-selective area (Pitcher et al., 2011). Using a quantitative metric (Fowlkes–Mallows index, Fowlkes and Mallows, 1983), we computed the similarity between hierarchical clusterings in two groups of subjects and in two groups of functional runs (**Supplementary Figure 9**). In both cases, the clustering similarity gradually decreased as the number of clusters increased, however, the similarity was always significantly higher than analytic chance where the two clusterings were assumed to be completely unrelated (for every level of clustering: $q \ll 0.05$, adjusted for multiple comparisons (100 t-tests)). Such reproducibility in clustering was more pronounced when the number of clusters was less than 25, thereby confirming that our cutoff point for the hierarchical clustering was reasonable. The clustering similarity was also computed for pairs of simulated data in which the real functional activities were replaced by Gaussian white noise (**Supplementary Figure 9**). Again, the clustering similarity was much higher for the real data, compared to the simulated/noise data (for every level of clustering: $q \ll 0.05$, adjusted for multiple comparisons (100 permutation tests)). The high degree of similarity between clusterings in odd and even functional runs suggested that the two, independent parts of the movie elicited highly similar patterns of clustering. Thus, our clustering was somewhat insensitive/invariant to the contents of the movie.

Finally, we evaluated the parcellation map at the level of 80 clusters. At this level, the clusters were small in size (**Supplementary Figure 4b**), yet they were spatially contiguous on the cortical surface (**Supplementary Figure 10a**). One of the clusters (the black patch in the map80; a subdivision of the large blue patch in the map25) was located within aSTS (anterior superior temporal sulcus), in a region which appeared to have a distinct representation in the PCA map (**Supplementary Figure 10a**). Thus, this cluster and its surrounding cortical tissue may be characterized by a specific functional property. A region-of-interest analysis revealed that the aSTS cluster was in fact a face-selective area. For this cluster, the preferred movie frames were mostly the images of faces/people (**Supplementary Figure 10b**), and the non-preferred movie frames consisted of a heterogeneous set of images including scenes

and body parts (**Supplementary Figure 10c**). Face selectivity in human aSTS has been reported in recent fMRI studies as well (Pinsk et al., 2009; Pitcher et al., 2011). The face cluster in human aSTS is presumably the homologue of the face patch AF – anterior fundus (Tsao et al., 2008) in macaque aSTS. The map of 80 clusters also showed a cluster that spatially corresponded to the motion-selective area MT (**Supplementary Figure 11**).

Functional parcellation of monkey visual cortex

The same data-driven approach was used for parcellating visual cortex in monkeys. Functional data were collected by scanning two macaque monkeys while they were viewing the same movie presented to the human subjects. In the monkey scans, we used a contrast agent (MION) and a gradient insert (AC88) to increase the functional sensitivity (see Methods). Functional data were preprocessed and registered to an atlas brain (macaque F99). To optimize functional-anatomical registration, the functional volumes were spatially undistorted before registration (**Supplementary Figure 12**). Data from four hemispheres were averaged in order to have a better estimate of the response profile in the vertices.

Figure 6a shows a flat patch of F99's right hemisphere. Using a lobar parcellation, the occipito-temporal vertices of the flat patch were selected. For these vertices, the PCA map was obtained by applying PCA to the averaged functional data (**Figure 6b**). Similar to the human PCA map, the monkey PCA map demonstrated a large scale organization across the entire occipito-temporal cortex. **Supplementary Figure 13a** shows the map of the first principal component. In this map, the values in the most medial and anterior regions of the occipito-temporal cortex were substantially distinct from those in the other regions. Such distinction might be related to differential sensitivity for a low-level visual feature (such as luminance and/or contrast), which is normally reflected in the first principal component. A control analysis confirmed that the PCA maps could not be explained by variation of MR signal intensity across the cortical surface (**Supplementary Figure 13b**).

For the clustering analysis, the vertices were projected into a 10-dimensional PCA space, then they were grouped together using a hierarchical clustering algorithm. **Figure 6c** shows the functionally defined clusters on the flat patch. At the level of two clusters, one cluster was localized at the foveal/parafoveal regions of the occipital cortex, similar to what we found in the human data. At the progressively later stages of hierarchical clustering, the newly formed clusters were spatially contiguous and systematically

arranged on the cortical surface even at the level of 25 clusters. The macaque map with 25 clusters ('map25m') could be considered as a tentative parcellation scheme for the occipito-temporal cortex in macaques.

As mentioned in the Introduction, the organization of visual cortical areas in macaque IT cortex has not been well understood. Thus, we were particularly interested to know whether our functional parcellation could provide insights about such organization. **Figure 7a** shows a collection of clusters from map25m that are located within the IT cortex. The posterior-anterior arrangement of these clusters resembled the arrangement of areas in the Felleman and Van Essen scheme (**Figure 7b**, Felleman and Van Essen, 1991). Six clusters, occupying the lower bank of STS and the adjacent lateral convexity of IT cortex, could be analogous to areas PITd, PITv, CITd, CITv, AITd, and AITv. An additional long cluster was located in the upper bank of STS, which may correspond to the 'superior temporal polysensory' (STP) area (Desimone and Gross, 1979). More posteriorly, there were clusters that appeared to be in the location of 'ventral occipitotemporal' (VOT) area. Area TEO (as described in the Ungerleider and Desimone scheme, Ungerleider and Desimone, 1986; Boussaoud et al., 1991) roughly corresponds to VOT+PITv (Van Essen, 2004). These areas are retinotopically organized (Kolster et al., 2014).

In the map25m, two clusters (termed clusters A and B) had an interesting topological property; each of them was completely embedded within a larger cluster – cluster A within putative PITd, and cluster B within putative CITv (**Figure 7a,c**). The preferred movie frames for clusters A and B are shown in **Supplementary Figure 14a,b**, respectively. Cluster A evidently had a preference for face images. In a comparison with classically localized face patches, cluster A showed a partial overlap with the face patch ML – middle lateral (Tsao et al., 2008), which is the largest and the most selective face patch in macaque IT cortex (**Figure 7c**). In a complimentary analysis, we used an automated face detection algorithm in Matlab (Viola-Jones algorithm, Viola and Jones, 2004) to label every two-second epoch of the movie as 'face epoch' or 'non-face epoch'. Cluster A showed a significantly higher response to face than non-face epochs, whereas cluster B, which was located outside of the face patches, had a low (and an approximately equal) response to face and non-face epochs (**Figure 7d**; stats are shown in the figure). Interestingly, the fMRI response in cluster A was positively correlated with the proportion of face frames within two-second epochs of the movie (**Figure 7e**; stats are shown in the figure). Cluster B was located ventral to an anterior face patch, in a region that was anatomically equivalent to a color-biased patch (Lafer-Sousa and Conway, 2013). In quantitative analyses, the preferred and non-preferred images of

cluster B contained a comparable amount of color information, and they both showed a significant bias towards daylight colors (**Supplementary Figure 15**; details of analyses and stats are described in the figure and figure caption). The ‘daylight color bias’ was significantly stronger in the preferred images ($p < 0.01$, permutation test for comparing the two correlation coefficients shown in **Supplementary Figure 15**). Thus, while we did not find an overall bias towards color in this cluster, there seemed to be a preference for daylight colors (orange-blue) over non-daylight colors (green-magenta).

Recent fMRI studies have reported place-selective patches (so-called the ‘place patches’) in macaques ([Rajimehr et al., 2011](#); [Nasr et al., 2011](#); [Kornblith et al., 2013](#)). These patches include mPPA (the monkey homologue of PPA) located within the posterior middle temporal sulcus, and LPP (lateral place patch) located within the occipitotemporal sulcus. In an independent group of monkeys, mPPA and LPP were identified using a conventional localizer experiment (**Figure 7f**; data from a representative monkey). Two clusters in the map25 corresponded to mPPA and LPP (**Figure 7g**). The LPP cluster included a patch that was isolated from the main part of the cluster and was located in the medial temporal cortex. This patch may be analogous to MPP (medial place patch) that is typically activated by microstimulation of LPP ([Kornblith et al., 2013](#)).

In the human data, our clustering approach enabled us to identify an area with a novel functional property. Similarly, this approach could help reveal a previously uncharacterized area in monkeys. The first sign of such area arose serendipitously when we were testing cortical activation maps in macaques for the labeled frames of the movie. To semantically label the movie frames, we recruited 100 human subjects through an online survey (Amazon Mechanical Turk) to classify each frame of the movie (sampled every two seconds) as either outdoor or indoor scene, and indicate whether the frame contained animate objects (faces, bodies, people, and animals) or not (see Methods for further details). In the comparison between animate versus inanimate movie frames, we found a network of cortical areas that was functionally more active for animate stimuli (**Supplementary Figure 16**). As expected, these areas included face patches and regions along the upper bank of STS that are thought to be involved in the processing of social stimuli in monkeys ([Sallet et al., 2011](#); [Lahnakoski et al., 2012](#)). Thus, highly accurate activation maps could be obtained by analyzing the fMRI responses for the labeled movie frames (see also [Bartels and Zeki, 2004a](#)). In the comparison between outdoor versus indoor scenes, we found an activation for outdoor scenes in the anterior-medial temporal cortex, in a region that was located further ventral to AITv (**Figure 8a**). This activation became much stronger when we

specifically contrasted “outdoor+animate scenes vs. indoor+animate scenes” (**Figure 8c**), and it was virtually absent in the contrast of “outdoor+inanimate scenes vs. indoor+inanimate scenes” (**Figure 8b**). The region manifesting such activation could possibly be a novel ‘place patch’ in macaques, because it showed a strong selectivity for specific outdoor scenes (i.e. outdoor scenes that contained animate objects). This region was named APP – anterior place patch, to emphasize its location in anterior-medial temporal cortex (**Figure 8c**).

APP was located within a large cluster in the map25m (**Figure 8d**, the brown patch). This cluster had two stripe-shaped segments: one segment located along the medial temporal cortex, and the other one located along the superior lip of STS immediately dorsal to STP. By continuing the hierarchical clustering, the entire cluster was partitioned into smaller regions, so that the most anterior subregion in the medial temporal segment matched topographically to APP (see the map of 92 clusters in **Figure 8d**). For this subregion, the preferred movie frames were outdoor+animate scenes (**Figure 8e**), and the non-preferred movie frames were indoor+animate scenes (**Figure 8f**). This analysis confirmed that the partitioning in the map, even at a fine scale, was based on movie-driven activations.

In the final analysis, we compared the parcellation maps between human and monkey (**Supplementary Figure 17**). Using the inter-species activity correlation (ISAC) method ([Mantini et al., 2012](#)), we computed the functional correspondence between map25 of humans and map25m of monkeys. The correlation matrix in **Supplementary Figure 17a** shows temporal correlation (similarity of time-courses) between clusters in map25 and clusters in map25m. If two clusters responded similarly to the movie, there would be a high correlation value for that pair in the correlation matrix. We examined three cases in which we observed highly significant positive correlations (**Supplementary Figure 17a,b**). In case #1, a cluster located in the foveal V2/V3/V4 in human cortex showed high correlation with clusters located in the corresponding foveal/parafoveal regions in monkey cortex (**Supplementary Figure 17c**, left panel). In case #2, face clusters in human showed high correlation with a face cluster and its surrounding region in monkey (**Supplementary Figure 17c**, middle panel). In case #3, a cluster located in the human motion-selective area (hMT+) showed high correlation with the monkey face cluster and also two other clusters that could be attributed to motion areas and mid-peripheral representations in the monkey extrastriate cortex (**Supplementary Figure 17c**, right panel). The correlation between motion and face clusters was also evident in the intra-species correlation matrices (**Supplementary Figure 17c**, plots in the rightmost panels). Recent evidence suggests that biological movements and magnocellular inputs contribute

strongly to the responses in macaque face patches during the free viewing of a natural, dynamic scene (Russ and Leopold, 2015).

Discussion:

In this fMRI study, we introduced a new, data-driven framework for functionally parcellating the entire visual cortex in humans and monkeys (**Supplementary Figure 18**). In this framework, we used a rich movie stimulus to drive the visual cortex and elicit a large variation in the patterns of response across the occipito-temporal voxels/vertices. Under natural viewing conditions, local fMRI responses show remarkable synchrony across subjects in a large expanse of cortex including visual areas in the occipital, ventral temporal, and lateral temporal cortices (Hasson et al., 2004). In contrast to univariate analyses, we took a multivariate approach to analyze raw time-varying patterns of response. Such approach began with averaging the time-courses of response in each vertex, across subjects. The averaging was done in a common anatomical space after spherical transformation of individual subjects' data into that common space. Other methods of averaging (such as 'hyperlignment' (Haxby et al., 2011)) may improve the estimation of group-average fMRI responses. In the next step, we used PCA to reduce the dimensionality of the activity space. PCA has been successfully used in previous fMRI studies for finding the most meaningful gradients/dimensions of semantic representation in visual cortex (Haxby et al., 2011; Huth et al., 2012). Other methods of dimensionality reduction (such as independent component analysis – ICA) can be also used (e.g. Bartels and Zeki, 2004b; Moeller et al., 2009; Norman-Haignere et al., 2015) – though such methods would normally require additional parameter settings. The actual parcellation occurred in the next step in which we applied a clustering algorithm in a low-dimensional PCA space. We used the hierarchical clustering algorithm, which had an advantage of defining clusters/parcels at different scales/resolutions. Finally, these clusters were projected onto the cortical surface to evaluate their spatial topography. We also carefully examined the functional specialization of many clusters in humans and monkeys. The parcellation maps showed clusters that corresponded to previously known visual cortical areas (e.g. six category-selective areas in humans). The maps also predicted new clusters/areas. Using a hypothesis-driven approach (e.g. using well-controlled, blocked and event-related fMRI experiments), one could explore in more detail the functional specificity of these clusters. For one cluster in humans and one cluster in monkeys, we took such an approach to characterize their functions.

Here we only included the occipito-temporal cortex in the functional slice prescription, partly because we wanted to maximize the spatial resolution in our imaging setup. The clustering analysis can be performed for the entire cerebral cortex including parietal and frontal lobes. Such analysis could reveal large cortical networks ('supra-areal organization') (Buckner and Yeo, 2014). Nodes within those networks could serve as 'seed points' in functional and structural connectivity analyses. On the other hand, clustering of vertices within a specific region could reveal 'sub-areal organization' (fine-grained distinctions) in that region. A similar approach has been used to find subregions within FFA and PPA (Çukur et al., 2013; Çukur et al., 2016).

In both humans and monkeys, the cortical parcellation was based on group-average time-courses. Due to the lack of stimulus repetitions in the movie, the statistical power was inherently low. Therefore, an extensive signal-averaging across subjects enabled us to have a better estimate of fMRI responses. The clusters obtained with this parcellation could be considered the most robust clusters that presumably play key roles in visual processing. However, it is possible that each subject has an idiosyncratic parcellation map. By scanning individual subjects in multiple sessions, one can assess the individual differences in the parcellation maps and in the layout of visual areas. Single-subject parcellation could also reveal small areas in 'balkanized' regions of cortex (e.g. face patches in anterior temporal cortex). These areas tend to be lost during the averaging process due to a high degree of variability in their locations. In the classical localizer experiments, the size and topography of areas in an individual subject depends on the amount of thresholding in the activation maps. This arbitrariness in threshold setting, which is a serious problem when comparing areas across subjects, can be avoided when parcellating the cortex by a clustering algorithm.

After parcellating the visual cortex, we used a specific type of region-of-interest analysis (a 'reverse-correlation' approach) to find the frames of the movie that produced the highest response in a given cluster. This analysis revealed the preferred movie frames that were consistent with the predicted functional selectivity of those clusters (e.g. a cluster within the fusiform gyrus (the putative FFA cluster) showed preference for movie frames containing faces). Linking the fMRI responses to the movie frames can be useful for addressing other interesting questions as well. The parcellation algorithm forces a set of data-points to be segmented into distinct clusters with 'hard borders' between them. However, the functional selectivity may not be homogeneous within a cluster. Instead, the selectivity may change smoothly within and across the clusters. To clarify whether such smooth transitions exist, one can

investigate the preferred movie frames for a region of cortex located around the border between two clusters. Furthermore, by systematically comparing the preferred movie frames across clusters, subtle variations in stimulus selectivity (tuning function) might be identified for clusters that show preference for a particular object category. For instance, there might be a systematic difference between the preferred scene images of PPA versus OPA.

The parcellation maps showed clusters in early visual cortex that spatially corresponded to the representation of eccentricity bands. This finding is consistent with the idea that ‘eccentricity bias’ is the major organizing principle in the occipito-temporal cortex ([Malach et al., 2002](#); [Hasson et al., 2002](#); [Hasson et al., 2003](#)). Interestingly, widespread correlation patterns of resting-state fMRI signal across early visual cortex also reflect topographic (eccentricity-based) organization ([Arcaro et al., 2015](#)). Thus, shared eccentricity representations may outweigh functional differences across anatomically-defined areas such as V2, V3, and V4.

In addition to visual areas, the temporal lobe includes auditory cortex, non-visual association cortex, and ‘semantic’ areas that are typically located in superior temporal gyrus and anterior temporal cortex. Accordingly, our parcellation maps showed large, less-differentiated clusters in those regions. These clusters may contain specific areas if they are mapped with appropriate auditory and/or word stimuli ([Huth et al., 2016](#)).

In humans, we found a cluster in lateral temporal cortex (the LT cluster), which demonstrated an ostensible preference for static images of human-object interactions. Although this region has been rarely charted in previous studies, there are some evidence suggesting that regions in posterior middle temporal gyrus (pMTG) respond preferentially to the images of tools and bodies/hands ([Beauchamp et al., 2002](#); [Orlov et al., 2010](#); [Bracci et al., 2012](#)), and that pMTG is implicated in the processing of tool use ([Lewis, 2006](#); [Gallivan et al., 2013](#); [Gallivan and Culham, 2015](#)). The presence of tools and hands in a picture may imply a form of human-object interaction, namely tool use. However, an intriguing result in our study was that the LT cluster was not sensitive to dynamic stimuli. One possible interpretation is that this region receives motion-subtracted input (‘implied motion’ signal), so that it efficiently extracts pure/invariant shape information in the action stimuli. The activity in this region might contribute to the ‘snapshot perception’ in patients who have akinetopsia (lesion in the motion area MT/V5 ([Zeki, 1991](#))).

Further studies are needed to shed light on the specificity and mechanisms of action representation in this region.

In monkeys, we characterized a functional role for an anterior place patch (APP) located in anterior-medial temporal cortex. This patch was more responsive to outdoor scenes including animate objects. Previous monkey fMRI studies had suggested the existence of such area ([Verhoef et al., 2015](#)) – though its specific function was not known. Unlike posterior place patches that process physical aspects of scenes and show low-level feature selectivity (such as sensitivity to high spatial frequencies ([Rajimehr et al., 2011](#)) and textures ([Kornblith et al., 2013](#); see also [Cant and Xu, 2012](#))), APP may process more semantic aspects of scenes and integrate information about a scene with information about animate objects embedded in that scene. A recent electrophysiological study has reported that place-selective neurons in anterior temporal cortex (area TEd) show a response bias for environmental/landscape stimuli that are defined by 3D depth cues ([Vaziri et al., 2014](#)). Outdoor scenes arguably contain more 3D depth cues than indoor scenes, and those 3D shape information may contribute to the response in APP.

Previous studies have used resting-state fMRI data to parcellate cortex in humans. Yeo et al. identified a set of cortical networks using a clustering analysis applied on data from a population of 1,000 healthy subjects ([Yeo et al., 2011](#)). Other studies employed various computational techniques to parcellate individual subjects' cortices (e.g. [Wig et al., 2014](#); [Laumann et al., 2015](#); [Wang et al., 2015a](#)). Using boundary mapping ([Cohen et al., 2008](#)) and graph theory, Nelson et al. parcellated the lateral parietal cortex into six distinct 'modules' ([Nelson et al., 2010](#)). The boundary mapping technique has been also used to parcellate the whole brain ([Gordon et al., 2016](#)). Using a module detection algorithm, Goulas et al. parcellated the lateral frontal cortex ([Goulas et al., 2012](#)). Using *k*-means clustering, Kahnt et al. parcellated the orbitofrontal cortex ([Kahnt et al., 2012](#)). Using multimodal MR images from the Human Connectome Project (HCP) database and a semi-automated gradient-based parcellation approach, Glasser et al. delineated 180 areas in human cerebral cortex ([Glasser et al., 2016](#)). In one fMRI study ([Vul et al., 2012](#)), human subjects were scanned while viewing rapid event-related presentations of 69 unique images drawn from 9 object categories. Using data-driven clustering of voxels, this study found face, place, and body clusters/systems in the ventral visual pathway. Finally, Kiani et al. parcellated macaque prearcuate cortex into spatially segregated subregions ('mesoscale subnetworks'), based on the responses of single- and multi-units recorded from arrays of electrodes ([Kiani et al., 2015](#)).

What is the importance of cortical parcellation? A great deal of evidence suggests that cerebral cortex in primates is functionally compartmentalized (see [Kanwisher, 2010](#) for review). A full characterization of the anatomical layout of cortical areas would be a fundamental step in understanding the function of those areas if we assume a tight relationship between structure and function in cortex. Characterizing the cortical maps could also help predict what behavioral and perceptual changes would occur in pathological cases where certain regions of cortex are affected by macroscopic damage/atrophy. Comparing the parcellation maps across species (e.g. between human and monkey) could be useful for inferring cortical homologies. Defining cortical homologies has important implications for (i) generalizing the neurobiological findings across species, and (ii) understanding how cortex has been organized in the course of development and evolution. ■

Methods:

Subjects

Five human subjects (with normal or corrected-to-normal vision) participated in the movie experiment. The same human subjects were tested in the second, block-design experiment. Written informed consent was obtained from all human subjects prior to the scanning session. The human subjects were scanned in MIT Martinos Imaging Center.

Two juvenile male rhesus monkeys (*Macaca mulatta*) were tested in the movie experiment. Surgical details and the training procedure for monkeys have been described elsewhere ([Kornblith et al., 2013](#)). The monkey subjects were scanned in Caltech Brain Imaging Center.

For labeling the movie frames, 100 participants ('Providers') were recruited online through Amazon Mechanical Turk (MTurk). We required that participants had performed at least 1,000 previous MTurk human intelligence tasks (HITs) and that at least 99% of previous HITs were accepted by their experimenters ('Requesters').

All experimental procedures conformed to US National Institutes of Health guidelines. The human and monkey protocols were approved by MIT Institutional Review Board (IRB) and Caltech Institutional

Animal Care and Use Committee (IACUC), respectively (MIT protocol # 1302005519; Caltech protocol # 1574).

Visual Stimuli

Stimuli were generated on an Apple Mac computer and presented via an LCD projector (1,024×768 pixels resolution, 60-Hz refresh rate) onto a rear-projection screen. Matlab and Psychophysics Toolbox (<http://psychtoolbox.org/>) were used to program the experiments.

In the movie experiment, a complete movie (*Baraka*, 1992, 96 minutes, 24 fps, directed by Ron Fricke) was presented to the human and monkey subjects. *Baraka* is a non-narrative documentary film that presents footage of people, places and things from around the world. The movie was divided into 10 sections, 9.6 minutes each. These sections were successively presented in 10 functional runs. The movie was presented in a full-screen mode with no sound. A blank epoch was included at the beginning and at the end of each run (6 seconds at the beginning; 14 seconds at the end). The human subjects were instructed to freely watch the movie (i.e. no fixation was required). The monkey subjects were periodically rewarded by juice drops for maintaining their gaze within the screen window. Eye position was monitored in humans and monkeys using an infrared pupil tracking system (EyeLink 1000).

In the block-design experiment, the stimuli were either dynamic (a 5-second movie clip) or static (a 5-second static frame/picture; the frame was extracted from the middle of the movie clip). 300 movie clips were selected from 20 categories of human-object interactions from UCF101 action recognition dataset (Soomro et al., 2012). The stimuli ($10^\circ \times 7.5^\circ$ in size) were colored and were presented on a uniform gray background. Each subject was scanned in 10 runs. Each run consisted of 10 dynamic and 10 static stimulus blocks that were presented alternatively. Each stimulus block consisted of 3 stimuli/trials, each lasting 5 seconds (block duration = 15 seconds). Throughout the functional scans, human subjects were instructed to continuously fixate a small fixation point at the center of screen. A 15-second blank epoch ('fixation-only' block) was presented at the beginning, middle, and end of each run.

For labeling the movie frames, we extracted a set of 2710 non-blank frames/images from the movie. The frames were sampled every 2 seconds. 10 groups of MTurk participants labeled the frames. In each group, each participant labeled 271 frames, thus 10 participants per group were needed to label the

whole frames. For a given frame, we assigned a label that was chosen by the majority of participants. There were 6 possible labels for each frame, based on answers to two questions: 1) whether the scene in the frame was outdoor (e.g. sky, mountain, city street), indoor (e.g. room), or unclear, and 2) whether the frame contained animate objects (people, faces, body parts, animals, and artworks depicting faces and/or bodies) or not. **Table 1** shows the percentage of labeled frames in each category.

Table 1: Percentage of labeled movie frames

Category	Percentage
outdoor+animate	41.93%
outdoor+inanimate	34.05%
indoor+animate	19.41%
indoor+inanimate	4.06%
unclear+animate	0.12%
unclear+inanimate	0.44%

Imaging Procedures

All subjects were scanned in a 3T horizontal bore scanner/magnet (Siemens Tim Trio). T2*-weighted gradient-echo EPI sequences were used for functional imaging (human: 32-channel surface coil, TR = 2,000 ms, TE = 30 ms, flip angle = 90°, 2 mm isotropic voxels, 30 semi-axial slices with no gap, and parallel imaging (acceleration factor 2); monkey: 8-channel phased array coil, TR = 2,000 ms, TE = 16 ms, flip angle = 90°, 1 mm isotropic voxels, 50 axial slices with no gap, and parallel imaging (acceleration factor 2)). The first 4 volumes of each run were discarded to allow for MR signal equilibration. To increase functional sensitivity in the monkey scans (to compensate for smaller voxels), we used a head-only gradient insert (Siemens AC88) and an exogenous MION-based contrast agent (Feraheme; 8-10 mg/kg IV).

T1-weighted MP-RAGE sequences (1 mm isotropic voxels in humans; 0.5 mm isotropic voxels in monkeys) were used for high-resolution anatomical imaging from the same subjects. Monkeys were

anesthetized during the anatomical imaging. In each scanning session, one or more field maps were also acquired to correct for local magnetic field inhomogeneity and improve alignment/co-registration of the functional scans with the anatomical scans.

Data Analysis

Functional and anatomical data were preprocessed and analyzed using FreeSurfer and FS-FAST (<http://surfer.nmr.mgh.harvard.edu/>). For each subject, the cortical surfaces were computationally reconstructed by analyzing the anatomical MR images. All functional MR volumes were first skull-stripped using FSL's brain extraction tool to create a mask of brain-only voxels, then motion-corrected using AFNI's motion correction algorithm. In monkeys, the field maps and FSL's FUGUE algorithm were used to geometrically undistort/unwarp the motion-corrected volumes. The next step of preprocessing was intensity normalization. Within the brain mask, the mean intensity of all voxels across all time points was computed. The intensity value at each voxel at each time point was then divided by the mean intensity and multiplied by 100. The functional volumes were rigidly registered to the same-subject anatomicals, then spherically transformed/resampled to a common anatomical space (surfaces of a reference subject in humans and F99 in monkeys). The functional values were spatially smoothed on the surface using a 2-D Gaussian kernel (3 mm FWHM in humans, and 2 mm FWHM in monkeys).

To obtain the time-course of response in each vertex of the cortical surface, we calculated the residuals in each functional run after detrending the signal (using a second-order polynomial fitting) and pre-whitening (to remove temporal autocorrelations in the signal). We also removed the effects of nuisance regressors including subject's head motion, global mean signal waveform, and the mean time-series in white matter, ventricles, and CSF. In each subject, the time-courses of 10 functional runs were concatenated to obtain a matrix of [vertices x time-points] in each hemisphere for the entire movie. The polarity of the MION-based MR responses was inverted, so that human BOLD and monkey MION data could be conveniently compared. In humans, data from all subjects were averaged separately for right and left hemispheres. In monkeys, data from both subjects (four hemispheres) were averaged. The average data matrix was used for the PCA analysis.

To obtain the activation maps, we did a univariate General Linear Model analysis on detrended and whitened signals. Stimulus-related regressors were defined as boxcar functions convolved with the SPM

hemodynamic response function (in humans) or a gamma hemodynamic response function (in monkeys). The parameters related to subject's head motion were also included in the GLM, so that they were regressed out. The average beta weights were calculated for each stimulus condition. Vertex-wise statistical tests were conducted for a contrast of two conditions, and thresholded statistical maps were obtained.

For clustering, we used an agglomerative hierarchical clustering algorithm. The squared Euclidean distance was used as a distance metric, and the Ward's method was used for linkage. The similarity between two hierarchical clusterings (clustering similarity – CS) was determined based on Fowlkes–Mallows index (Fowlkes and Mallows, 1983):

$$CS_k = \frac{\sum_{i=1}^k \sum_{j=1}^k m_{i,j}^2 - n}{\sqrt{\left(\sum_{i=1}^k \left(\sum_{j=1}^k m_{i,j} \right)^2 - n \right) \left(\sum_{j=1}^k \left(\sum_{i=1}^k m_{i,j} \right)^2 - n \right)}}$$

where n is the number of objects in the two clusterings, k is the number of clusters, and $m_{i,j}$ is the number of objects in common between the i th cluster in one clustering and the j th cluster in the other clustering. CS can be calculated for every k , and it ranges between 0 and 1. A higher value for CS means that the two clusterings are systematically related to each other. The mathematical procedure for calculating the analytic chance, where the two clusterings are assumed to be completely unrelated, has been described in Fowlkes and Mallows, 1983.

To find the preferred and non-preferred movie frames of a cluster, we first obtained the mean time-course across vertices of the cluster. The mean time-course was z-normalized, and its significant peaks were detected. The resulting time-points were sorted based on the magnitude of response. We then reconstructed short (2-second) movie segments composed of the frames that led to the highest and lowest responses, assuming a standard hemodynamic lag of 5 seconds for BOLD response (Menon et al., 1995; Boynton et al., 1996) and 4 seconds for MION response (Vanduffel et al., 2001). In the figures, the middle frames from the selected movie segments are shown. These preferred (or non-preferred) images

are ordered consecutively first from left to right in a row then from top to bottom – where the top-left image elicited the highest (or lowest) response.

Acknowledgements:

We would like to thank Nicole Schweers for help in collecting monkey data, Christina Triantafyllou for setting up the pulse sequences in the human scanning, Doug Greve, Sheeba Arnold, Shahin Nasr, Maryam Vaziri Pashkam, Elias Issa, Michael Arcaro, Jonathan Polimeni, Alexander Kell, Elahe' Yargholi, Donna Dierker, Kaitlin Bohon, Christoph Witzel, Karthik Srinivasan for help in data analysis and providing scripts/datasets, and Roger Tootell, Nancy Kanwisher, Jim DiCarlo, Bevil Conway for helpful comments. This research was supported by McGovern Institute for Brain Research, Martinos Imaging Center at MIT, and NIH grant R01 EY019702 to D.T. The authors declare no competing financial interests.

References:

1. Arcaro, M.J., Honey, C.J., Mruczek, R.E., Kastner, S., and Hasson, U. (2015). Widespread correlation patterns of fMRI signal across visual cortex reflect eccentricity organization. *eLife* 4.
2. Baldassano, C., Beck, D.M., and Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. *NeuroImage* 75, 228-237.
3. Bartels, A., and Zeki, S. (2004a). Functional brain mapping during free viewing of natural scenes. *Human brain mapping* 21, 75-85.
4. Bartels, A., and Zeki, S. (2004b). The chronoarchitecture of the human brain--natural viewing conditions reveal a time-based anatomy of the brain. *NeuroImage* 22, 419-433.
5. Bartels, A., Zeki, S., and Logothetis, N.K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cerebral cortex* 18, 705-717.
6. Beauchamp, M.S., Lee, K.E., Haxby, J.V., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34, 149-159.
7. Born, R.T., and Bradley, D.C. (2005). Structure and function of visual area MT. *Annual review of neuroscience* 28, 157-189.
8. Boussaoud, D., Desimone, R., and Ungerleider, L.G. (1991). Visual topography of area TEO in the macaque. *The Journal of comparative neurology* 306, 554-575.
9. Boynton, G.M., Engel, S.A., Glover, G.H., and Heeger, D.J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 16, 4207-4221.
10. Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., and Peelen, M.V. (2012). Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. *Journal of neurophysiology* 107, 1443-1456.
11. Buckner, R.L., and Yeo, B.T. (2014). Borders, map clusters, and supra-areal organization in visual cortex. *NeuroImage* 93 Pt 2, 292-297.

12. Cant, J.S., and Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32, 7685-7700.
13. Cohen, A.L., Fair, D.A., Dosenbach, N.U., Miezin, F.M., Dierker, D., Van Essen, D.C., Schlaggar, B.L., and Petersen, S.E. (2008). Defining functional areas in individual human brains using resting functional connectivity MRI. *NeuroImage* 41, 45-57.
14. Cukur, T., Huth, A.G., Nishimoto, S., and Gallant, J.L. (2013). Functional subdomains within human FFA. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33, 16748-16766.
15. Cukur, T., Huth, A.G., Nishimoto, S., and Gallant, J.L. (2016). Functional Subdomains within Scene-Selective Cortex: Parahippocampal Place Area, Retrosplenial Complex, and Occipital Place Area. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 36, 10257-10273.
16. Desimone, R., and Gross, C.G. (1979). Visual areas in the temporal cortex of the macaque. *Brain research* 178, 363-380.
17. Dilks, D.D., Julian, J.B., Paunov, A.M., and Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33, 1331-1336.
18. Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science* 293, 2470-2473.
19. Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598-601.
20. Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex* 1, 1-47.
21. Fowlkes, E.B., and Mallows, C.L. (1983). A Method for Comparing 2 Hierarchical Clusterings. *J Am Stat Assoc* 78, 553-569.
22. Gallivan, J.P., and Culham, J.C. (2015). Neural coding within human brain areas involved in actions. *Current opinion in neurobiology* 33, 141-149.
23. Gallivan, J.P., McLean, D.A., Valyear, K.F., and Culham, J.C. (2013). Decoding the neural mechanisms of human tool use. *eLife* 2, e00425.
24. Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., *et al.* (2016). A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171-178.
25. Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., and Petersen, S.E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral cortex* 26, 288-303.
26. Goulas, A., Uylings, H.B., and Stiers, P. (2012). Unravelling the intrinsic functional organization of the human lateral frontal cortex: a parcellation scheme based on resting state fMRI. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32, 10238-10252.
27. Grill-Spector, K., and Weiner, K.S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature reviews Neuroscience* 15, 536-548.
28. Halgren, E., Dale, A.M., Sereno, M.I., Tootell, R.B., Marinkovic, K., and Rosen, B.R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human brain mapping* 7, 29-37.
29. Hasson, U., Harel, M., Levy, I., and Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37, 1027-1041.
30. Hasson, U., Levy, I., Behrmann, M., Hendler, T., and Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34, 479-490.

31. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634-1640.
32. Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., and Ramadge, P.J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404-416.
33. Henriksson, L., Karvonen, J., Salminen-Vaparanta, N., Railo, H., and Vanni, S. (2012). Retinotopic maps, spatial tuning, and locations of human visual areas in surface coordinates characterized with multifocal and blocked fMRI designs. *PLoS one* 7, e36859.
34. Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., and Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 453-458.
35. Huth, A.G., Nishimoto, S., Vu, A.T., and Gallant, J.L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76, 1210-1224.
36. Janssens, T., Zhu, Q., Popivanov, I.D., and Vanduffel, W. (2014). Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 34, 10156-10167.
37. Julian, J.B., Fedorenko, E., Webster, J., and Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage* 60, 2357-2364.
38. Kahnt, T., Chang, L.J., Park, S.Q., Heinzle, J., and Haynes, J.D. (2012). Connectivity-based parcellation of the human orbitofrontal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32, 6240-6250.
39. Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America* 107, 11163-11170.
40. Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 17, 4302-4311.
41. Kiani, R., Cueva, C.J., Reppas, J.B., Peixoto, D., Ryu, S.I., and Newsome, W.T. (2015). Natural grouping of neural responses reveals spatially segregated clusters in prearcuate cortex. *Neuron* 85, 1359-1373.
42. Kolster, H., Janssens, T., Orban, G.A., and Vanduffel, W. (2014). The retinotopic organization of macaque occipitotemporal cortex anterior to V4 and caudoventral to the middle temporal (MT) cluster. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 34, 10168-10191.
43. Kolster, H., Peeters, R., and Orban, G.A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30, 9801-9820.
44. Kornblith, S., Cheng, X., Ohayon, S., and Tsao, D.Y. (2013). A network for scene processing in the macaque temporal lobe. *Neuron* 79, 766-781.
45. Kriegeskorte, N., Bodurka, J., and Bandettini, P. (2008). Artfactual time-course correlations in echo-planar fMRI with implications for studies of brain function. *Int J Imaging Syst Technol* 18, 345-349.
46. Lafer-Sousa, R., and Conway, B.R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nature neuroscience* 16, 1870-1878.
47. Lahnakoski, J.M., Glerean, E., Salmi, J., Jaaskelainen, I.P., Sams, M., Hari, R., and Nummenmaa, L. (2012). Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the

- distributed brain network for social perception. *Frontiers in human neuroscience* 6, 233.
48. Laumann, T.O., Gordon, E.M., Adeyemo, B., Snyder, A.Z., Joo, S.J., Chen, M.Y., Gilmore, A.W., McDermott, K.B., Nelson, S.M., Dosenbach, N.U., *et al.* (2015). Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron* 87, 657-670.
 49. Levy, I., Hasson, U., Avidan, G., Hendler, T., and Malach, R. (2001). Center-periphery organization of human object areas. *Nature neuroscience* 4, 533-539.
 50. Lewis, J.W. (2006). Cortical networks related to human use of tools. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* 12, 211-231.
 51. Malach, R., Levy, I., and Hasson, U. (2002). The topography of high-order human object areas. *Trends in cognitive sciences* 6, 176-184.
 52. Mantini, D., Hasson, U., Betti, V., Perrucci, M.G., Romani, G.L., Corbetta, M., Orban, G.A., and Vanduffel, W. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nature methods* 9, 277-282.
 53. Menon, R.S., Ogawa, S., Hu, X., Strupp, J.P., Anderson, P., and Ugurbil, K. (1995). BOLD based functional MRI at 4 Tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals. *Magnetic resonance in medicine* 33, 453-459.
 54. Moeller, S., Nallasamy, N., Tsao, D.Y., and Freiwald, W.A. (2009). Functional connectivity of the macaque brain across stimulus and arousal states. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29, 5897-5909.
 55. Nasr, S., Devaney, K.J., and Tootell, R.B. (2013). Spatial encoding and underlying circuitry in scene-selective cortex. *NeuroImage* 83, 892-900.
 56. Nasr, S., Liu, N., Devaney, K.J., Yue, X., Rajimehr, R., Ungerleider, L.G., and Tootell, R.B. (2011). Scene-selective cortical regions in human and nonhuman primates. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31, 13771-13785.
 57. Nelson, S.M., Cohen, A.L., Power, J.D., Wig, G.S., Miezin, F.M., Wheeler, M.E., Velanova, K., Donaldson, D.I., Phillips, J.S., Schlaggar, B.L., *et al.* (2010). A parcellation scheme for human left lateral parietal cortex. *Neuron* 67, 156-170.
 58. Nishimoto, S., Huth, A.G., Bilenko, N.Y., and Gallant, J.L. (2017). Eye movement-invariant representations in the human visual system. *Journal of vision* 17, 11.
 59. Norman-Haignere, S., Kanwisher, N.G., and McDermott, J.H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron* 88, 1281-1296.
 60. Ojemann, J.G., Akbudak, E., Snyder, A.Z., McKinstry, R.C., Raichle, M.E., and Conturo, T.E. (1997). Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *NeuroImage* 6, 156-167.
 61. Orlov, T., Makin, T.R., and Zohary, E. (2010). Topographic representation of the human body in the occipitotemporal cortex. *Neuron* 68, 586-600.
 62. Pinsk, M.A., Arcaro, M., Weiner, K.S., Kalkus, J.F., Inati, S.J., Gross, C.G., and Kastner, S. (2009). Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. *Journal of neurophysiology* 101, 2581-2600.
 63. Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., and Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage* 56, 2356-2363.
 64. Rajimehr, R., Devaney, K.J., Bilenko, N.Y., Young, J.C., and Tootell, R.B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS biology* 9, e1000608.
 65. Rajimehr, R., and Tootell, R. (2008). Organization of Human Visual Cortex. In *The Senses: A*

- Comprehensive Reference, R.H. Masland, and T.D. Albright, eds. (New York: Academic Press), pp. 595-614.
66. Russ, B.E., and Leopold, D.A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *NeuroImage* 109, 84-94.
 67. Sallet, J., Mars, R.B., Noonan, M.P., Andersson, J.L., O'Reilly, J.X., Jbabdi, S., Croxson, P.L., Jenkinson, M., Miller, K.L., and Rushworth, M.F. (2011). Social network size affects neural circuits in macaques. *Science* 334, 697-700.
 68. Schwarzlose, R.F., Baker, C.I., and Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25, 11055-11059.
 69. Soomro, K., Zamir, A.R., and Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:12120402.
 70. Steinbach, M., Ertöz, L., and Kumar, V. (2004). The Challenges of Clustering High Dimensional Data. In *New Directions in Statistical Physics: Econophysics, Bioinformatics, and Pattern Recognition*, L.T. Wille, ed. (Berlin: Springer), pp. 273-309.
 71. Tootell, R.B., Silverman, M.S., Switkes, E., and De Valois, R.L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science* 218, 902-904.
 72. Tootell, R.B., Switkes, E., Silverman, M.S., and Hamilton, S.L. (1988). Functional anatomy of macaque striate cortex. II. Retinotopic organization. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 8, 1531-1568.
 73. Tsao, D.Y., Moeller, S., and Freiwald, W.A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America* 105, 19514-19519.
 74. Ungerleider, L.G., and Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *The Journal of comparative neurology* 248, 190-222.
 75. Van Essen, D.C. (2004). Organization of Visual Areas in Macaque and Human Cerebral Cortex. In *The Visual Neurosciences*, L.M. Chalupa, and J.S. Werner, eds. (Cambridge: MIT Press), pp. 507-521.
 76. Vanduffel, W., Fize, D., Mandeville, J.B., Nelissen, K., Van Hecke, P., Rosen, B.R., Tootell, R.B., and Orban, G.A. (2001). Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron* 32, 565-577.
 77. Vaziri, S., Carlson, E.T., Wang, Z., and Connor, C.E. (2014). A channel for 3D environmental shape in anterior inferotemporal cortex. *Neuron* 84, 55-62.
 78. Verhoef, B.E., Bohon, K.S., and Conway, B.R. (2015). Functional architecture for disparity in macaque inferior temporal cortex and its relationship to the architecture for faces, color, scenes, and visual field. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 35, 6952-6968.
 79. Viola, P., and Jones, M.J. (2004). Robust real-time face detection. *Int J Comput Vision* 57, 137-154.
 80. Vul, E., Lashkari, D., Hsieh, P.J., Golland, P., and Kanwisher, N. (2012). Data-driven functional clustering reveals dominance of face, place, and body selectivity in the ventral visual pathway. *Journal of neurophysiology* 108, 2306-2322.
 81. Wandell, B.A., Dumoulin, S.O., and Brewer, A.A. (2007). Visual field maps in human cortex. *Neuron* 56, 366-383.
 82. Wang, D., Buckner, R.L., Fox, M.D., Holt, D.J., Holmes, A.J., Stoecklein, S., Langs, G., Pan, R., Qian, T., Li, K., *et al.* (2015a). Parcellating cortical functional networks in individuals. *Nature neuroscience* 18, 1853-1860.
 83. Wang, L., Mruczek, R.E., Arcaro, M.J., and Kastner, S. (2015b). Probabilistic Maps of Visual

Topography in Human Cortex. *Cerebral cortex* 25, 3911-3931.

84. Weiner, K.S., and Grill-Spector, K. (2011). Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *NeuroImage* 56, 2183-2199.
85. Wig, G.S., Laumann, T.O., Cohen, A.L., Power, J.D., Nelson, S.M., Glasser, M.F., Miezin, F.M., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2014). Parcellating an individual subject's cortical and subcortical brain structures using snowball sampling of resting-state correlations. *Cerebral cortex* 24, 2036-2054.
86. Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., *et al.* (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* 106, 1125-1165.
87. Zeki, S. (1991). Cerebral akinetopsia (visual motion blindness). A review. *Brain : a journal of neurology* 114 (Pt 2), 811-824.

Figure Captions:

Figure 1. Structural data, functional data, and PCA analysis

A, Structural data were used to reconstruct the cortical surface in each hemisphere and make flattened patches. The red-green overlay shows major sulci-gyri of cortex, respectively. Sulcal abbreviations: CaS, calcarine; POS, parieto-occipital; IPS, intraparietal; CeS, central; CoS, collateral; OTS, occipito-temporal; STS, superior temporal; LS, lateral. B, Using an automated anatomical parcellation in Freesurfer, all the occipito-temporal voxels/vertices were extracted (~ 40,000 vertices per hemisphere). The anatomical parcels in the occipito-temporal lobe (within the yellow boundary) are listed on the right. C, Functional data were based on a 96-minute natural movie stimulus (2810 time-points). After preprocessing, the time-courses of activity were averaged across subjects to obtain an average time-course in each vertex. D, A functional activity space was constructed. Each axis represents the functional activity at a given time-point. Data-points are vertices. E, The dimensionality of the activity space was reduced using PCA. The vertices were transformed into a new coordinate system, with the principal components (PCs) as new dimensions. The first three dimensions are shown. A color (an RGB value) was assigned for each vertex, based on its position in the 3-dimensional PCA space. F, The colored vertices were visualized on the flat patch of cortical surface (the PCA map).

Figure 2. The first principal component correlates with temporal variation of luminance in the movie

A, Every two-second segment of the movie contained 48 frames. Each movie frame was geometrically represented as a point in a high-dimensional pixel space where dimensions were intensity/luminance values of pixels of the grayscale frame. The 48 points formed a trajectory of luminance variation in the course of two seconds. The luminance variation index (LVI) was defined as the sum of the Euclidean distances between each frame and the subsequent frame in the pixel space. B, The correlation between LVI and principal component (PC) weights. For each time-point, LVI and the weights of PC1, PC2, and PC3 were obtained. Data-points in X-axis (LVI) and Y-axis (PC weight) were divided into 40 bins via an equal-frequency binning procedure. LVIs and PC weights were averaged in those bins, and the average data-points were plotted. Two data-points with very low and very high values were excluded as outliers. Pearson r was used to compute correlation coefficient in these plots and other correlation plots throughout the paper. C, The map of PC1 scores on the flat patch of cortical surface.

Figure 3. Hierarchical clustering of occipito-temporal vertices

A, The first ten principal components, indicated by a bracket, captured most of the variance in fMRI data. For the clustering purpose, the occipito-temporal vertices were represented in a 10-dimensional PCA space. B, Hierarchical clustering tree/dendrogram. The clusters were color-coded. At each level of clustering, the newly-formed cluster received a distinct color if it was the smaller part of the parent cluster. C, The functionally-defined clusters were projected onto the cortical surface, so that their spatial organization could be evaluated. The maps show clusters at nine different levels (number of clusters = 2, 3, 4, 5, 6, 7, 8, 9, and 25). The map with 25 clusters was named 'map25'.

Figure 4. The parcellation map reveals retinotopic and category-selective clusters

A, map25 clusters in early visual cortex shown on the flat patch. B, The average eccentricity map ($n = 53$) shown on the flat patch. Data from [Wang et al., 2015b](#). C, Selected clusters of map25 that corresponded to category-selective areas. These clusters are labeled with white font. The black retinotopic borders are based on average retinotopic mapping data ($n = 15$) from [Henriksson et al., 2012](#). The orange

borders/labels show classically localized category-selective areas derived from an average map ($n = 40$, Julian et al., 2012). Abbreviations in panels A-C: phPIT, putative human homologue of macaque PIT; phTEO, putative human homologue of macaque TEO; XFC, extended foveal cortex; FFA, fusiform face area; pSTS, posterior STS face area; PPA, parahippocampal place area; OPA, occipital place area; EBA, extrastriate body area; BA+, additional body areas; LT*, lateral temporal cluster (the asterisk denotes that this cluster/area is being reported for the first time); STS, superior temporal sulcus; MTG, middle temporal gyrus; ITS, inferior temporal sulcus. D, Example movie frames of the movie segments that produced the highest activation in the LT cluster (see Methods for more details). Each row shows one movie segment. E, The z-normalized fMRI response in the LT cluster as a function of motion energy in the movie. Using a block-matching algorithm in Matlab, motion energy (magnitude of motion) was estimated for every two-second segment of the movie. Each segment contained 48 movie frames. Each frame was spatially divided into non-overlapping macroblocks. The movement vectors were calculated by comparing corresponding blocks between two frames separated by one second (i.e. frame # 25 and frame # 1, frame # 26 and frame # 2, and so on). The total motion energy in a two-second movie segment was obtained by averaging the amounts of displacement in the movement vectors across all blocks and all frame comparisons. The movie segments were classified into 10 groups based on motion energy values (small motion \rightarrow large motion), and an average response in the LT cluster was computed for each group. Positive and negative responses are shown by green and red colors, respectively.

Figure 5. The LT cluster responds preferentially to static images of human-object interactions

A, Example frames of movie-clip stimuli used in the block-design experiment. The stimulus set included 20 categories of human-object interactions. B, A fixed-effects average map of three subjects showing fMRI activations for static (yellow) vs. dynamic (cyan) stimuli. The map is shown on the lateral view of inflated surface (STS: superior temporal sulcus). The colorbar here and in the other activation maps represents P values. The border of LT cluster was derived from the map25. The border of motion-selective area MT was derived from a motion localizer map (see Supplementary Figure 11). C, The activation maps in individual subjects (S1, S2, S3). D-E, Scatter plots showing vertex-wise comparison of fMRI percent signal change (PSC) for static vs. dynamic stimuli in MT (D) and LT (E). Red, green, and blue data-points correspond to three subjects in panel C.

Figure 6. Functional parcellation of macaque visual cortex

A, Flat patch of macaque F99's right hemisphere. The red-green overlay shows major sulci-gyri of cortex, respectively. Sulcal abbreviations: CaS, calcarine; LuS, lunate; IOS, inferior occipital; POS, parieto-occipital; IPS, intraparietal; STS, superior temporal; OTS, occipito-temporal; HS, hippocampal; CiS, cingulate; CeS, central; LS, lateral; AS, arcuate; PS, principal. B, The PCA map in the occipito-temporal cortex. The occipito-temporal vertices were defined based on a lobar parcellation of macaque F99 in the Caret software (<http://brainvis.wustl.edu/wiki/index.php/Caret:About>). C, Hierarchical clustering was applied in the PCA space, and the functionally-defined clusters were projected onto the cortical surface. The maps show clusters at twelve different levels (number of clusters = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 25). The map with 25 clusters was named 'map25m'.

Figure 7. Selected clusters of map25m and their anatomical/functional significance

A, Clusters within macaque IT cortex that corresponded to anatomically-defined areas in Felleman and Van Essen (1991) scheme. The full layout of Felleman and Van Essen areas is shown in panel B. The IT clusters included VOT, PITd, PITv, CITd, CITv, AITd, AITv, and STP. C, Within PITd and CITv, there were smaller clusters that were named cluster A and cluster B, respectively. The location of IT face patches was determined based on a probabilistic map of three monkeys, cutoff 66% (Janssens et al., 2014). These face patches included ML (middle lateral), MF (middle fundus), AL (anterior lateral), and AF (anterior fundus). STS: superior temporal sulcus, PMTS: posterior middle temporal sulcus, OTS: occipito-temporal sulcus. D, The z-normalized fMRI response in clusters A and B for face and non-face epochs of the movie. A two-second epoch of the movie was labeled as a face epoch if all of its frames contained frontal and/or profile faces, and it was labeled as a non-face epoch if none of its frames contained faces. In total, there were 327 face epochs and 350 non-face epochs. E, The correlation between fMRI response in cluster A and proportion of 'face frames' in two-second epochs of the movie. The first and last data-points show averaged responses for non-face and face epochs, respectively (see panel D). The remaining data-points show averaged responses within 10 equal-size bins along the X-axis. The red line is linear regression fit. F, A conjunction map showing macaque place patches (mPPA and LPP) on the inflated surface (STS: superior temporal sulcus). The map of each contrast was thresholded at $p < 10^{-8}$,

then a conjunction map was obtained by choosing vertices that had a significant activation above the threshold in all three contrasts (i.e. place vs. face, place vs. object, and place vs. scrambled place). The conjunction map was slightly smoothed. The source of data was [Verhoef et al., 2015](#). G, Clusters of map25m that spatially corresponded to mPPA, LPP, and possibly MPP.

Figure 8. APP responds preferentially to outdoor scenes containing animate objects

A-C, All time-points of the movie were manually labeled as outdoor or indoor scenes containing animate or inanimate objects. The activation maps are shown for different contrasts on a lateral view of inflated cortex (STS: superior temporal sulcus). As indicated by the axis-rotation icon, the inflated cortex has been slightly tilted for a better visualization of ventral surface. D, A parcel in the map of 92 clusters matched topographically to APP (anterior place patch). This parcel was part of a larger cluster in map25m, which is displayed on a patch of flattened temporal cortex (FTC). For this parcel, preferred and non-preferred movie frames are shown in panels E and F, respectively.

Supplementary Figure 1. Functional slice prescription in the human scans

30 EPI slices were used to cover the occipito-temporal cortex.

Supplementary Figure 2. PCA maps for intact and shuffled data in right and left hemispheres

RH: right hemisphere, LH: left hemisphere. The map shown in panel A is the same as the map shown in panel F of Figure 1. In panels C and D, the time-course of response in each occipito-temporal vertex was circularly shifted by a random amount (see the formula in panel C), then the shuffled time-courses were used in PCA.

Supplementary Figure 3. SNR map in human

The map shows fsnr (functional signal-to-noise ratio) map in a representative subject/hemisphere. For each occipito-temporal vertex, fsnr was computed as the mean of regressed out time-course divided by the standard deviation of GLM residuals.

Supplementary Figure 4. Quantitative characteristics of PCA and clustering analyses

A, For principal components greater than an ‘elbow point’ of 10, the second derivative of PCA eigenvalues was near zero, meaning that each of those components added only a small and linear amount to the explained variance. Thus, the first 10 principal components were used. B, For cluster numbers greater than 25, the clusters were very small in size. For cluster numbers greater than 100, each cluster had only few vertices. Thus, those clusters were probably not distinguishable from noise. The smoothed curve was obtained by averaging nearby points (i.e. smoothing with a ‘moving average’ kernel).

Supplementary Figure 5. Preferred movie frames for putative category-selective clusters

Preferred movie frames are shown for face (A), scene (B), and body (C) clusters. The procedure for finding the preferred movie frames has been described in Methods.

Supplementary Figure 6. Further evidence for the correspondence between clusters and category-selective areas

A, The classically localized PPA (cPPA) was obtained from an average map in [Julian et al., 2012](#) (see also Figure 4C). cPPA had overlap with two patches. The posterior patch (named pPPA here) belonged to a cluster that also included a spatially distinct patch at the location of OPA. The pPPA/OPA cluster had a preference for scene frames (see Supplementary Figure 5B). The anterior patch (named aPPA here) also had a preference for scene frames (see images on the right). The black arrow indicates that pPPA and OPA are functionally correlated, as they are co-clustered. B, An average map showing fMRI activations for bodies versus faces, objects, and places. This average map was obtained from an independent group

of subjects ($n = 8$) using a mixed-effects analysis. The hot-spots of activity (denoted by asterisks) were named EBA1, EBA2 and EBA3 here. As shown in panel C, these subdivisions of EBA were arranged around area MT/V5, and they were co-localized with a cluster in map25. Part of this cluster, located within occipito-temporal sulcus (OTS), was named FBA (see [Weiner and Grill-Spector, 2013](#) for the definition of FBA based on anatomical landmarks).

Supplementary Figure 7. Blurry images do not activate the LT cluster

A, Examples of static images with relatively higher (A) and lower (B) power at low spatial frequencies (SF). These two images had the highest and lowest normalized power at low SF (< 1 cycle/degree) in our dataset. To obtain the normalized power in an image, the power spectrum of the image was converted to a 1-D plot using rotational averaging, then the low SF power was divided by the total power. C, The comparison between low SF (yellow) and high SF (cyan) images shows an fMRI activation for low SF (blurry) images outside of the LT cluster.

Supplementary Figure 8. Cross-validation analyses

A-B, map25 in right (A) vs. left (B) hemispheres. The map in panel A is also shown in Figure 3C. The black arrow indicates the pSTS face cluster in the right hemisphere map. This cluster is absent in the left hemisphere map, at this level of hierarchical clustering. C-D, map25 in an average data from odd-labeled subjects (subjects # 1, 3, 5) (C) vs. even-labeled subjects (subjects # 2, 4) (D). E-F, map25 in an average data from odd functional runs (E) vs. even functional runs (F).

Supplementary Figure 9. Quantitative comparison between hierarchical clusterings

For 100 levels of hierarchical clustering in right and left hemispheres, the clustering similarity was computed for odd/even runs and odd/even subjects (see Methods for the formula). The values of analytic chance and standard deviations associated with those values were also calculated; standard deviations were too small to be shown. Additionally, the clustering similarity was computed for pairs of

simulated data in which the real functional activities in occipito-temporal cortex were replaced by Gaussian white noise. Before clustering, the random noise data were spatially smoothed using a Gaussian filter ($\text{FWHM} = \sqrt{3^2 + 4^2} = 5 \text{ mm}$), considering the spatial smoothing of 3 mm used in real data and a hemodynamic point spread function of 4 mm (Engel et al., 1997). The shaded area around the black curves indicates one standard deviation, calculated based on 100 simulations.

Supplementary Figure 10. The parcellation map at the level of 80 clusters

A, The map80 shows fine-grained layout of areas in the right occipito-temporal cortex, with clusters that are small in size (see Supplementary Figure 4B). One of the clusters was located within aSTS (anterior superior temporal sulcus), in a region which appeared to have a distinct representation in the PCA map. B-C, Preferred (B) and non-preferred (C) movie frames for the aSTS cluster. This cluster evidently had a preference for images of faces/people.

Supplementary Figure 11. The location of area MT on the parcellation maps

A, Motion-selective areas MT (middle temporal) and OMA (occipital motion area) were localized in one subject based on a block-design comparison between moving dots and static dots. The stimuli were full-screen black/white random dot patterns, either moving coherently or jittering. The motion direction (for moving dots) and the position of dots (for static dots) changed every second throughout the blocks. B-C, A cluster in map25 (B) and its subdivisions in map80 (C). One subdivision spatially corresponded to area MT.

Supplementary Figure 12. An improved functional-anatomical registration in the monkey scans

A single horizontal slice from an anatomical (MP-RAGE) scan is shown on the left. The corresponding slice from a functional (GE-EPI) scan is shown on the right. The pial surface is indicated with a red outline. The good match/registration between functional and anatomical images was obtained through

1) spatial undistorting of EPI volumes, and 2) an improved registration algorithm using a boundary-based cost function ('bbregister' in Freesurfer).

Supplementary Figure 13. Maps of the first principal component scores (A) and SNR (B) on the monkey flat patch

See legend of Supplementary Figure 3 for details about computing the SNR values.

Supplementary Figure 14. Preferred movie frames for cluster A (panel A) and cluster B (panel B)

Clusters A and B are shown on the flat patch in Figure 7C.

Supplementary Figure 15. Analysis of color selectivity in cluster B

The movie frames, sampled every two seconds, were sorted based on the magnitude of response elicited by those frames in cluster B. ~ 2% of the frames (55 frames, out of 2710 non-blank frames) were chosen as preferred and non-preferred images. The color content of each (natural) image was estimated by a following procedure. For each pixel of the image, RGB values were converted to CIE then LMS values, where L was the long-wavelength cone response, M was the medium-wavelength cone response, and S was the short-wavelength cone response (see [Murphey et al., 2008](#) for the conversion formulas). Next, the normalized cone contrast was calculated using the equations $L = (L - L_0)/L_0$, $M = (M - M_0)/M_0$, and $S = (S - S_0)/S$, where (L_0, M_0, S_0) was the color of an achromatic gray background (the mean color of all pixels of all movie frames). Two orthogonal color axes were created: $(L - M)$ for long-to medium-wavelength color contrast, and $(L + M) - S$ for short-wavelength color contrast. Such color space, schematically demonstrated in panel A, is called DKL color space ([Derrington et al., 1984](#)). The colors of daylight are predominantly found along 'daylight axis'. Each image was represented as a point in the color space after averaging color contrast values across all pixels of the image (panels B-C). The red lines are linear regression fits. The color contrast values in preferred images were not significantly different from those in non-preferred images ($p > 0.05$, t-test). The color content of the images was also

evaluated using a 'colorfulness metric' described in [Hasler and Süssstrunk, 2003](#). The preferred and non-preferred images were not significantly different from each other in terms of average colorfulness ($p > 0.05$, t-test).

Supplementary Figure 16. A functional network involved in the processing of animate stimuli

The map shows the comparison between animate versus inanimate movie frames, on the lateral view of inflated surface. Only fMRI activations for animate stimuli are displayed (one-sided map). Sulcal abbreviations: STS, superior temporal; PMTS, posterior middle temporal; MTS, middle temporal; AMTS, anterior middle temporal.

Supplementary Figure 17. Comparison between human and monkey parcellation maps

A, The correlation matrix demonstrating the functional correspondence between clusters in map25 (X-axis) and clusters in map25m (Y-axis). For each cluster, the mean time-course of response was computed by averaging time-courses of all vertices in that cluster. Then, the temporal correlation was calculated for all pairs of human-monkey clusters. The black circles indicate pairs with high correlation values. In the colorbar, positive correlations are yellow -> red, and negative correlations are light blue -> dark blue. B, The significance matrix at the threshold of Bonferroni-corrected p-value ($p < 0.05/625$; $p < 0.00008$). The significant pairs are shown black. C, Example clusters in human map25 and their corresponding clusters in monkey map25m. The clusters are displayed on the flat patches. Intra-species correlation matrices (CMs) are shown in the rightmost panels (human: top panel, monkey: bottom panel).

Supplementary Figure 18. Data-driven framework for functional parcellation

References for Supplementary Figures:

1. Derrington, A.M., Krauskopf, J., and Lennie, P. (1984). Chromatic mechanisms in lateral

geniculate nucleus of macaque. *The Journal of physiology* 357, 241-265.

2. Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral cortex* 7, 181-192.
3. Hasler, D., and Susstrunk, S.E. (2003). Measuring colorfulness in natural images. *Proc. SPIE 5007, Human Vision and Electronic Imaging VIII*, 87-95.
4. Julian, J.B., Fedorenko, E., Webster, J., and Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage* 60, 2357-2364.
5. Murphey, D.K., Yoshor, D., and Beauchamp, M.S. (2008). Perception matches selectivity in the human anterior color center. *Current biology : CB* 18, 216-220.
6. Weiner, K.S., and Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychological research* 77, 74-97.

Figure 1

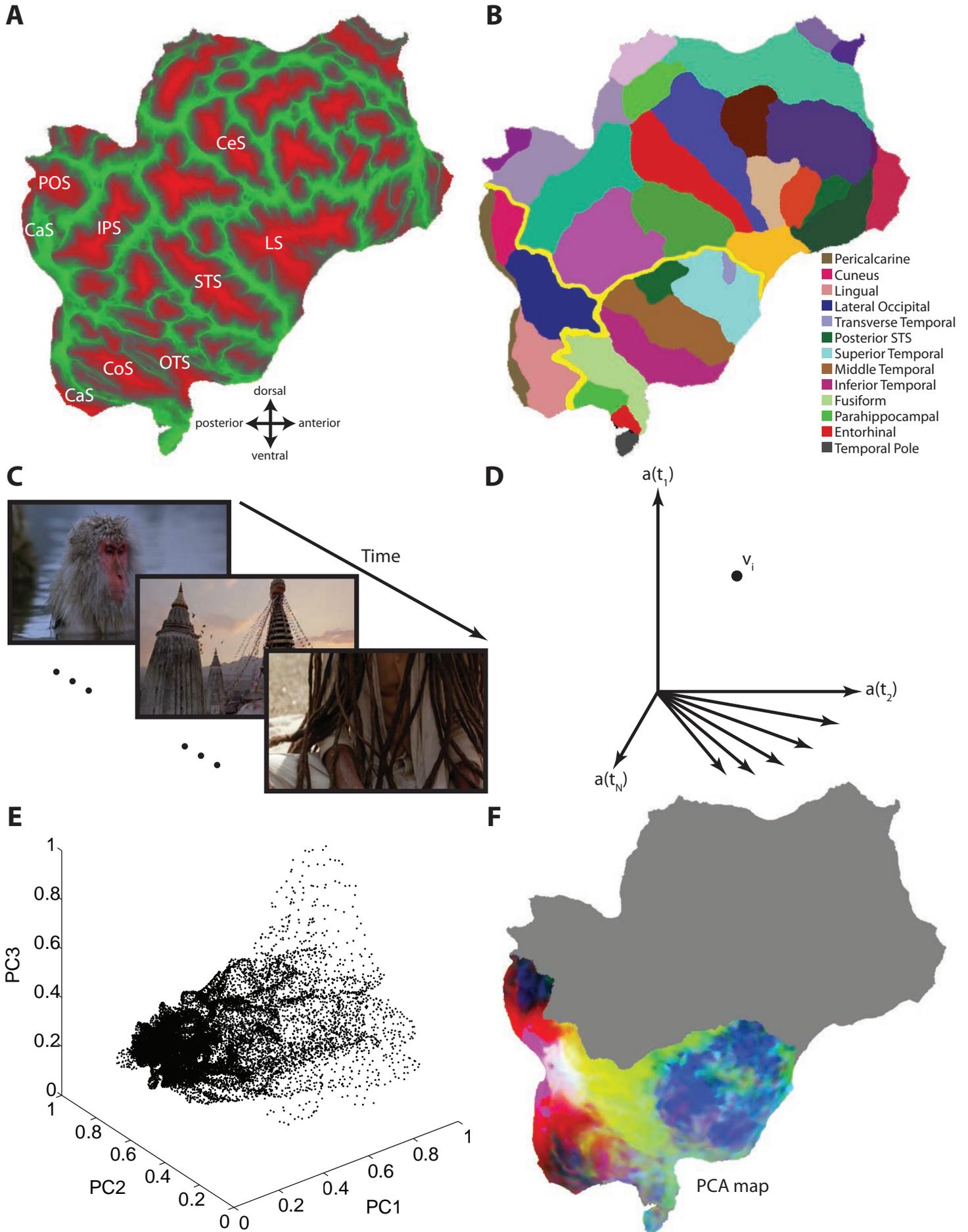


Figure 2

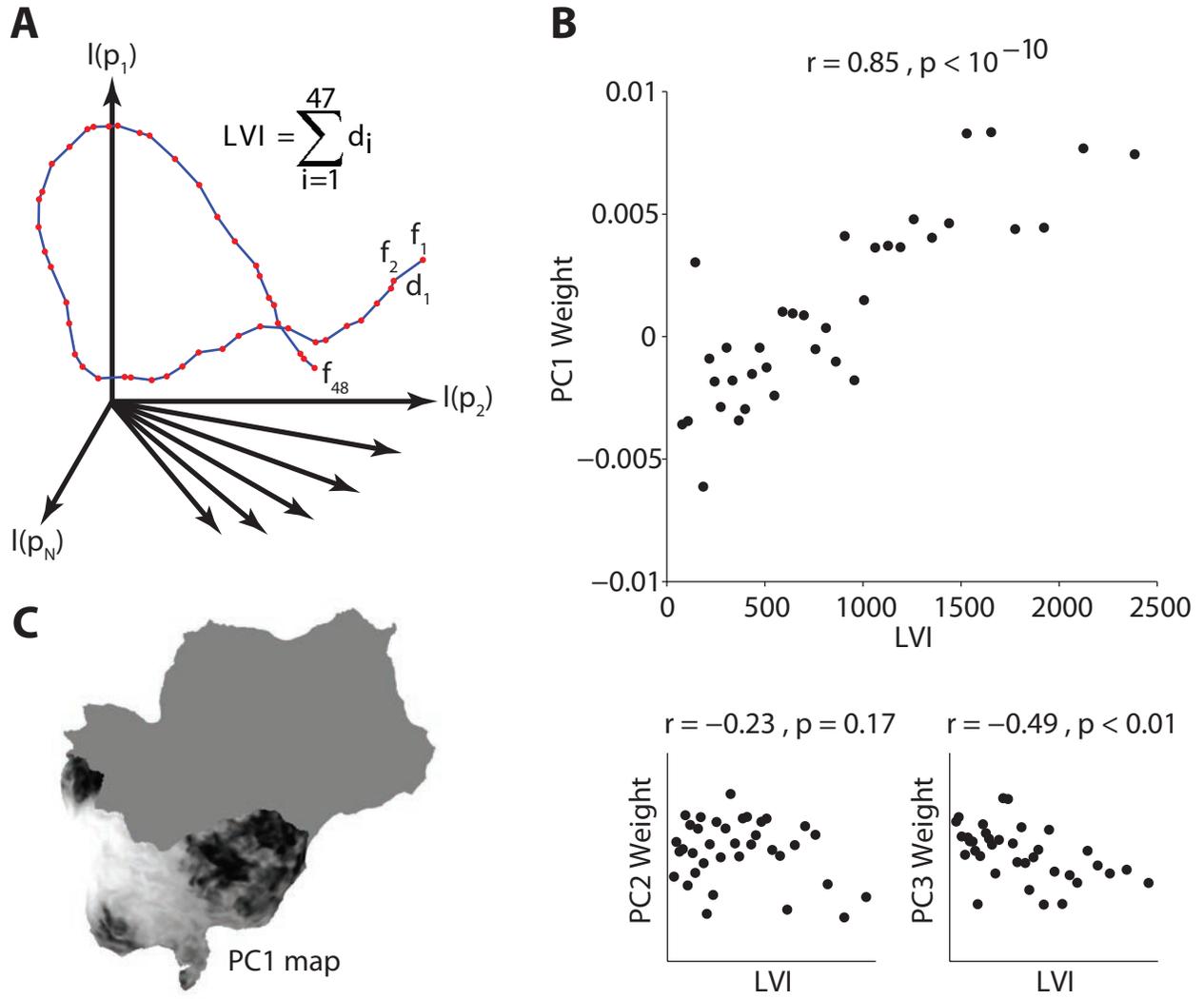


Figure 3

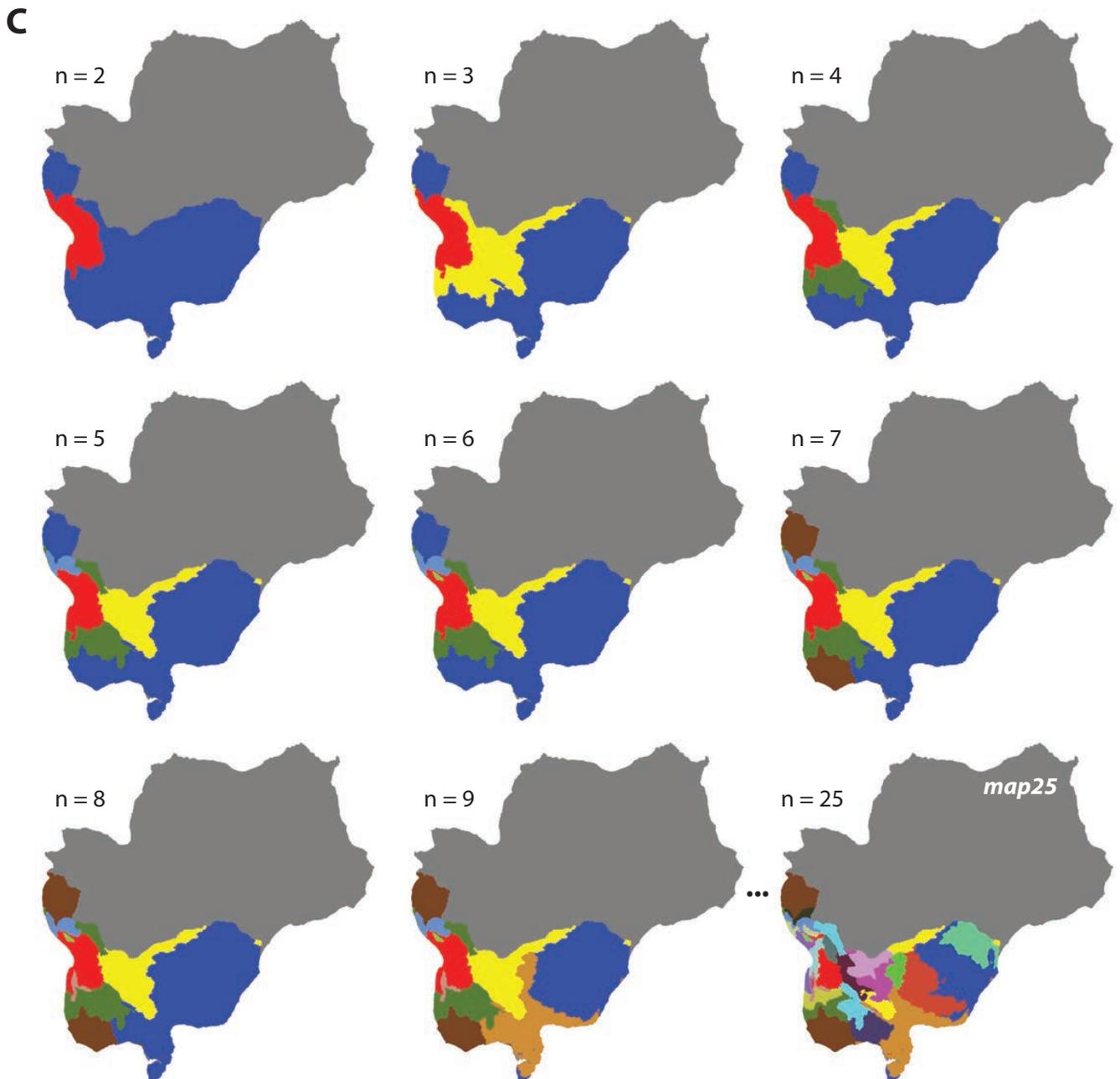
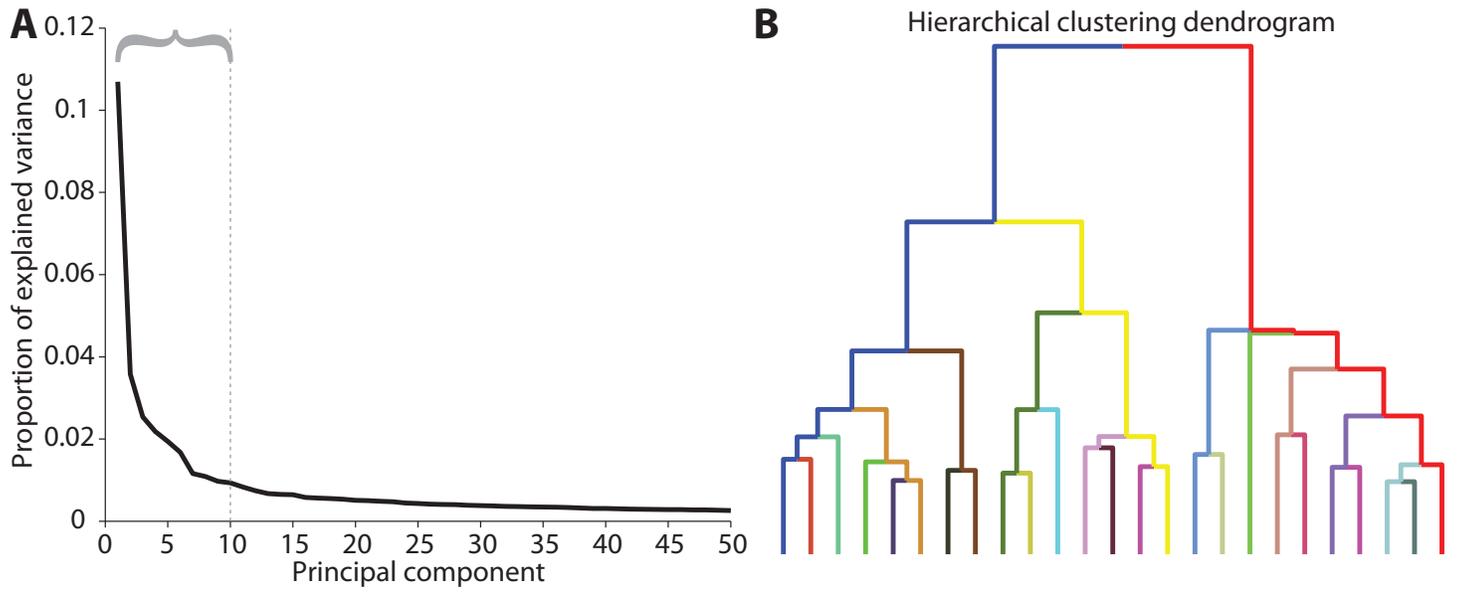


Figure 4

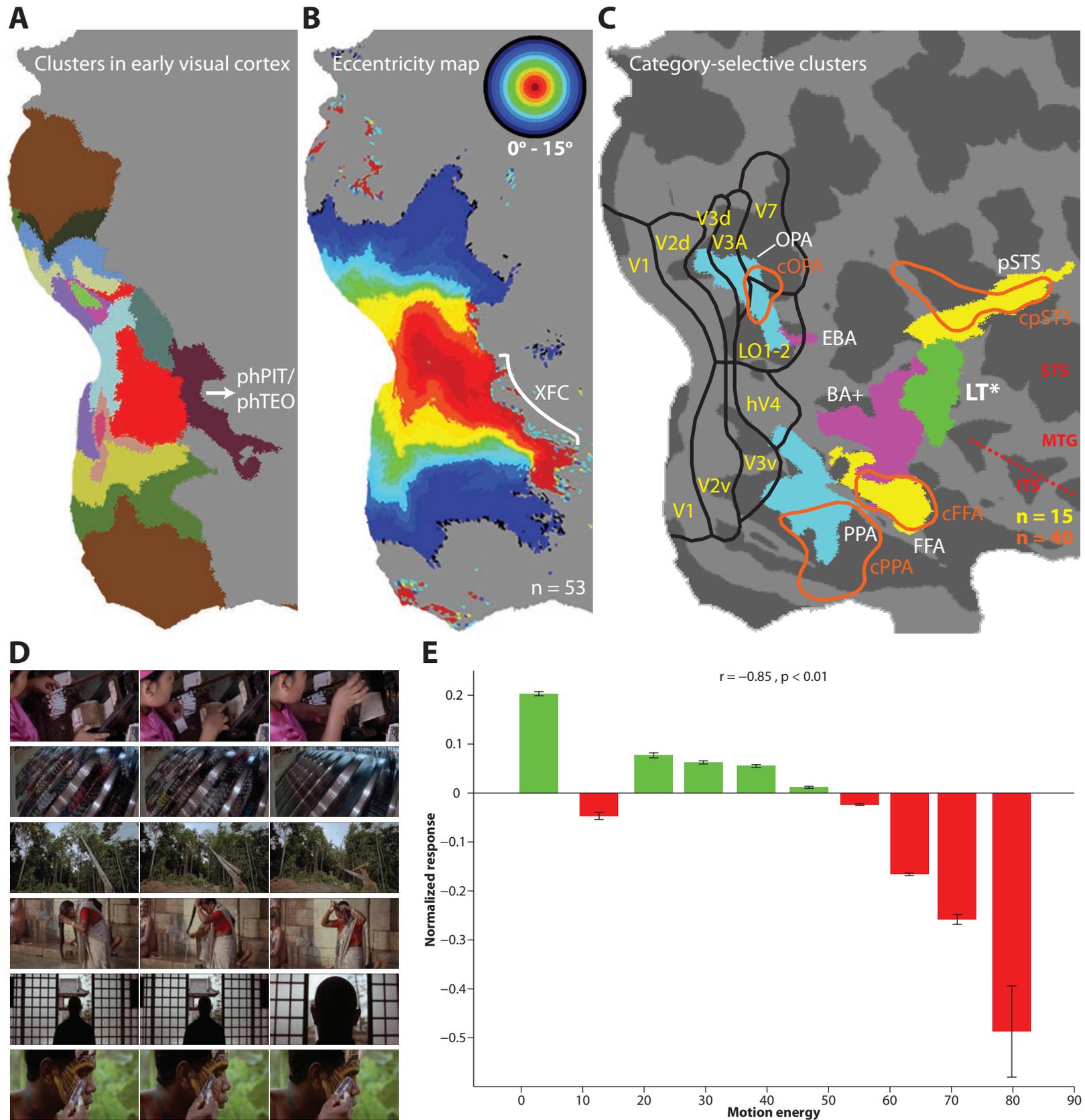


Figure 5

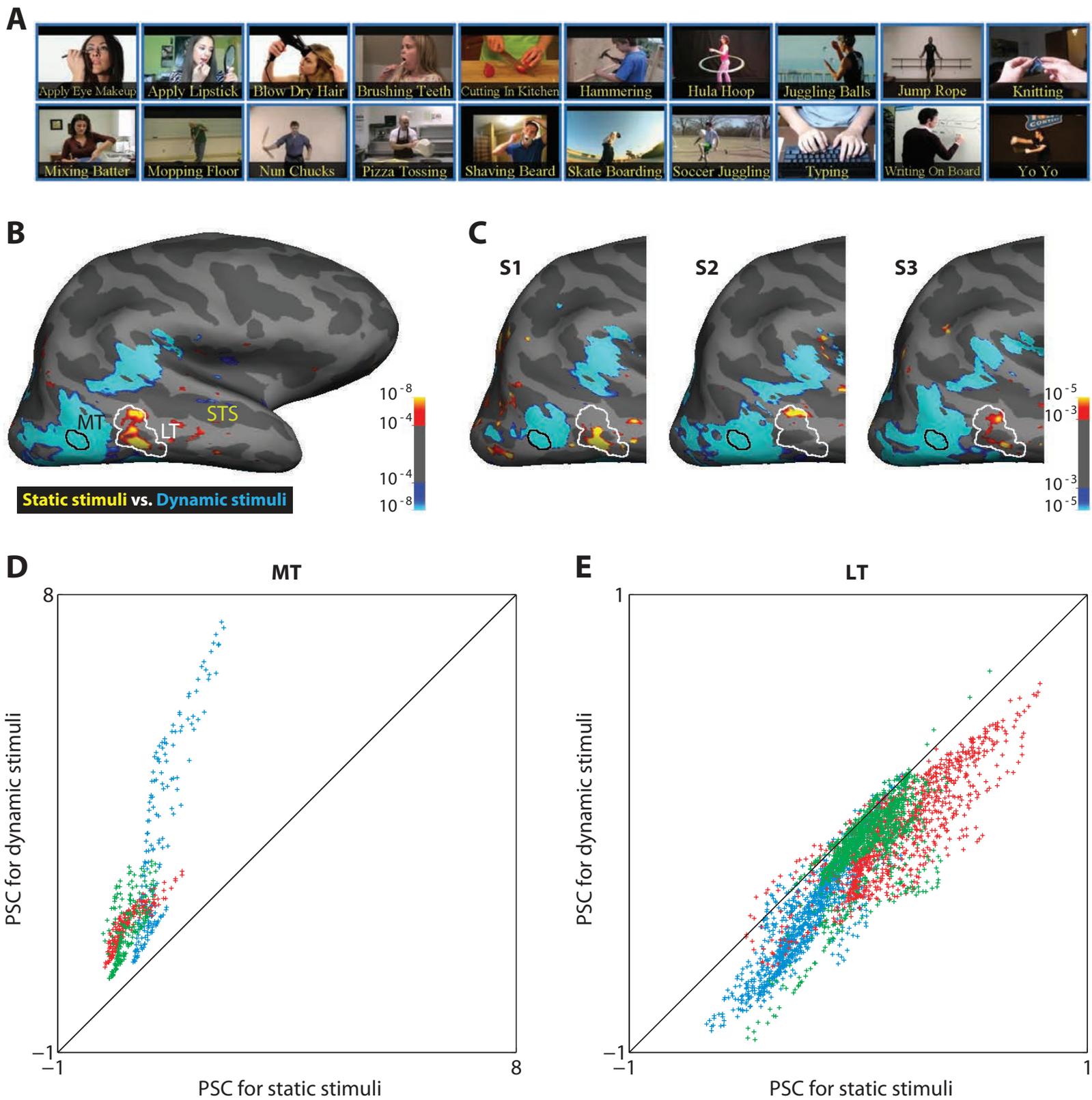


Figure 6

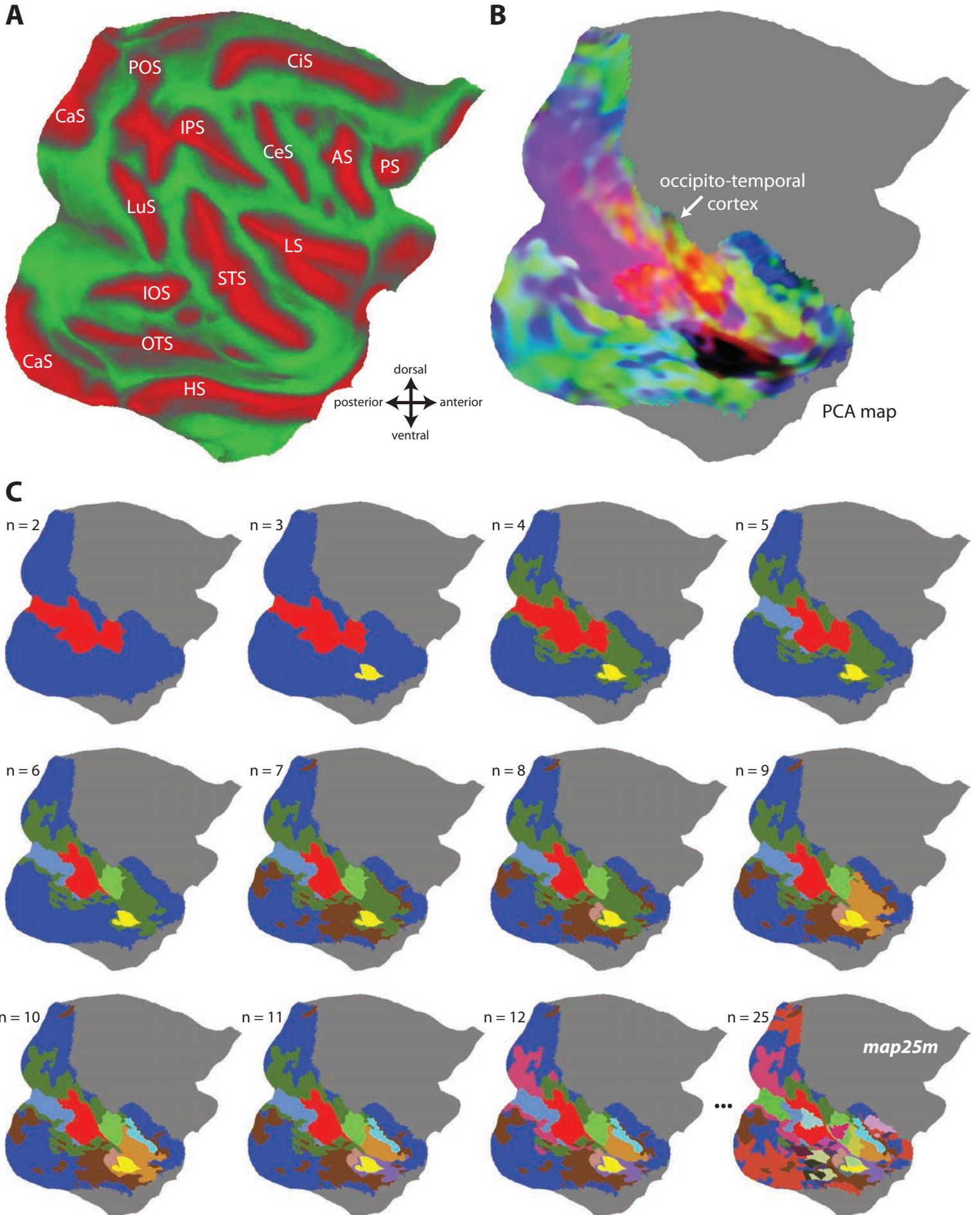


Figure 7

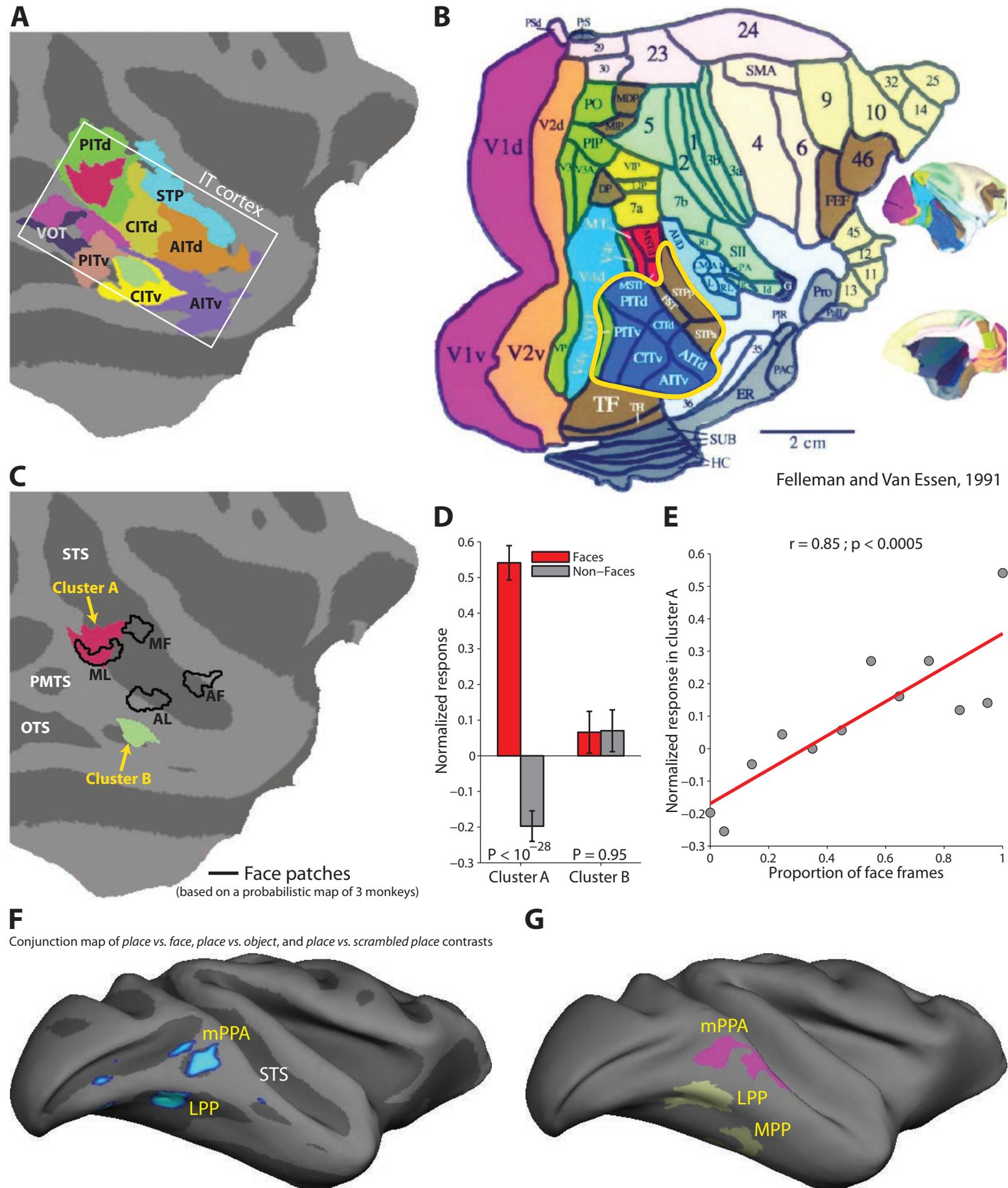
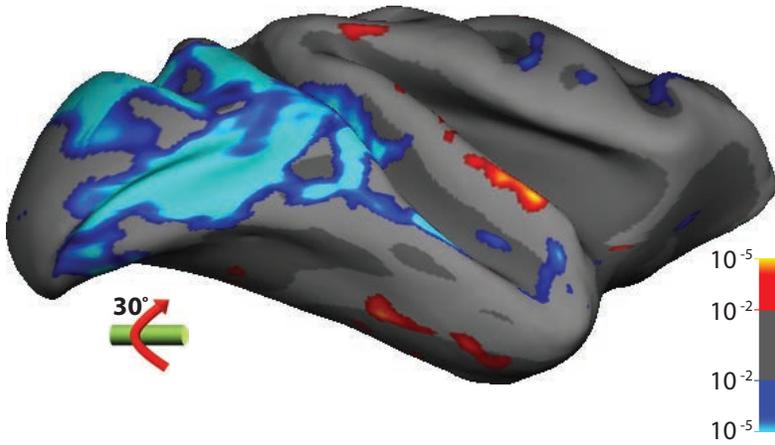
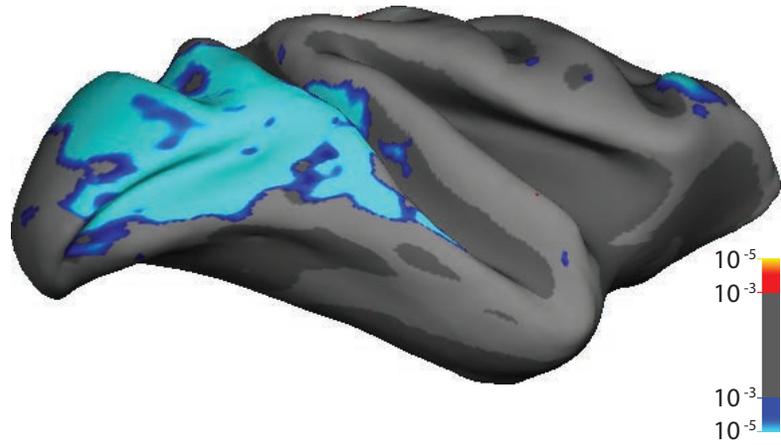


Figure 8

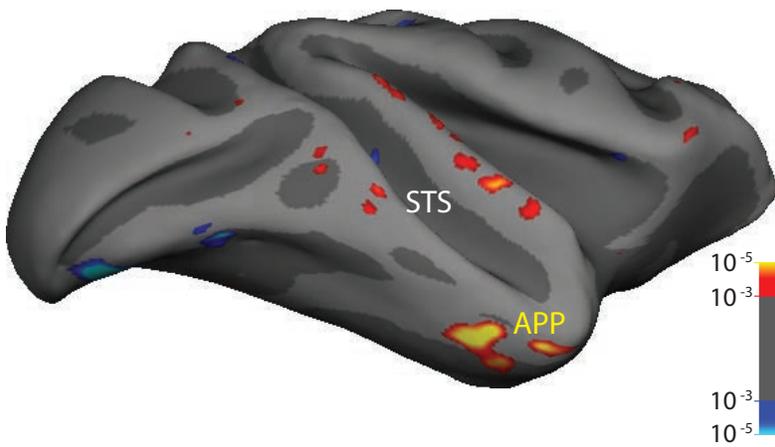
A Outdoor scenes vs. Indoor scenes



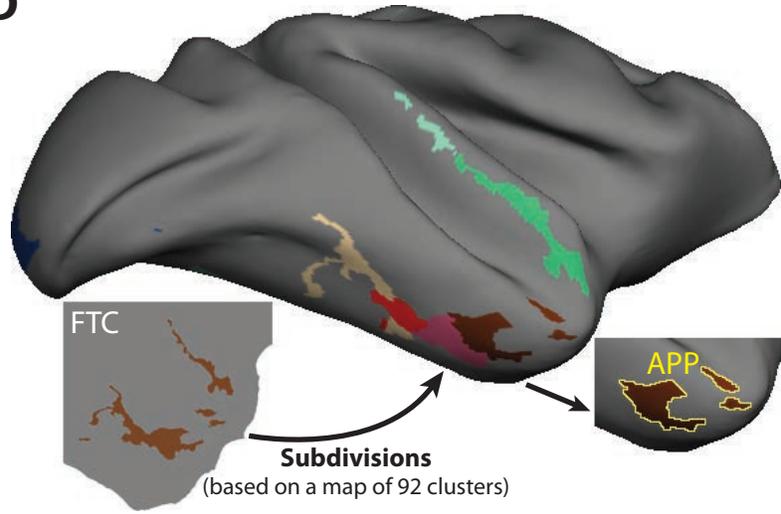
B Outdoor+inanimate scenes vs. Indoor+inanimate scenes



C Outdoor+animate scenes vs. Indoor+animate scenes



D



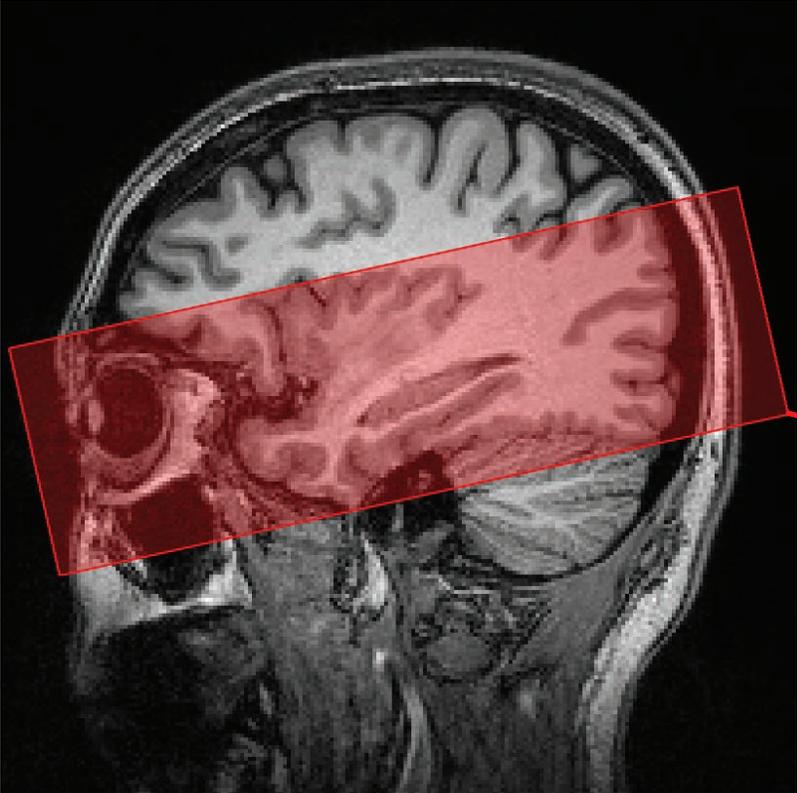
E



F

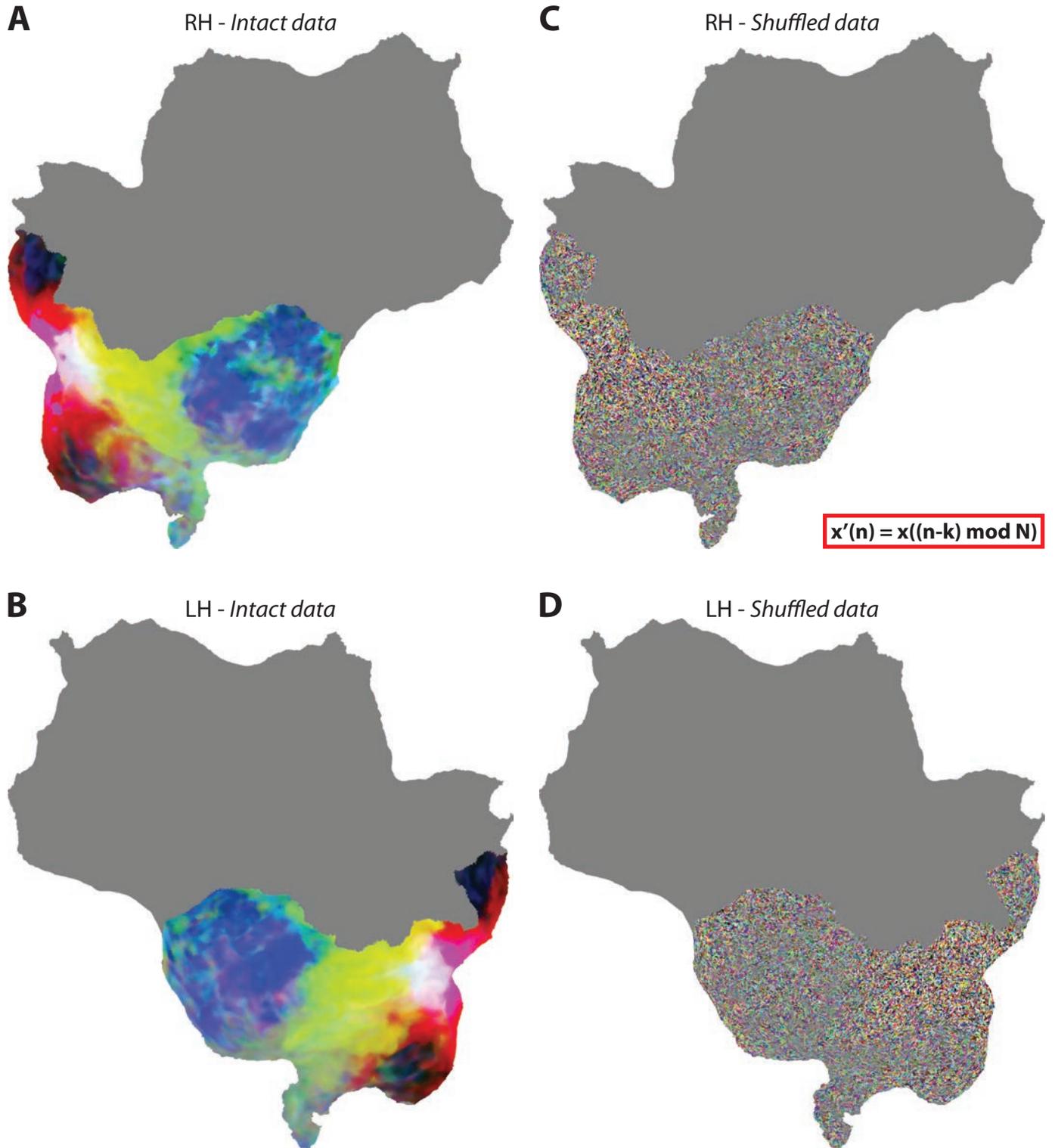


SFigure 1

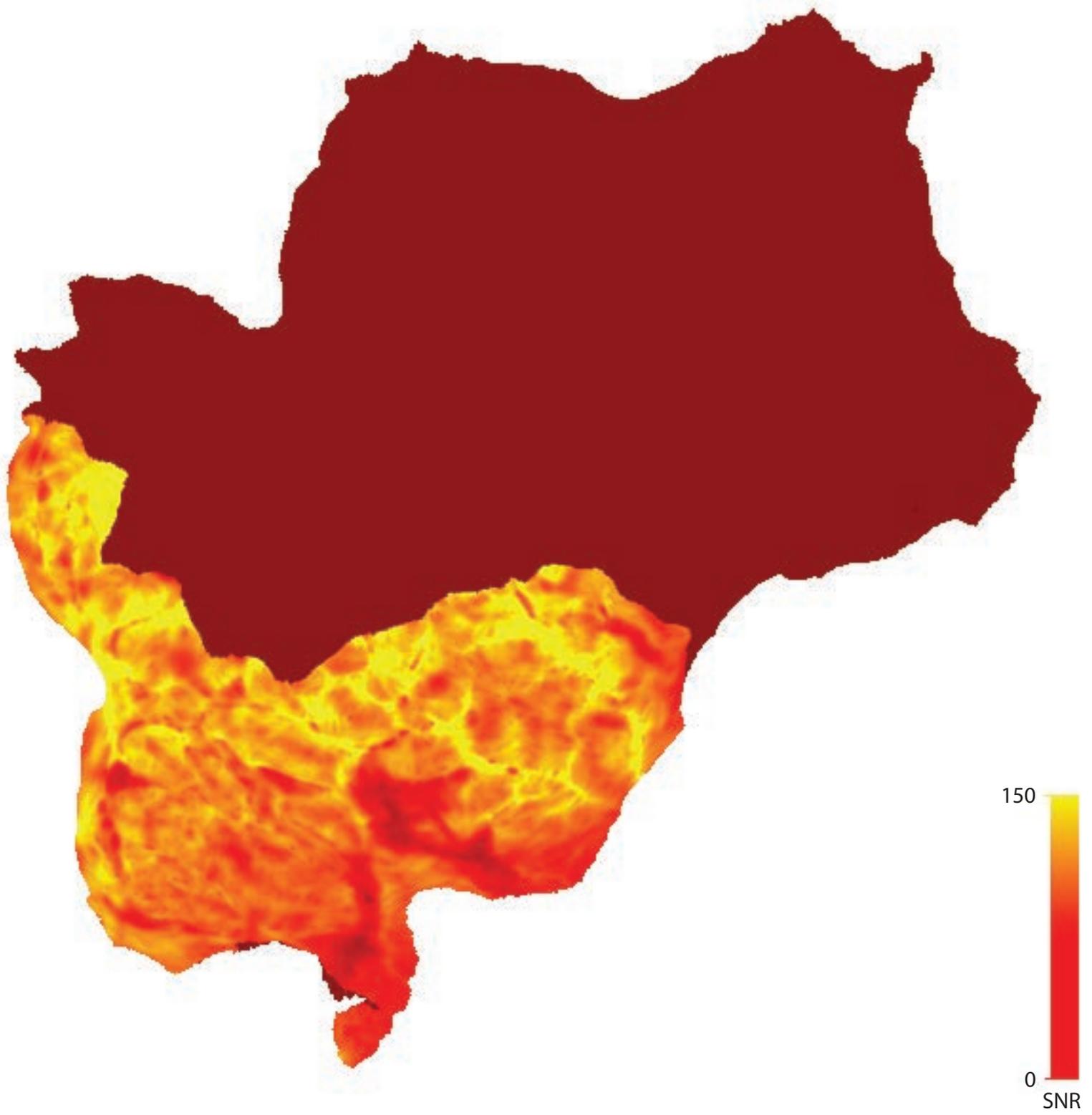


→ 30 EPI slices covering the occipito-temporal cortex

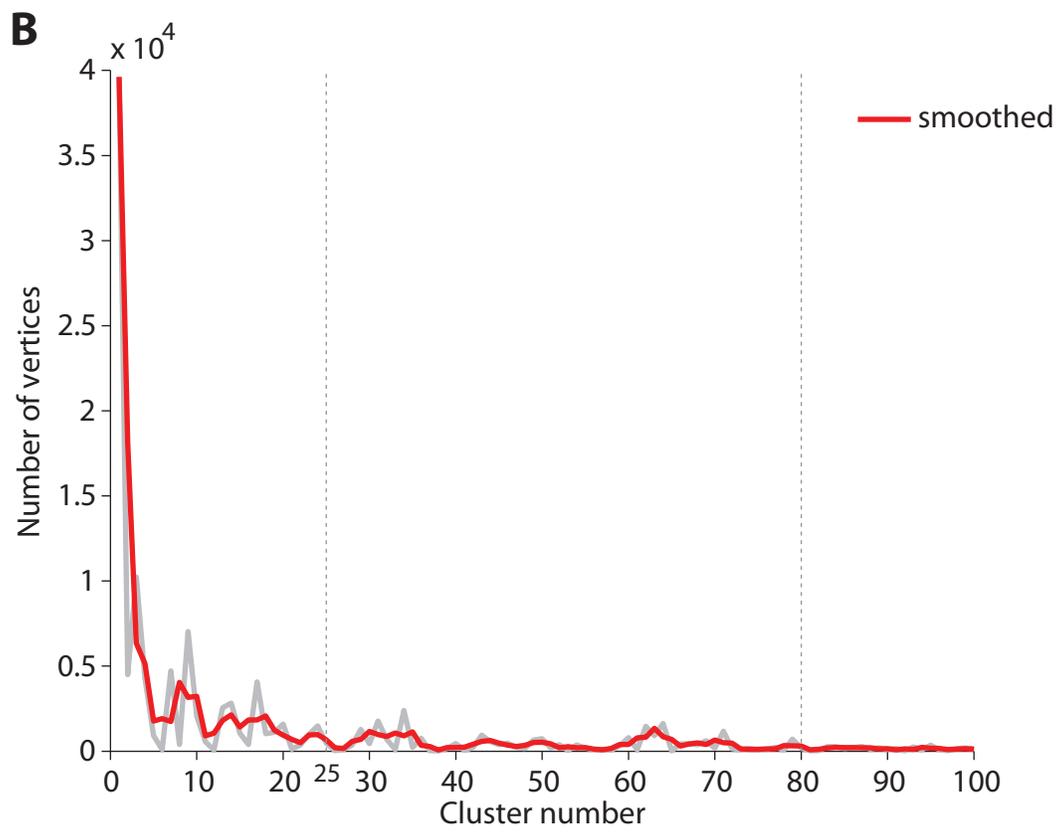
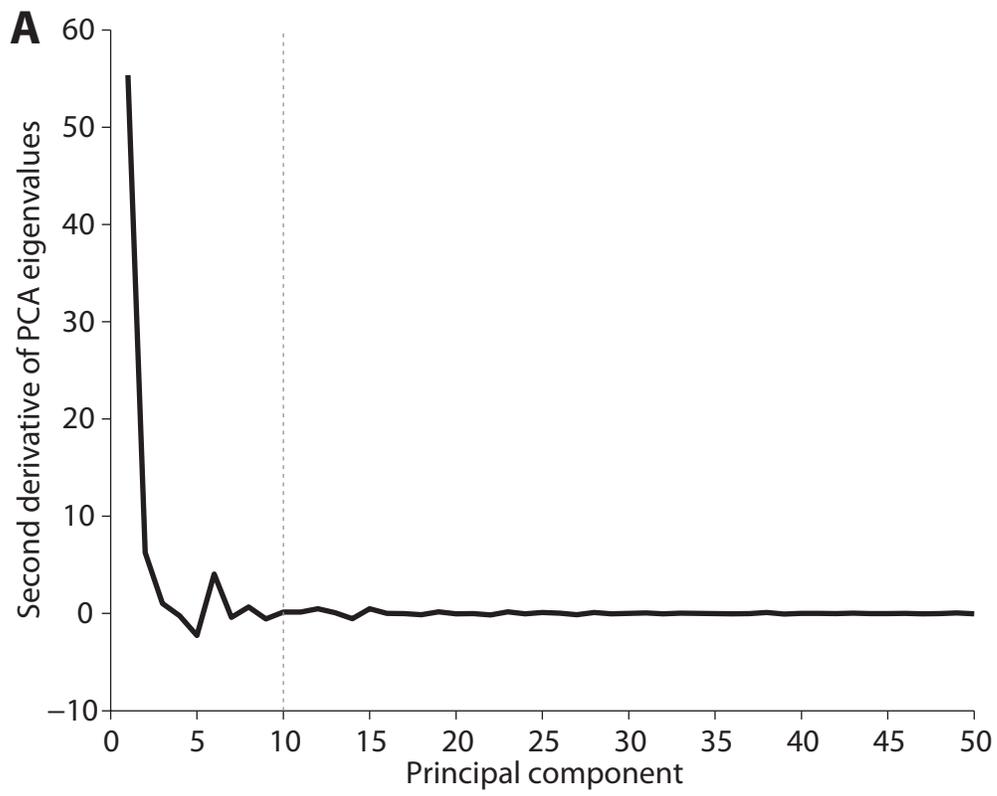
SFigure 2



SFigure 3



SFigure 4

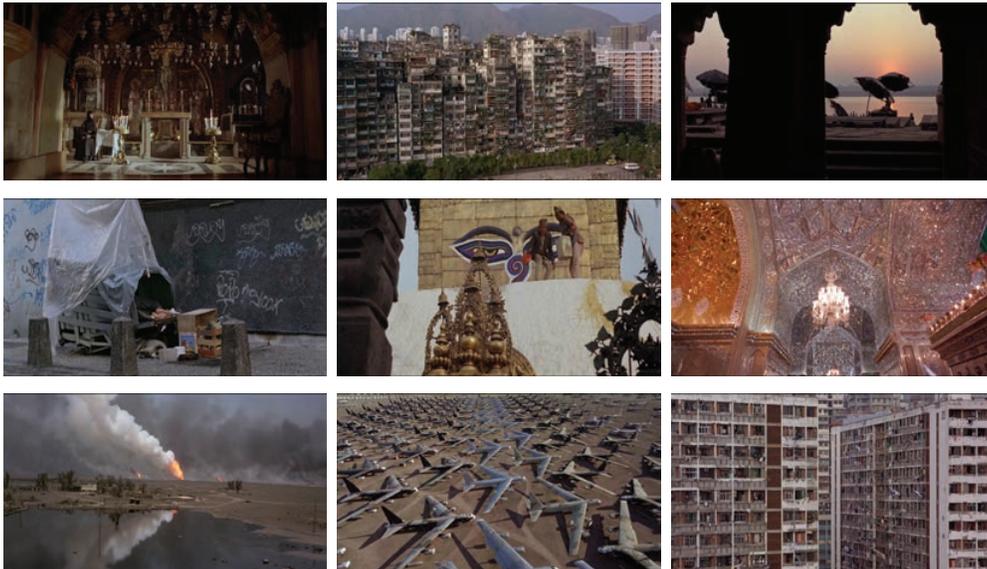


SFigure 5

A



B



C

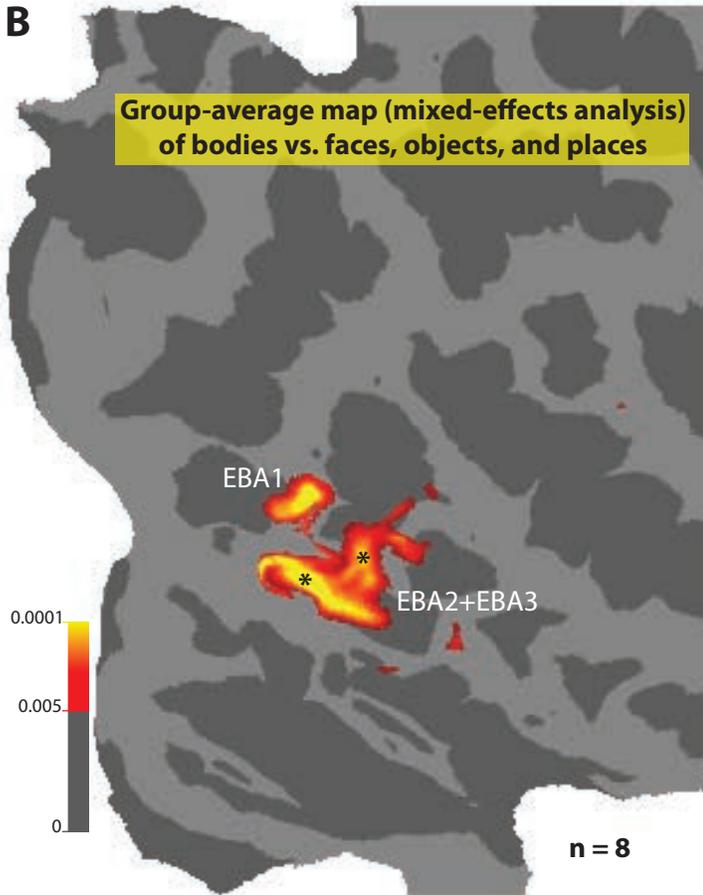


SFigure 6

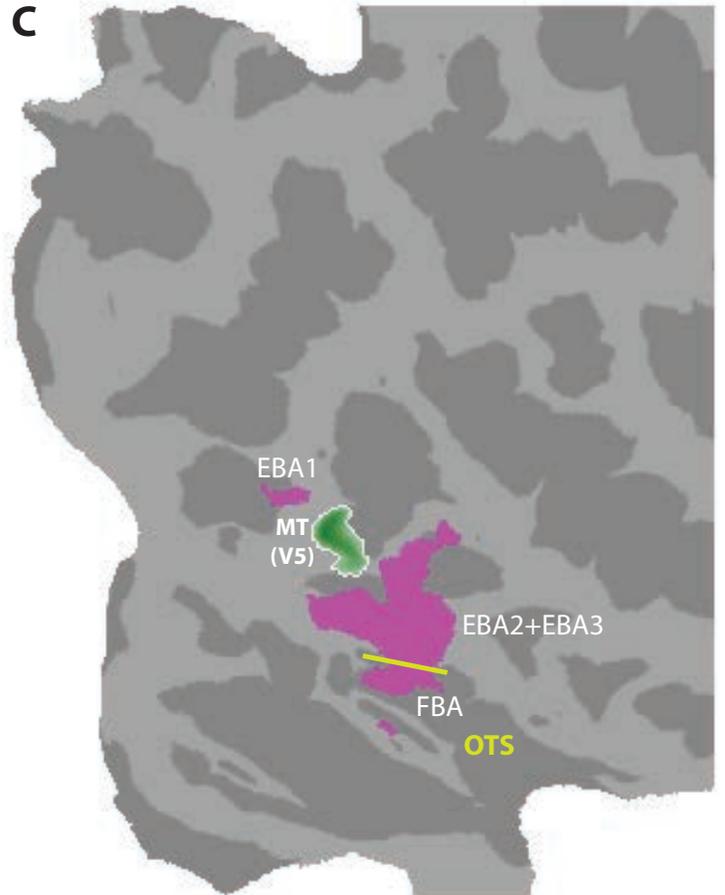
A



B



C

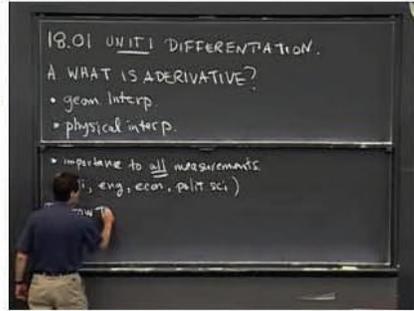


SFigure 7

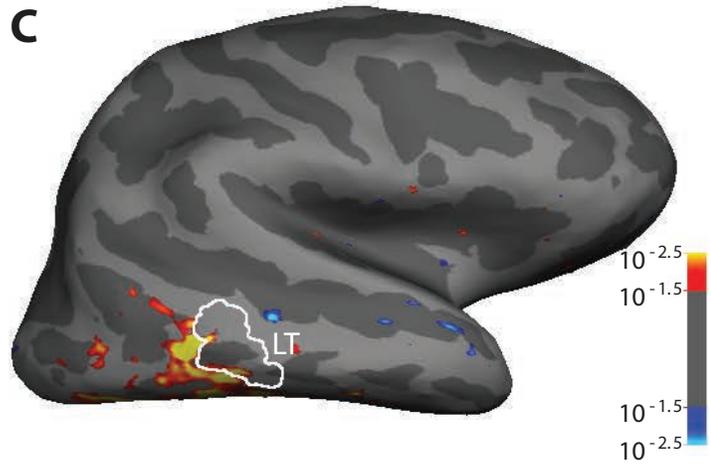
A



B

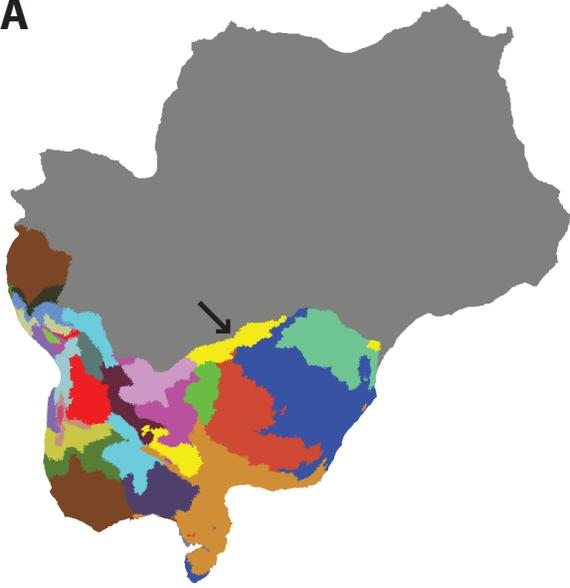


C



SFigure 8

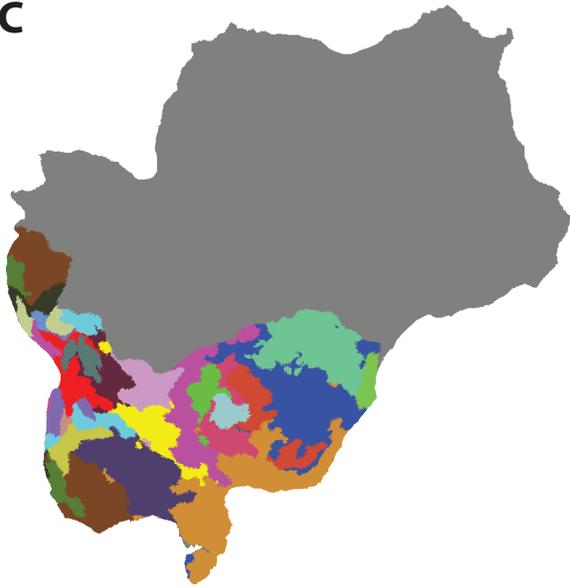
A



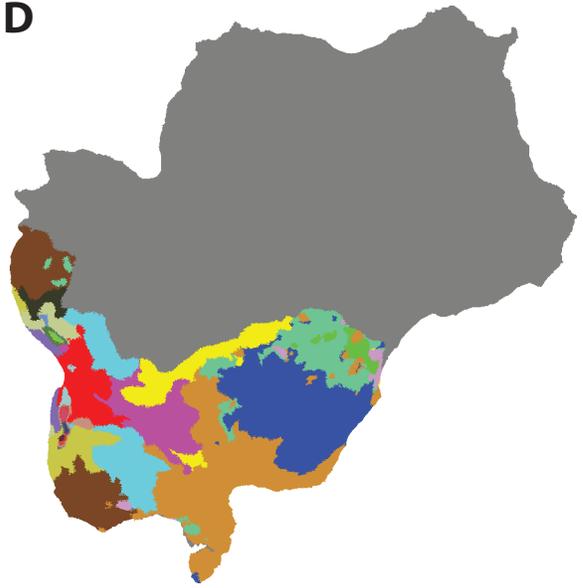
B



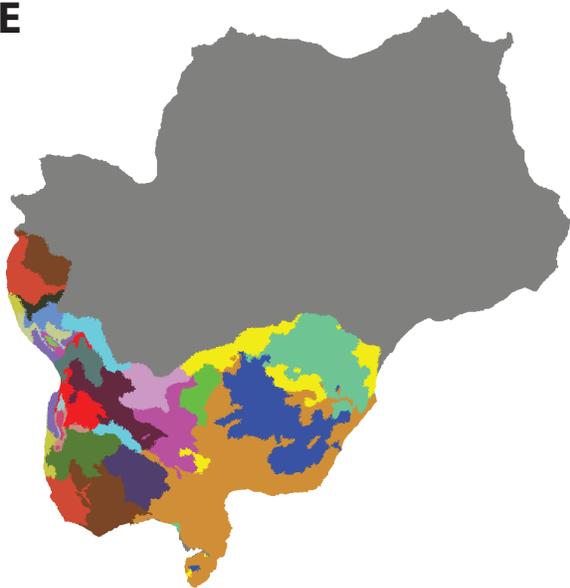
C



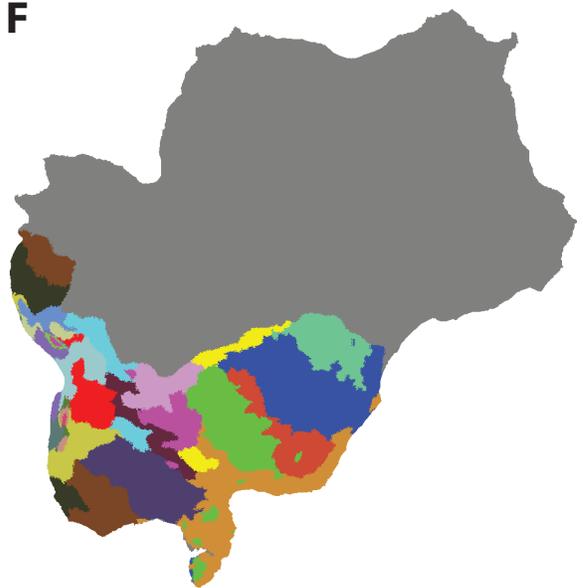
D



E

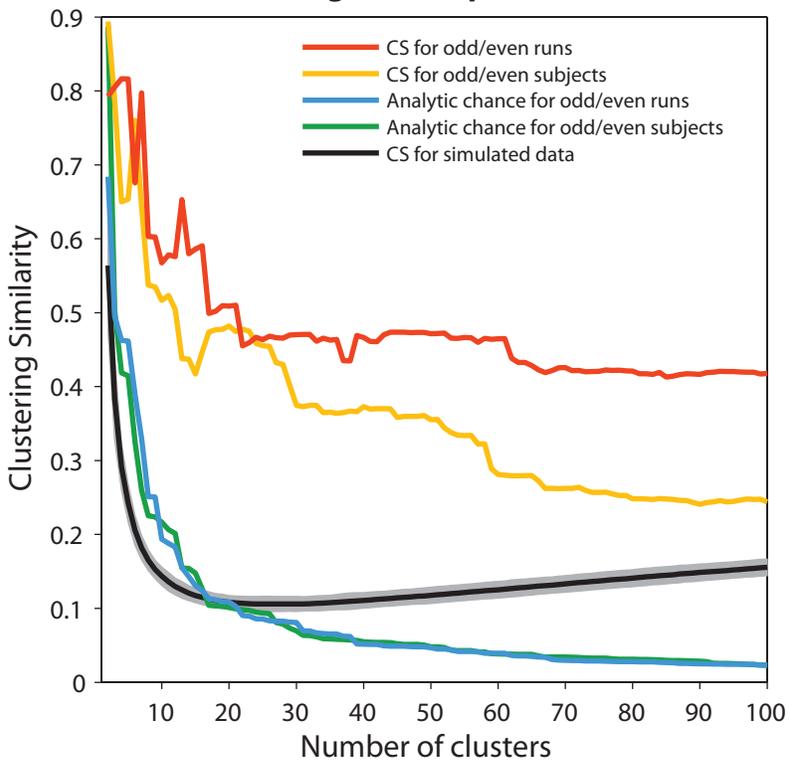


F

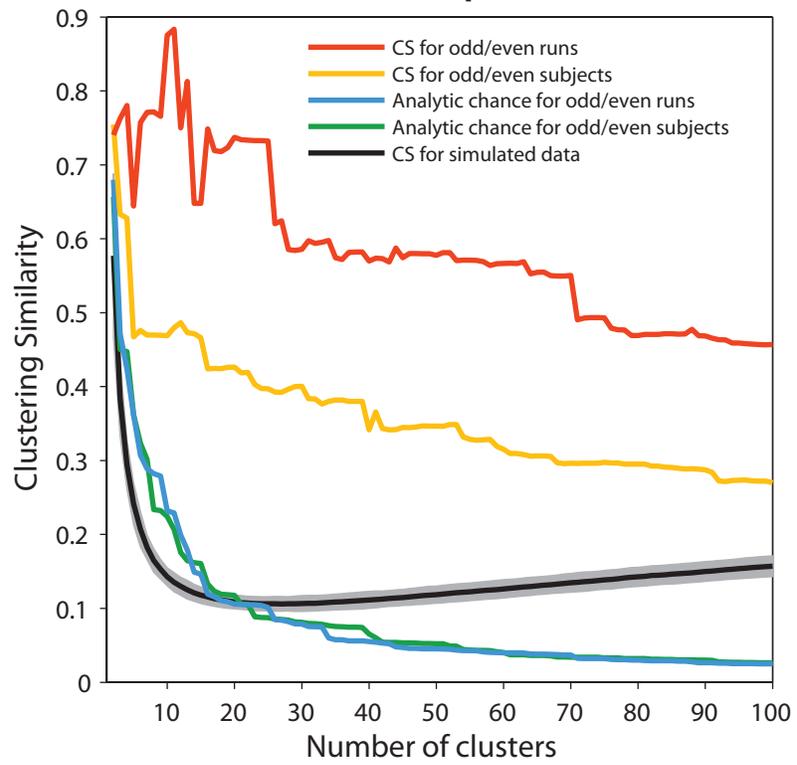


SFigure 9

Right hemisphere



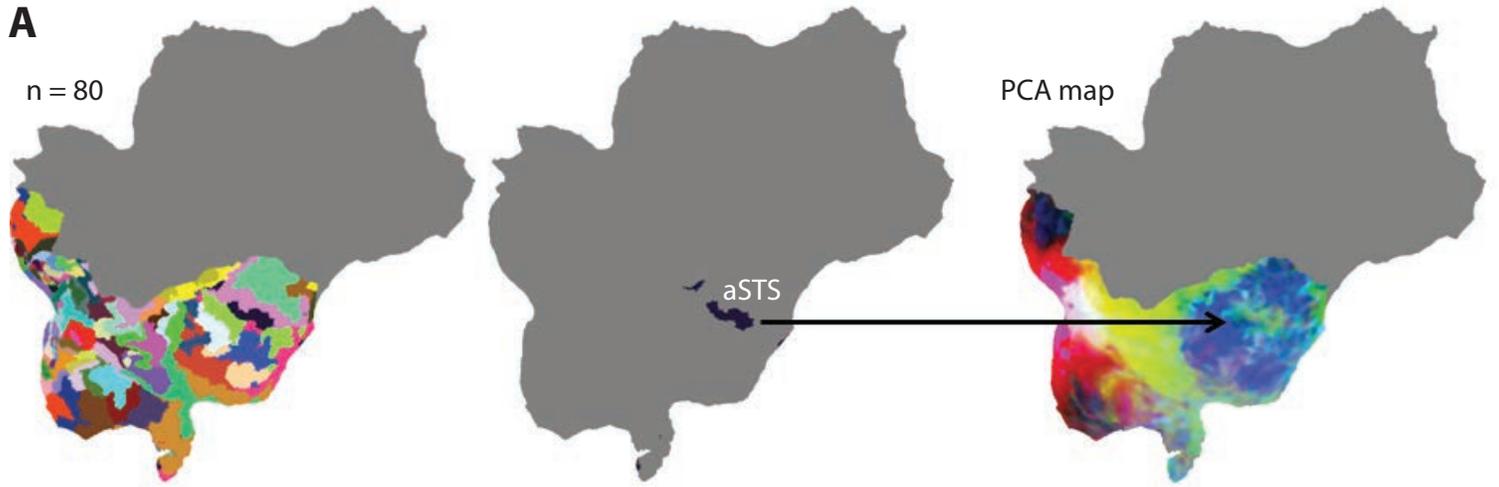
Left hemisphere



SFigure 10

A

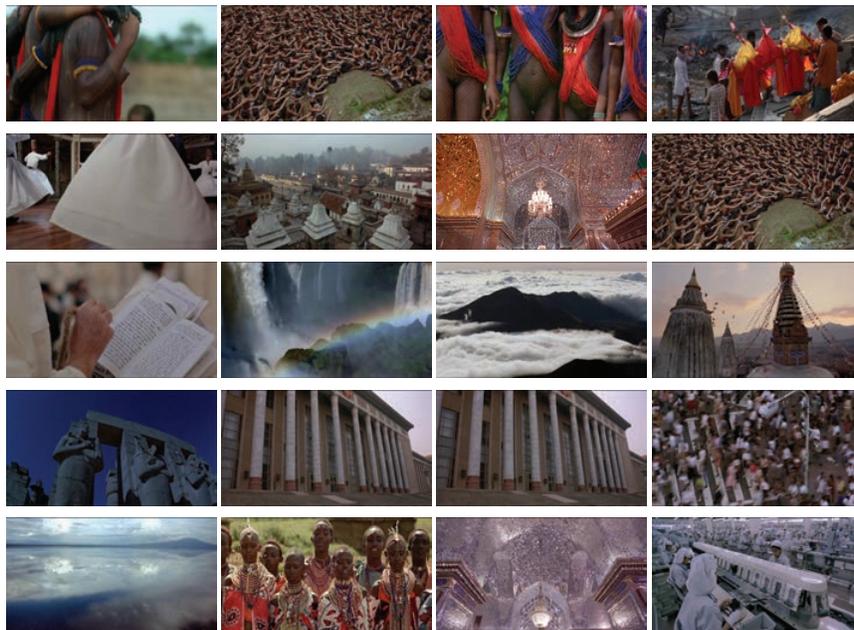
n = 80



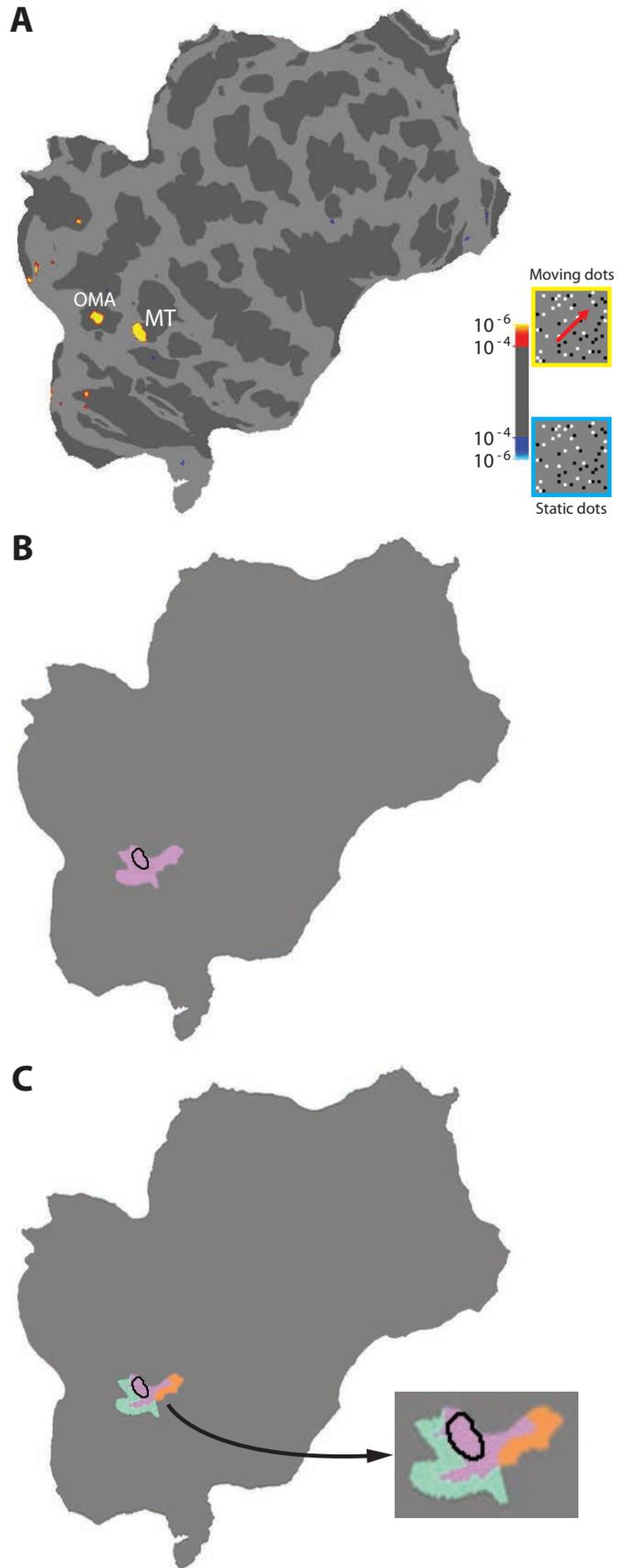
B



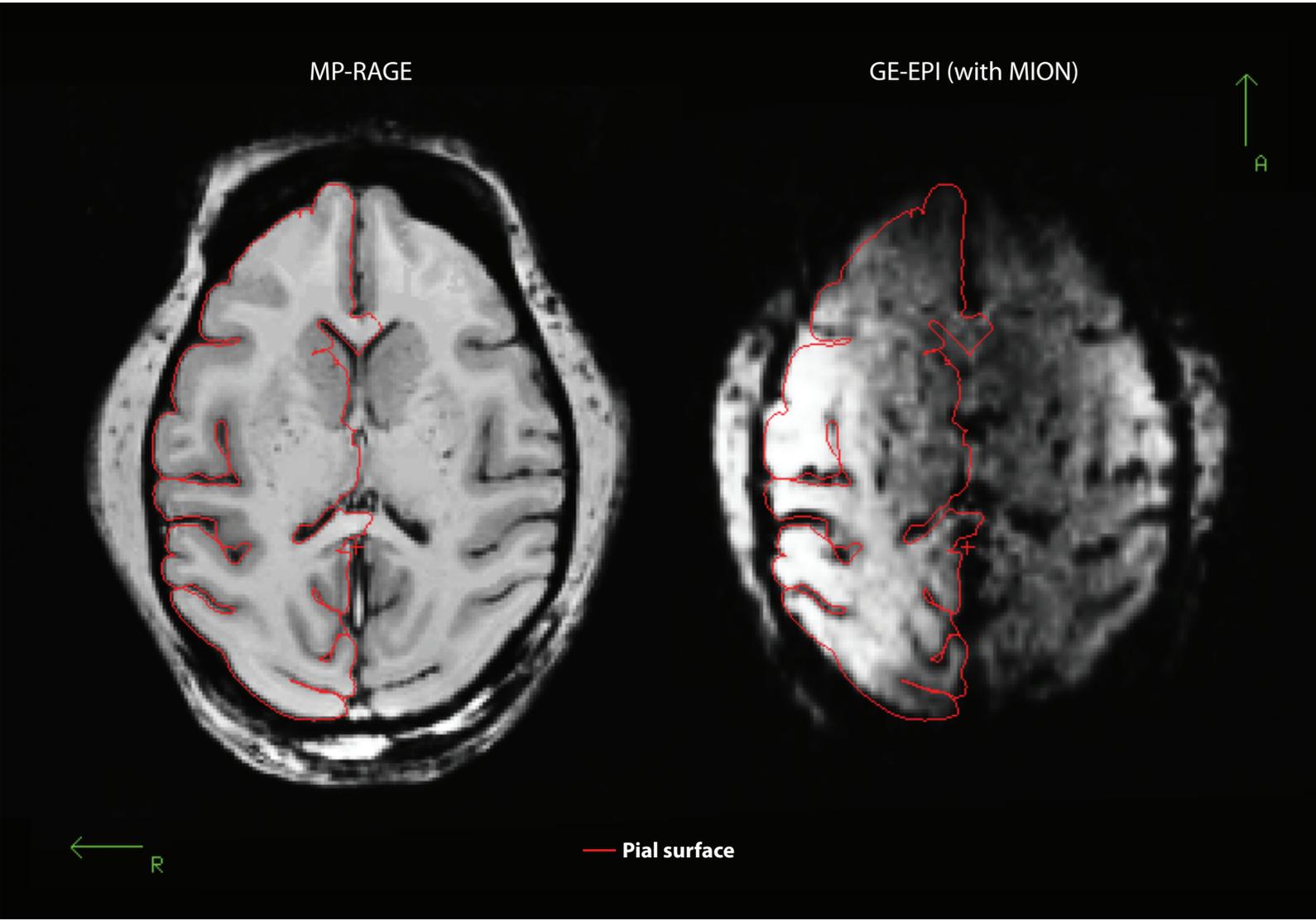
C



SFigure 11

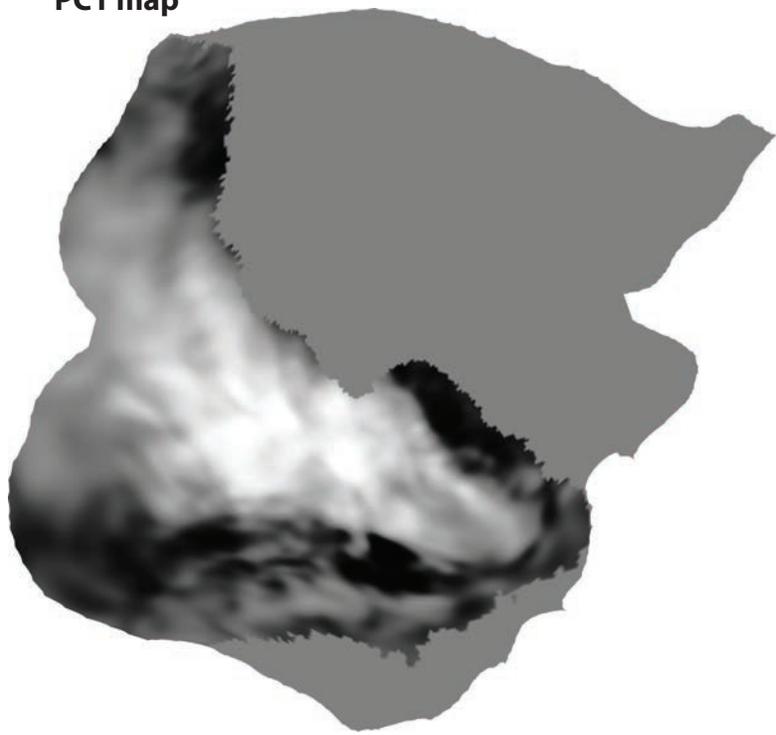


SFigure 12

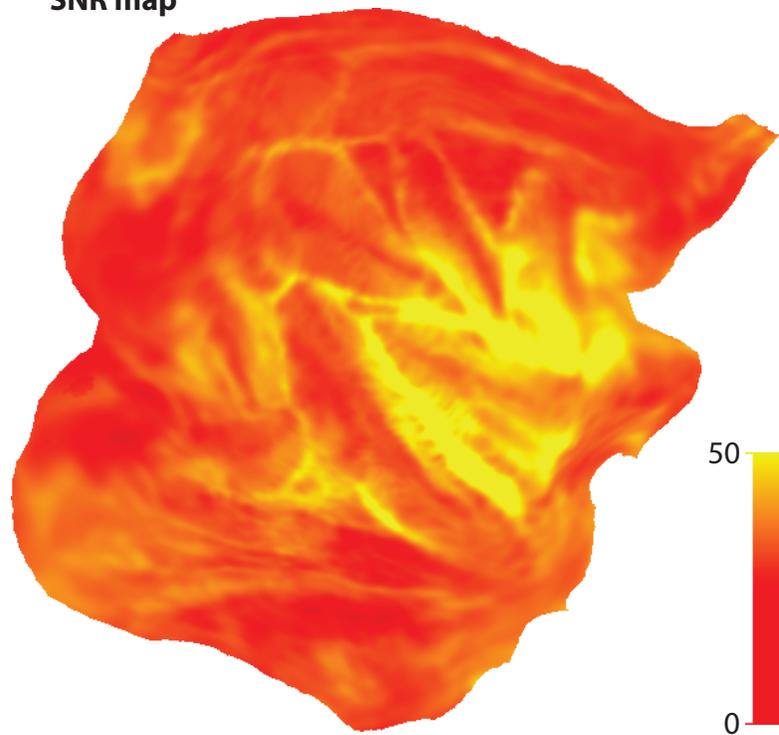


SFigure 13

A PC1 map



B SNR map

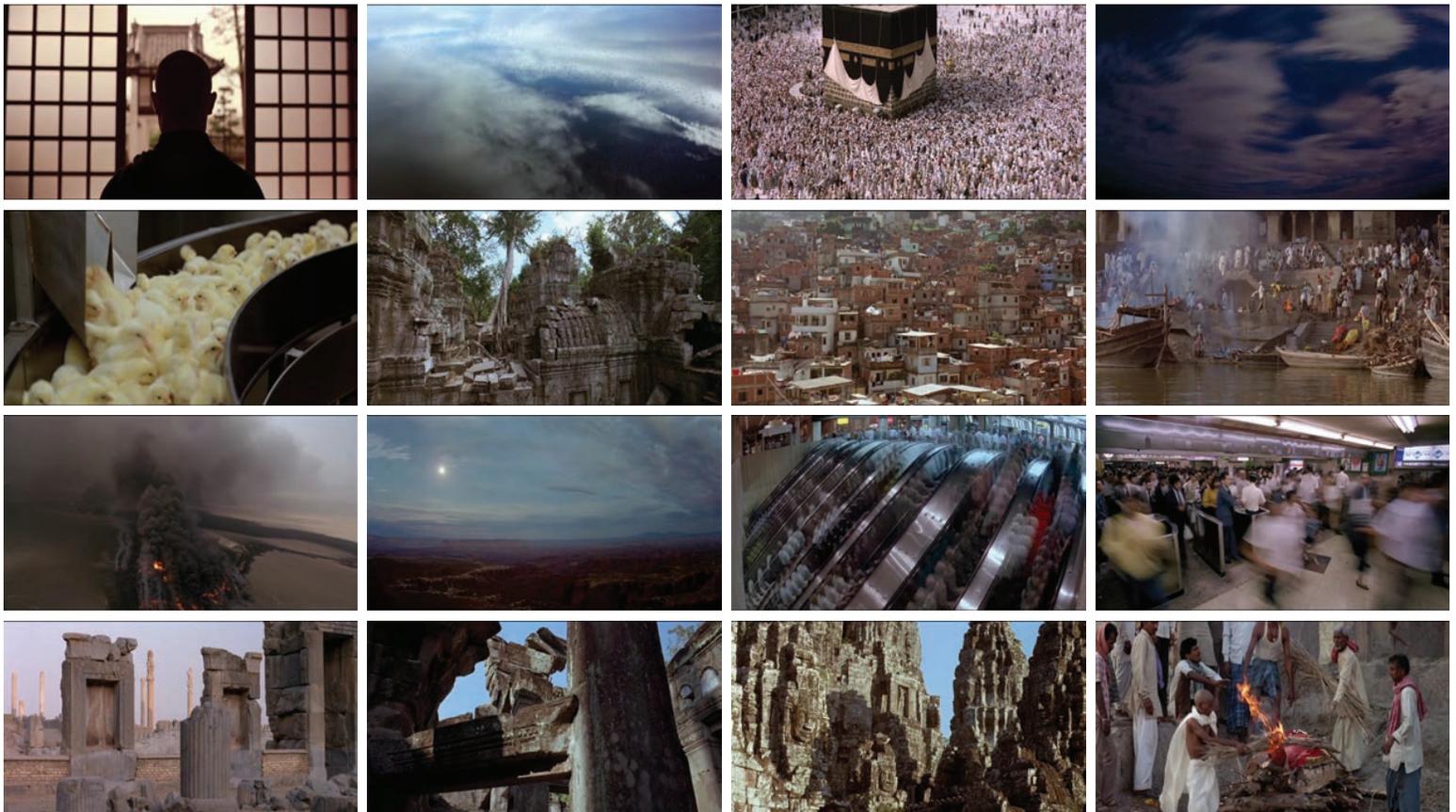


SFigure 14

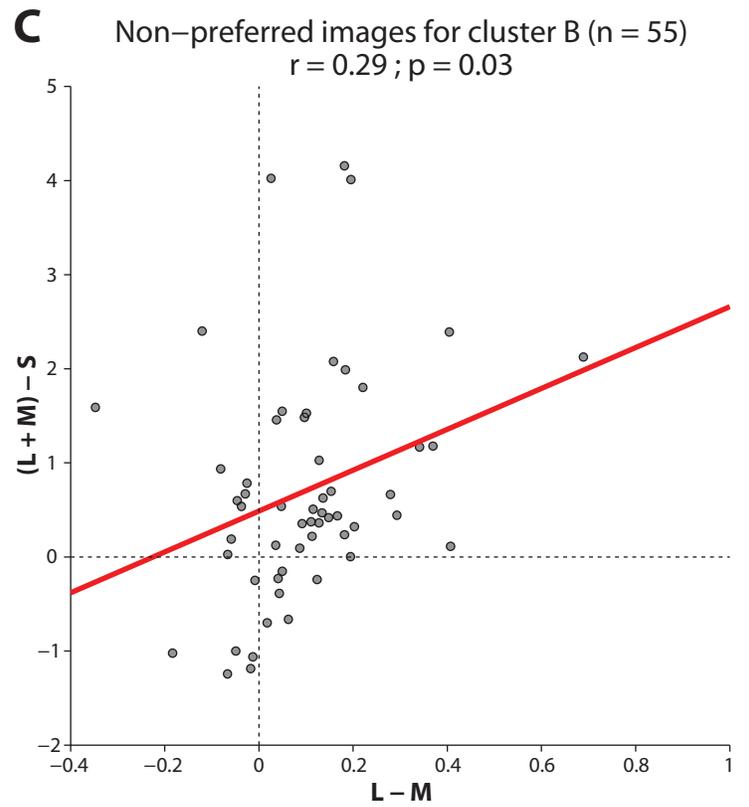
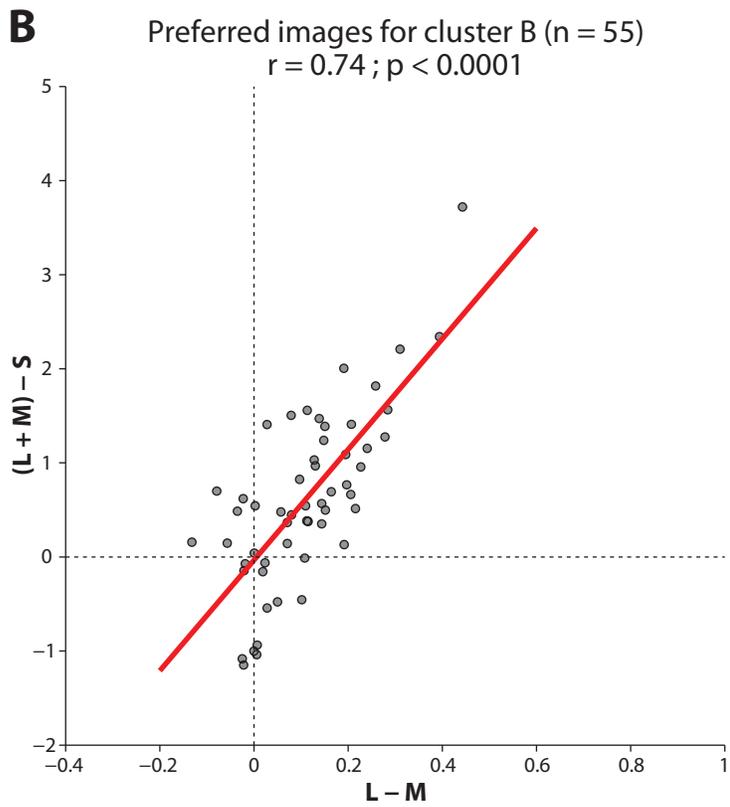
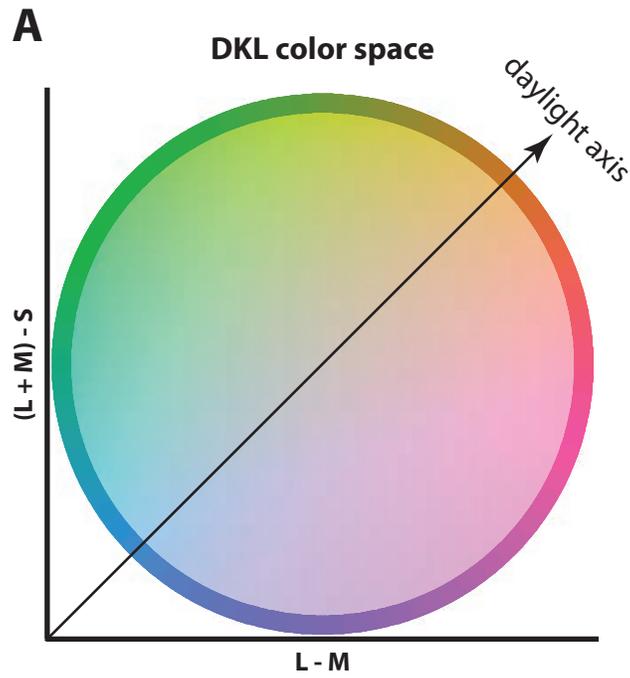
A



B

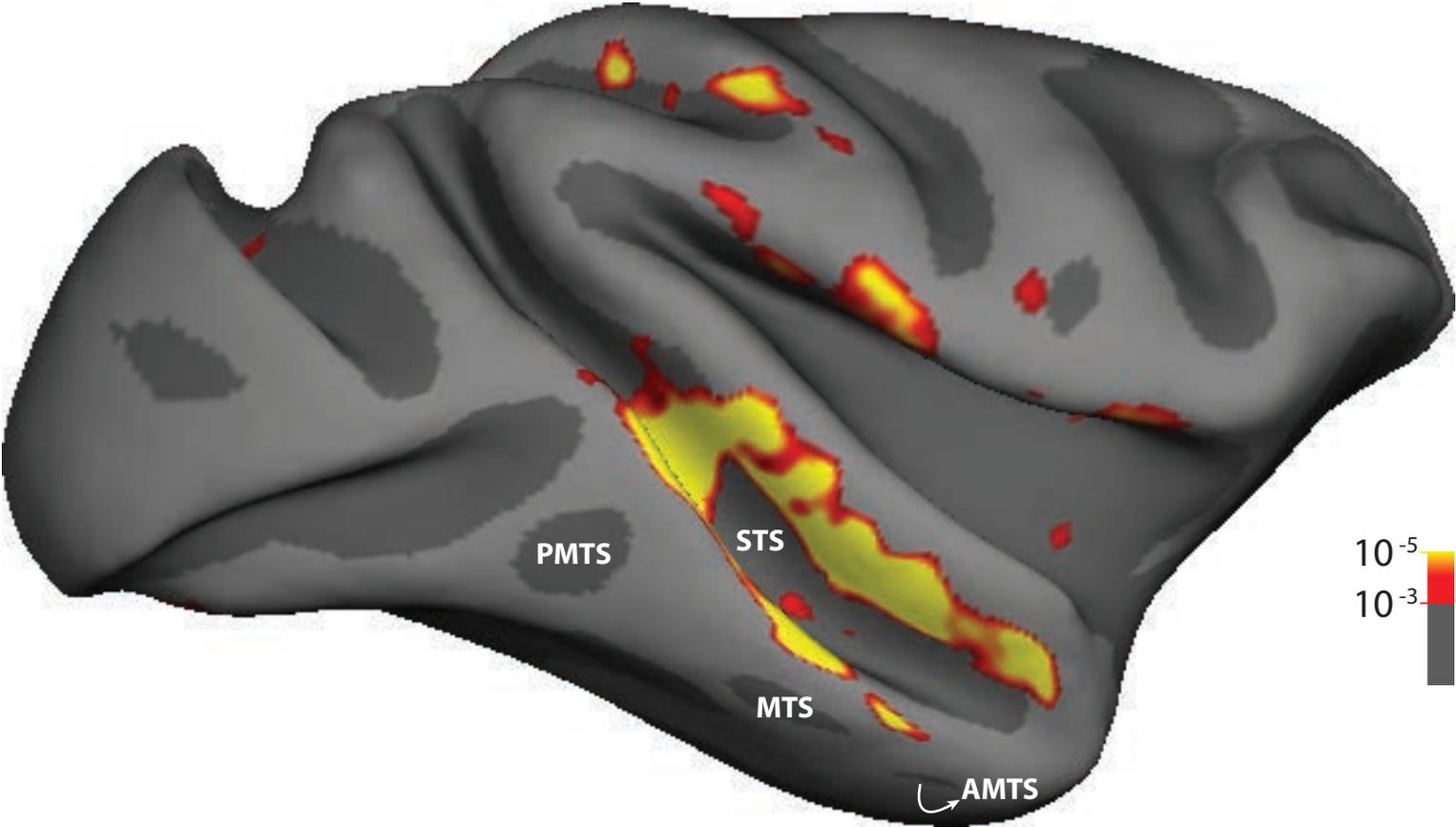


SFigure 15



SFigure 16

A functional network involved in the processing of animate stimuli



SFigure 17

