nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-020-00711-6

Topographic organization of the human subcortex unveiled with functional connectivity gradients

Ye Tian[1]✉, Daniel S. Margulies[2], Michael Breakspear[3,4] and Andrew Zalesky[1,5]✉

Brain atlases are fundamental to understanding the topographic organization of the human brain, yet many contemporary human atlases cover only the cerebral cortex, leaving the subcortex a terra incognita. We use functional MRI (fMRI) to map the complex topographic organization of the human subcortex, revealing large-scale connectivity gradients and new areal boundaries. We unveil four scales of subcortical organization that recapitulate well-known anatomical nuclei at the coarsest scale and delineate 27 new bilateral regions at the finest. Ultrahigh field strength fMRI corroborates and extends this organizational structure, enabling the delineation of finer subdivisions of the hippocampus and the amygdala, while task-evoked fMRI reveals a subtle subcortical reorganization in response to changing cognitive demands. A new subcortical atlas is delineated, personalized to represent individual differences and used to uncover reproducible brain–behavior relationships. Linking cortical networks to subcortical regions recapitulates a task-positive to task-negative axis. This new atlas enables holistic connectome mapping and characterization of cortico–subcortical connectivity.

The ubiquity of MRI has enabled high-throughput mapping of standardized brain parcellation atlases that represent consensus cortical boundaries across hundreds of scanned people[1]. Compared with the cerebral cortex, efforts to parcellate the human subcortex are scarcer. Indeed, many leading brain atlases delineate only cortical regions[2,3].

The subcortex is widely implicated in the pathophysiology of many neuropsychiatric disorders, and its role in cognition is unequivocal[4]. To better elucidate the role of the subcortex in health and disease, further work is needed to advance recent efforts to map the functional architecture of the subcortex[5,6]. This study aims to unravel the incredibly complex functional organization of the human subcortex by developing and validating a comprehensive subcortical atlas using 3 Tesla (3T) and 7T fMRI.

Cortical boundaries[7] within a brain atlas are often delineated at locations of abrupt spatial change in cortical microstructure, function and/or connectivity[2,8,9]. Cortical gradients[10] and the related connectopy concept[11] offer a complementary representation of cortical topography. Rather than delineating discrete spatial boundaries, topographic change is represented continuously along overlapping organizational axes[12].

Brain cartography studies tend to espouse one or the other of these two representations[1]. This is exemplified by current contention about whether the functional topography of the cerebellum is best represented by continuous gradients[13] or discrete boundaries[14]. However, the two paradigms are not necessarily contradictory. Spatial variation across expanses of the cerebrum may be evident but too gradual to warrant boundary delineation; thus, a continuum might be deemed to provide the most parsimonious representation. This leads to the following question of model selection: how sharp and conspicuous must a change be to warrant boundary delineation[15]? Here, we map functional connectivity gradients of the human subcortex and delineate subcortical boundaries only when justified by a formal model selection process.

Subcortical geometry is fundamentally different to the convoluted sheet-like structure of the cerebral cortex[16]. New paradigms are therefore required to facilitate subcortical cartography. To this end, we developed gradientography, a fMRI analog of diffusion MRI tractography that enables the quantification of subcortical connectivity gradients and a statistically principled formalism to guide boundary delineation. This allows us to comprehensively parcellate the subcortex on the basis of resting-state functional connectivity profiles and to produce a standardized atlas representing consensus boundaries among more than 1,000 healthy adults. The new atlas comprises four scales that recapitulate well-known anatomical nuclei at the coarsest scale and reveal 27 bilateral regions at the finest. The new atlas is replicated using ultrahigh field strength fMRI (7T) and is personalized to account for individual connectivity differences. An analysis of task-evoked fMRI (tfMRI) reveals a subtle reorganization of subcortical topography in response to changing cognitive demands. We use the new atlas to uncover correlational relationships between subcortical functional connectivity and individual variation in five behavioral domains and to reveal a new task-positive to task-negative organizational axis in cortico–subcortical connectivity[17]. This study provides an important step toward understanding the functional organization of the human subcortex.

Results

We investigated the functional connectivity architecture of the human subcortex using resting-state fMRI (rfMRI) data acquired at 3T ($n=1,080$) and 7T ($n=183$) and tfMRI data acquired at 3T ($n=725$) from healthy adults participating in the Human Connectome Project (HCP)[18]. Demographics are shown in Supplementary Table 1. Two 3T rfMRI sessions were acquired for

[1]Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne, Parkville, Victoria, Australia. [2]Centre National de la Recherche Scientifique (CNRS) UMR 8002, Integrative Neuroscience and Cognition Center, Université de Paris, Paris, France. [3]Discipline of Psychiatry, Faculty of Health and Medicine, University of Newcastle, Newcastle, New South Wales, Australia. [4]School of Psychology, Faculty of Science, University of Newcastle, Newcastle, New South Wales, Australia. [5]Department of Biomedical Engineering, Melbourne School of Engineering, The University of Melbourne, Parkville, Victoria, Australia. ✉e-mail: ye.tian2@unimelb.edu.au; azalesky@unimelb.edu.au

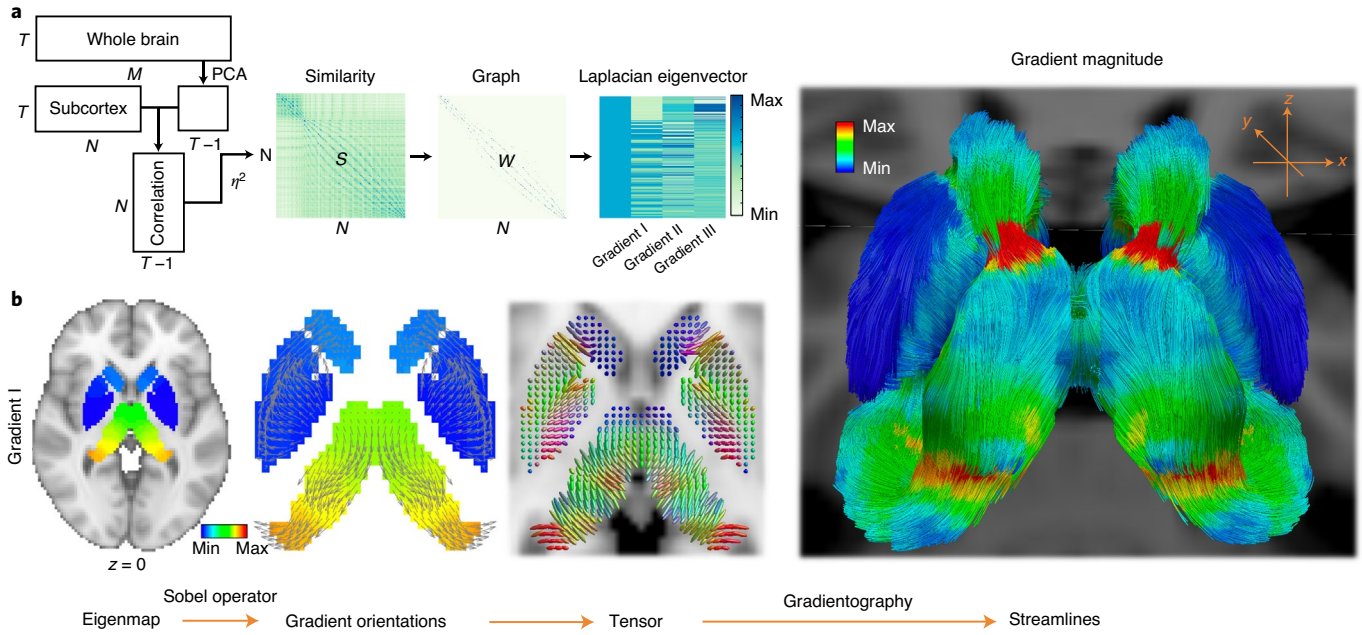

Fig. 1 | Gradientography in the human subcortex. a, A schematic of functional connectivity gradient mapping. The similarity in connectional fingerprints between pairs of subcortical voxels was measured using the η^2 coefficient. Pairs of voxels with similar fingerprints connected to similar brain areas. The similarity matrix (*S*) shown represents a group consensus for 1,080 healthy adults. The similarity matrix was transformed into a sparse graph with adjacency matrix *W*. The graph nodes correspond to subcortical voxels. Eigenvalues and eigenvectors were computed for the graph Laplacian. The eigenvectors with the second, third and fourth smallest eigenvalue are called gradients I, II and III, respectively. Each gradient provides a mapping of continuous variation in functional connectivity across the topography of the subcortex. *N*, number of subcortical voxels; *M*, number of whole-brain gray matter voxels; *T*, number of time frames. **b**, A schematic of the gradientography pipeline. Gradients were projected onto subcortical voxels to enable anatomical visualization. The axial slice shown is for gradient I. The Sobel operator was used to estimate the local gradient direction and magnitude for each subcortical voxel. The arrows shown point in the direction of the estimated gradients. Arrow lengths are normalized to unit length. Tensors were fitted to the gradient field. Tensors are colored according to the gradient direction (red, left–right; green, posterior–anterior; blue, superior–inferior). Streamlines were propagated through the tensor field using methods for diffusion MRI tractography. Streamlines are shown from a posterior viewpoint and colored according to the gradient magnitude. The red bands indicate peaks in the gradient magnitude that demarcate hippocampus–thalamus and caudate–thalamus boundaries.

each individual, with the first session used as the primary dataset (REST1) and the second used for replication (REST2). Gradients in functional connectivity were mapped using Laplacian eigenmaps and a newly developed procedure called gradientography. Formal model selection was used to determine whether gradient magnitude peaks were sufficiently large to warrant boundary delineation. The competing model posited a gradual spatial change in connectivity without discrete boundaries. This resulted in a new fMRI-based multiscale parcellation atlas of the subcortex, which we used to reveal new insights into subcortical function and architecture.

Connectivity gradients and gradientography. Whole-brain functional connectivity fingerprints[19] were mapped for each subcortical voxel and individual. Spatial gradients characterizing continuous modes of functional connectivity variation across the topography of the subcortex were computed using Laplacian eigenmaps and the connectopy method[11] (Fig. 1a). The eigenmap, or connectopy, explaining the most variance in functional connectivity was called gradient I (Fig. 1b), and the next two successive eigenmaps were called gradient II and gradient III, respectively (Supplementary Fig. 1). Further gradients explained substantially less variance and were therefore not considered (Supplementary Fig. 2). A new technique called gradientography was developed to visualize the connectivity gradients. Gradientography involves estimation of the principal gradient direction for each subcortical voxel, representing gradient directions and magnitudes with tensors, and then propagation of streamlines through the tensor field using established tools for diffusion MRI tractography. In this way, approximately 15,000 streamlines were initialized and propagated throughout the entire subcortical volume to map the principal organizational axes of the subcortex for gradient I (Fig. 1b) and for gradients II and III (Supplementary Fig. 1).

Gradient magnitude peaks indicated the locations of putative functional boundaries between regions. To visualize peak locations, streamlines were colored according to the gradient magnitude. Peak locations are shown as circumscribed red bands (Fig. 1b) and recapitulated boundaries between well-known anatomical nuclei. Local maxima in the gradient magnitude images showed moderate consistency among the three gradients (Supplementary Figs. 1 and 3); thus, only gradient I was used for boundary delineation. Gradient I streamlines were partitioned into dorsal and ventral groups at the location of the anatomical discontinuity between the caudate tail and the anterior extent of the thalamus. The dorsal group covered the globus pallidus (GP) and the striatum (Fig. 2a), whereas the amygdala, hippocampus and thalamus constituted the ventral group (Fig. 2b).

We next sought to quantify the size of each gradient magnitude peak to determine whether gradients were sufficiently abrupt to warrant boundary delineation. To this end, gradient magnitude images were projected onto two-dimensional trajectories called diversity curves[15]. A separate diversity curve was mapped for the dorsal and ventral streamline group. The dorsal diversity curve comprised three distinct local maxima indicating the locations of putative boundaries between the GP–putamen, the putamen–nucleus

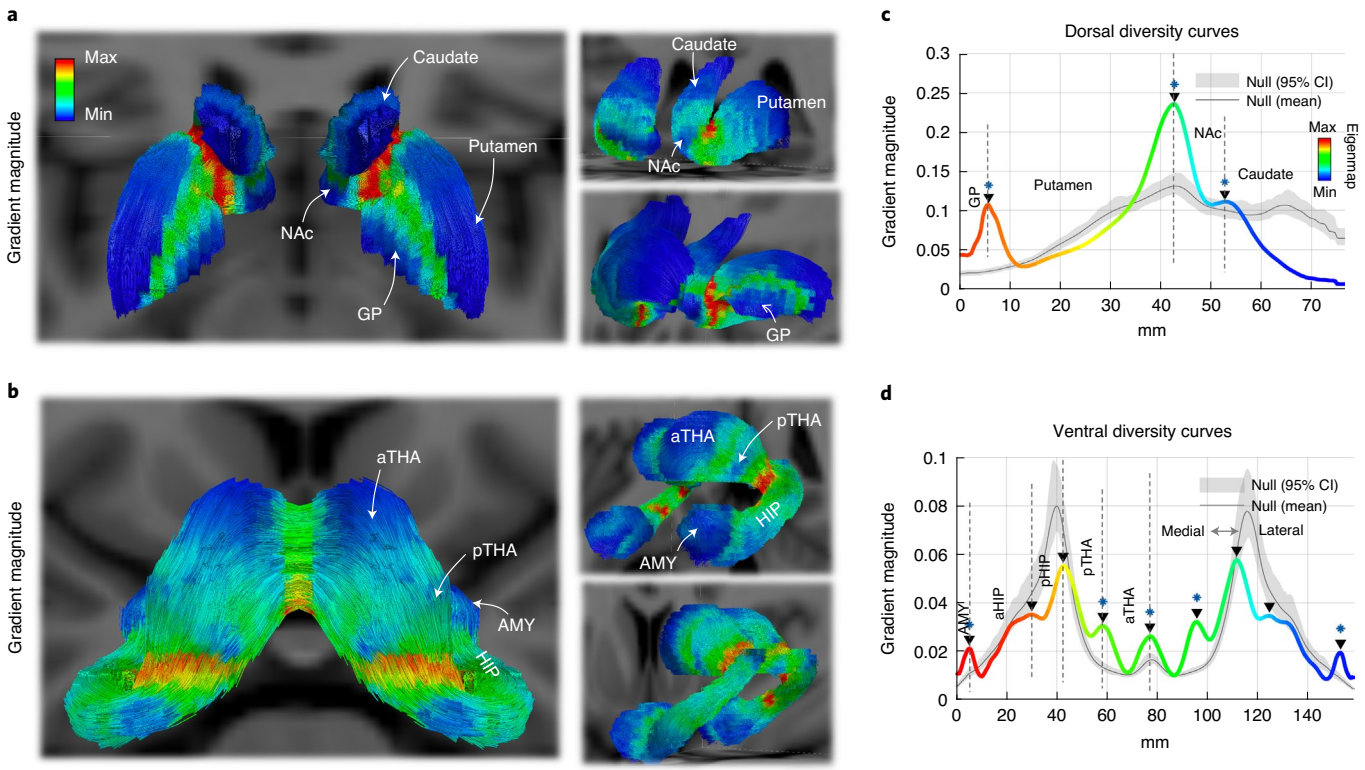

Fig. 2 | Model selection: which gradient magnitude peaks are sufficiently large to warrant boundary delineation? a,**b**, Streamlines were partitioned into bilaterally symmetric dorsal (**a**) and ventral (**b**) groups. The streamlines shown are colored according to the gradient magnitude. Dorsal streamlines traverse the GP, the putamen, the NAc and the caudate. Ventral streamlines traverse the left and right thalamus (THA; where a is anterior and p is posterior), the hippocampus (HIP) and the amygdala (AMY). **c**,**d**, Dorsal (**c**) and ventral (**d**) diversity curves were mapped to parameterize the gradient magnitude as a function of spatial position along the length of dorsal and ventral streamline trajectories. Black triangles indicate diversity curve peaks (local maxima) demarcating locations of putative functional boundaries. Diversity curves are colored according to eigenmap values for gradient I. Null diversity curves ($n=100$) for the null hypothesis of an absence of discrete boundaries are shown in gray, with shading used to show 95% confidence intervals (CI) and solid lines indicating the mean. Peaks in the null diversity curves are due to the effects of geometry. Peaks for which the null hypothesis could be rejected are demarcated with blue asterisks ($P<0.05$ one-sided, FDR-corrected for all 12 peaks).

accumbens (NAc) and the NAc–caudate (Fig. 2c). Nine local maxima were evident in the ventral diversity curve, which suggested boundary locations that distinguished the amygdala, the anterior and posterior hippocampus, the anterior and posterior thalamus, as well as the left and right thalamus (Fig. 2d). Locations of gradient magnitude peaks could not be explained by structural MRI signal properties and potential confounds, including the fMRI signal-to-noise ratio (SNR), the fMRI signal magnitude and the T1-weighted and T1/T2-weighted contrast (Supplementary Fig. 4).

A new subcortical atlas. The magnitude of local maxima in the diversity curves varied by a factor of more than ten, which suggests that putative boundaries between some regions were markedly more abrupt than others. How large should a peak be to warrant boundary delineation? Small peaks can emerge due to the convoluted geometry of the subcortex and other confounds (Supplementary Fig. 5). False-positive boundaries should not be delineated at such peaks. Therefore, for each peak, a formal model selection procedure was used to test the null hypothesis that the peak magnitude could be explained by geometry and/or other confounds. Rejection of the null hypothesis indicated that boundary delineation was warranted. Otherwise, if the peak magnitude was not sufficiently large to enable rejection of the null hypothesis, a continuous representation of spatial variation in connectivity was deemed to be the most parsimonious model[15]. The null hypothesis was tested using a geometry-preserving null model (Methods). Diversity curves generated using the null model are colored gray in Fig. 2c,d. The actual peaks significantly exceeded all null data ($P<0.05$, false discovery rate (FDR)-corrected for all 12 peaks), except for peaks separating the bilateral hippocampus/thalamus and the anterior–posterior hippocampus. However, given that the hippocampus and the thalamus are anatomically and functionally distinct and that the peak separating these two regions was unequivocal, we made an exception and delineated a boundary here, whereas the putative anterior–posterior hippocampus boundary was not delineated, which is consistent with the null data. While peaks in the null diversity curves often corresponded with locations to peaks in the actual diversity curves, including the functional boundaries between the putamen–NAc, the posterior hippocampus–posterior thalamus and the left–right thalamus, distinct peaks were also evident where the effect of geometry is minimal (Supplementary Fig. 5a), such as the amygdala–anterior hippocampus, the anterior–posterior thalamus and the GP–putamen. This indicates that subcortical geometry can predict and account for the location of some functional boundaries, but geometry per se cannot explain the magnitude of all boundaries, except for the hippocampus–thalamus (see "Further methodological considerations" in the Methods). Finally, the watershed transform algorithm was used to segment subcortical voxels into contiguous regions such that boundaries demarcated watershed lines.

This resulted in eight bilateral regions that constituted a relatively coarse atlas, referred to as the scale I atlas (Fig. 3a). The scale I atlas recapitulated well-known anatomical nuclei, namely, the amygdala,

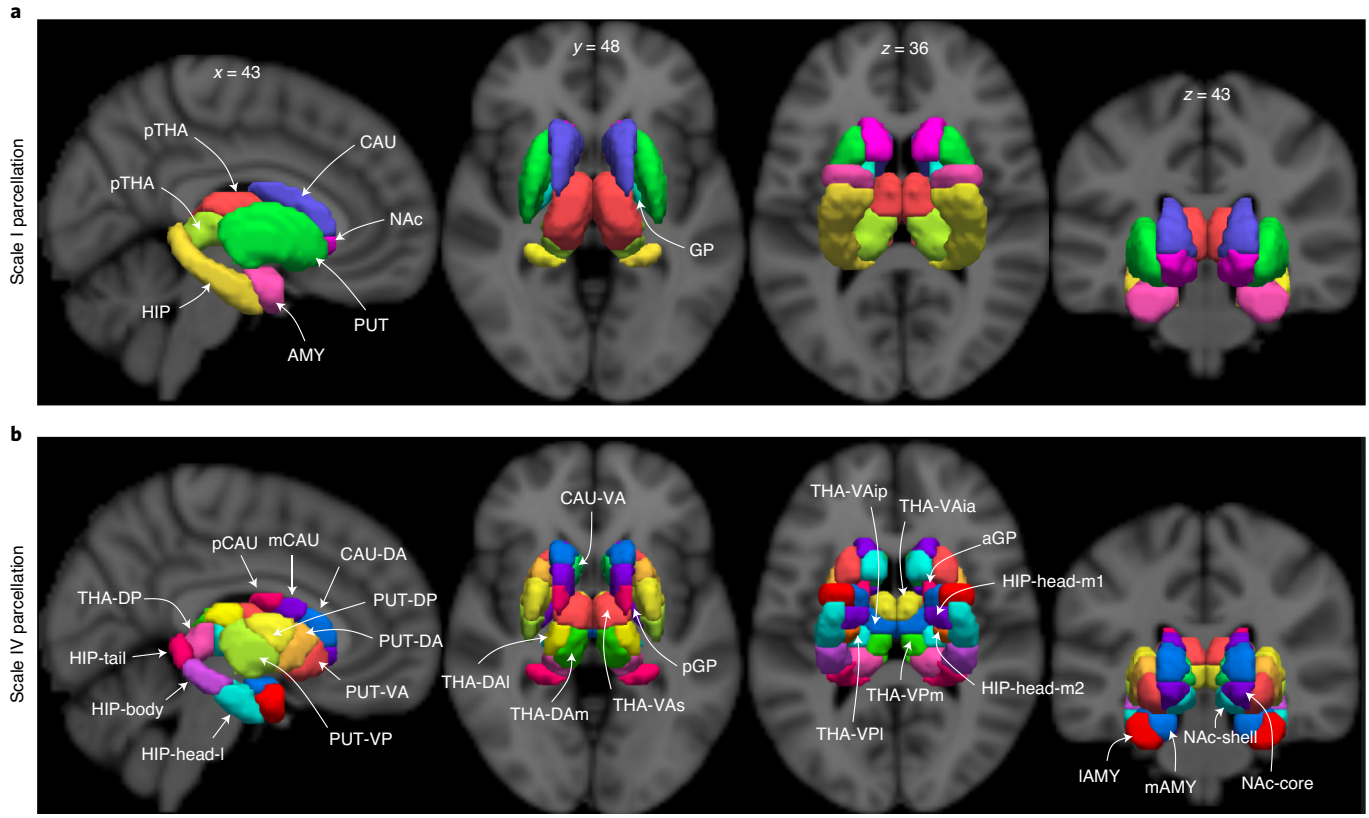

Fig. 3 | A multiscale group-consensus parcellation atlas derived from 3T-rfMRI acquired in 1,080 healthy adults. a, Scale I is the coarsest atlas and comprises eight bilateral regions that recapitulate anatomical nuclei, including the caudate nucleus (CAU), the NAc, the putamen (PUT), the amygdala, the hippocampus and the thalamus. An anteroposterior division within the thalamus is also present. **b**, Scale IV is the finest scale of the parcellation hierarchy and comprises 27 bilateral regions. Homologous regions are identically colored. d, dorsal; i, inferior; l, lateral; m, medial; s, superior; v, ventral (detailed nomenclature for all regions is provided in Supplementary Table 2).

the hippocampus, the thalamus, the GP, the NAc, the caudate and the putamen, as well as an anteroposterior thalamic partition, which corresponded to the spatial distribution of the two subpopulations of thalamic projection cells (matrix and core)[20]. Gradient mapping, boundary delineation and model selection were then iteratively performed separately for each region constituting the scale I atlas. This resulted in a finer atlas, called the scale II atlas, comprising 16 bilateral regions. In turn, the null hypothesis was independently tested for each region in the scale II atlas. Recursively continuing this process until the null hypothesis could not be rejected for any regions yielded four progressively more detailed atlases of the subcortex (scale I: 16 regions (Fig. 3a); scale II: 32 regions; scale III: 50 regions; scale IV: 54 regions (Fig. 3b); Supplementary Fig. 6). The new atlas is hierarchically organized, with the entire subcortex residing at the base of a hierarchical tree and the smallest regions constituting the scale IV atlas defining the leaves (Fig. 4a,b). Subcortical gradients were also hierarchically organized (Fig. 4b,c and Supplementary Fig. 7). A schematic of the recursive procedure is provided in Supplementary Fig. 8.

Parcellation of the amygdala terminated at scale II, and striatal parcellation continued until scale III, while only the thalamus and the hippocampus progressed to scale IV. A nomenclature scheme for all delineated regions was developed (Supplementary Table 2). In particular, regions that could be unequivocally matched to homologous anatomical structures were labeled as such, while the remaining regions were labeled with reference to standard anatomical planes (for example, anterior–posterior, medial–lateral, dorsal–ventral and superior–inferior).

Scales I and II generally recapitulated anatomical boundaries, whereas some divergence emerged at finer scales between anatomy and the boundaries delineated here based on functional connectivity. The eight thalamic subregions constituting scale IV showed moderate spatial correspondence with histologically delineated thalamic nuclei (Supplementary Fig. 9a). The hippocampus was segmented along its long axis into anterior (head) and posterior (body and tail) components (Supplementary Fig. 9b), which is consistent with previous functional-based hippocampal parcellations[21]. Although the head of the hippocampus was further divided into medial and lateral components, the overall long-axis organization diverged from the characteristic mediolateral organization that is histologically found[22].

Connectivity gradients (gradients I–III) for the hippocampus, the striatum and the thalamus were projected onto the cortical surface on the basis of the cortical vertices with which they were most strongly connected[11,23] (Supplementary Fig. 10a–c).

Parcellation homogeneity. Having delineated a new subcortical atlas, we next sought to investigate the validity and reproducibility of the atlas. A key marker of parcellation validity is within-parcel homogeneity of functional signals[1]. Homogeneity was measured on the basis of within-parcel synchrony of fMRI time series acquired in the second fMRI session (REST2), whereas the first session (REST1) was used for parcellation. Homogeneity was separately computed for each region and then averaged across all regions to yield parcellation homogeneity estimates for scales I–IV (Supplementary Fig. 11a).

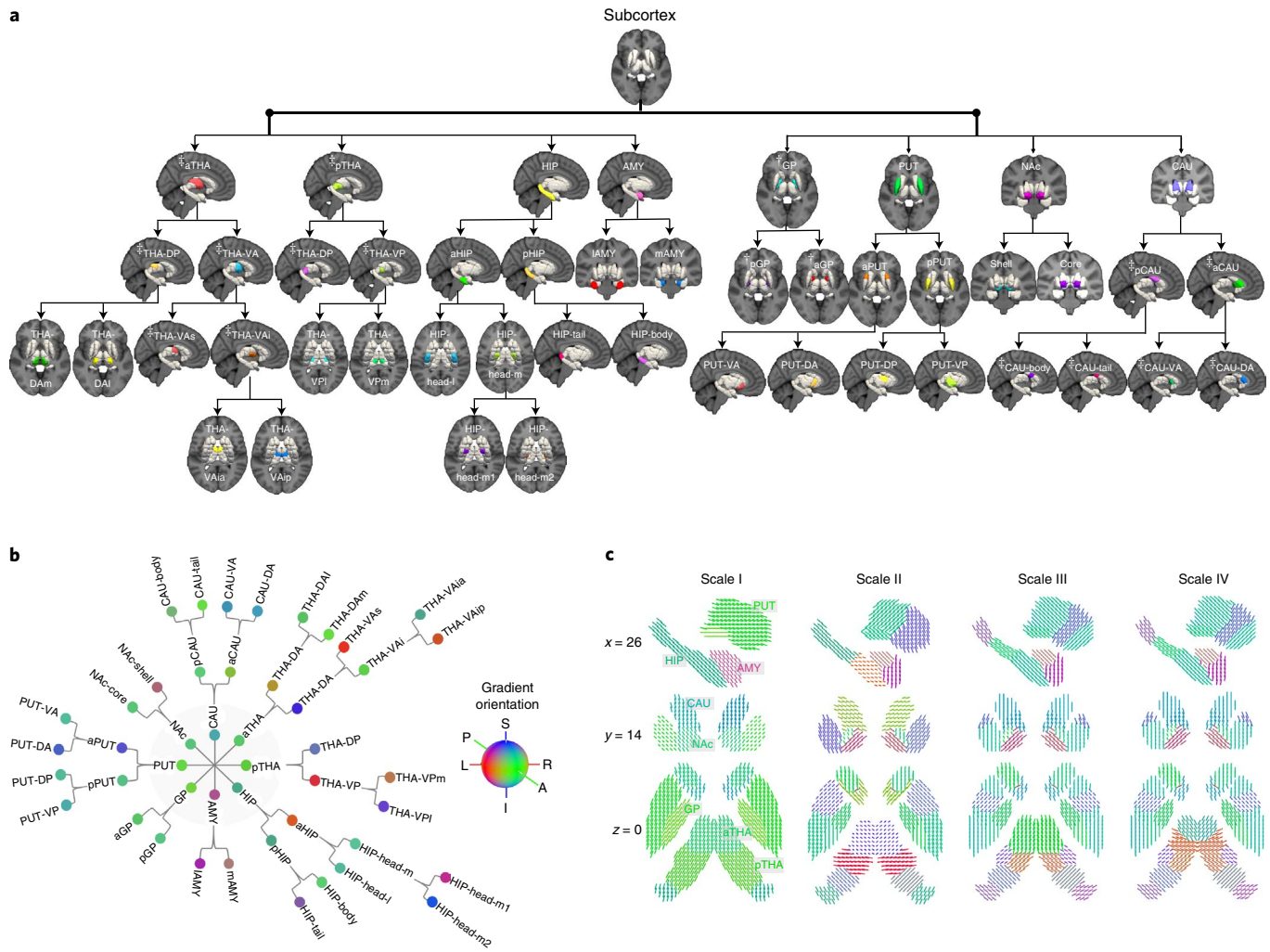

Fig. 4 | Organization of the human subcortex. Model selection and null hypothesis testing was used to determine whether gradient magnitude peaks were sufficiently large to warrant boundary delineation. This process was recursively repeated for each new region delineated, unveiling a multiscale parcellation architecture. The null hypothesis of an absence of boundaries could not be rejected for any of the regions constituting scale IV; thus, further (finer) parcellation was not warranted. **a**, A top-down representation of the parcellation hierarchy. Regions are shown atop a reference structural MRI image. Homologous regions are identically colored. †The caudate was removed for visualization purposes; ‡the putamen and the GP were removed for visualization purposes. **b**, A parcellation hierarchy visualized in the form of a circular tree. The central node represents the entire subcortex and the other nodes represent distinct regions. Regions are arranged within four concentric circles, where the innermost circle (gray) is the first level (scale I). Nodes are colored according to principal gradient orientations within each region (red, left–right; green, posterior–anterior; blue, superior–inferior). The principal gradient orientation was estimated by averaging the local gradient directions across all voxels within each region. **c**, Region-averaged gradient directions represented with quiver lines for scales I–IV. The same images without region averaging are shown in Supplementary Fig. 7. Detailed nomenclature for all regions is provided in Supplementary Table 2.

As expected[8], homogeneity monotonically increased from scale I to scale IV (Fig. 5a). To test whether parcellation homogeneity was significantly greater than expected due to chance, homogeneity was also estimated for ensembles of random subcortical parcellations (Supplementary Fig. 11b). Random parcellations were generated by delineating boundaries at arbitrary locations, subject to preservation of key geometric properties (Methods). Scales II–IV were significantly more homogeneous than expected due to chance ($P<0.01$), whereas this was not the case for scale I ($P=0.23$; Fig. 5a). Scale I regions (Fig. 3a) encompassed multiple functionally distinct substructures, which potentially explains the relatively low within-parcel homogeneity of this scale. Parcellation homogeneity was compared between the new atlas and existing imaging- and histology-based subcortical parcellations (Supplementary Table 3). Striatal, hippocampal and thalamic parcellations were isolated from the new atlas to enable comparisons with existing parcellations of these regions. The new atlas was significantly more homogeneous than existing imaging-based parcellations of the entire subcortex (subcortex-I: $P<0.001$; subcortex-II: $P<0.001$), the hippocampus ($P<0.001$), the thalamus ($P<0.001$) or the striatum ($P<0.001$), as well as a histological atlas of the hippocampus (Hippocampus-His: $P<0.001$; Fig. 5b and Supplementary Fig. 12). These results were replicated in an independent dataset comprising ten healthy adults (six males, mean age of 26 ± 2.1 years; Supplementary Fig. 13).

In auxiliary analyses, we tested whether task-evoked activity mapped onto the new atlas and respective atlas boundaries. To this end, the standard deviation in task-evoked activity was computed across voxels constituting each region and then averaged over all regions. This was repeated for scales I–IV. The lower the standard deviation, the more circumscribed task-evoked activity was to

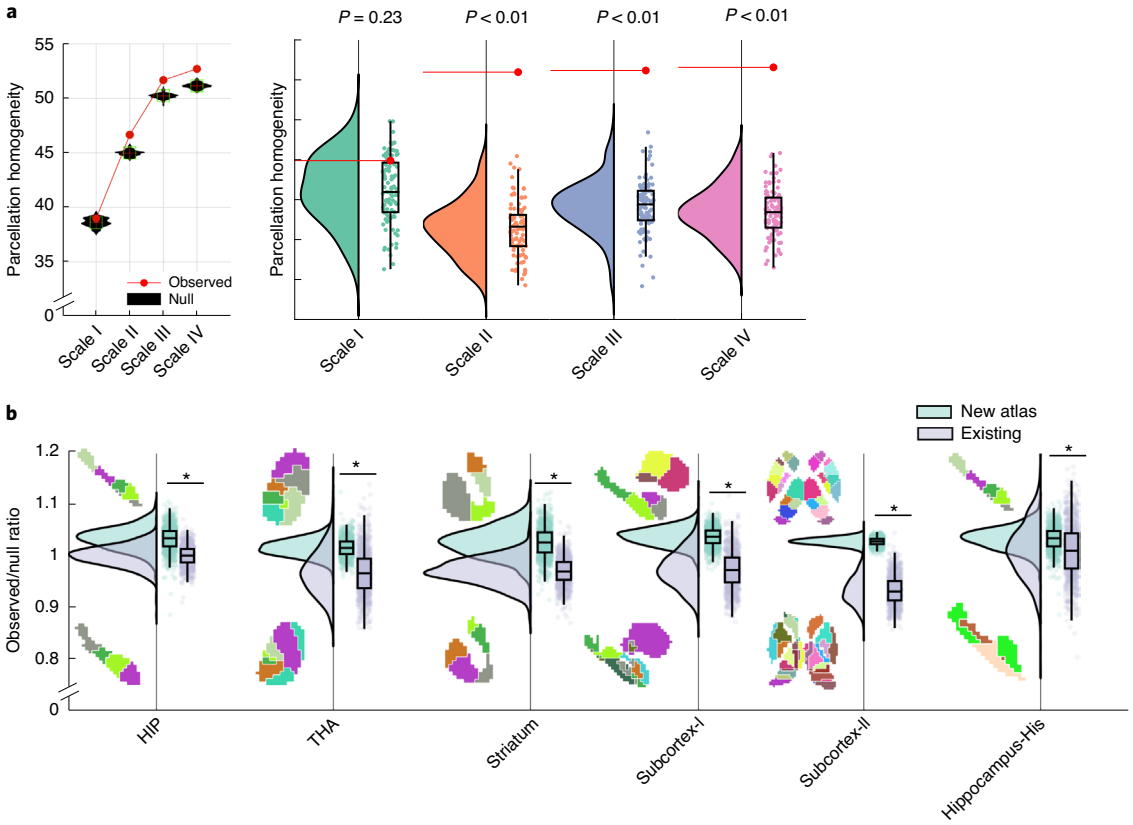

Fig. 5 | Parcellation homogeneity. a, Parcellation homogeneity of the new atlas was benchmarked to random parcellations of the subcortex comprising an identical number of regions and a comparable distribution of region sizes. Violin plots show the distribution of parcellation homogeneity measured within ensembles of 100 random parcellations. The *P* value (one-sided) shown for each scale is the proportion of random parcellations comprising the ensemble that were more homogeneous than the observed parcellation. The horizontal red lines indicate the observed parcellation homogeneity values. The bottom and top edges of the boxes indicate the 25th and 75th percentiles of the distribution, respectively. The central mark indicates the median. The whiskers extend to the most extreme data points that are not considered outliers (1.5-times the interquartile range). The vertical axis is omitted to aid visualization. The null hypothesis of parcellation homogeneity that is no better than chance could be rejected for all scales except scale I. **b**, Parcellation homogeneity comparisons between the new atlas and existing parcellation atlases of the entire subcortex and specific subcortical nuclei (Supplementary Table 3). Parcellation homogeneity was separately computed for each individual (REST2, $n = 1{,}021$) and normalized by the random parcellation homogeneity, yielding an observed-to-null ratio. The box plots show the distribution of this ratio across individuals for the new (turquoise) and existing (violet) parcellation atlases. The bottom and top edges of the boxes indicate the 25th and 75th percentiles of the distribution, respectively. The central mark indicates the median. The whiskers extend to the most extreme data points that are not considered outliers (1.5-times the interquartile range). Homogenous parcellations have a ratio that exceeds one. Asterisks indicate that the observed-to-null ratio is significantly higher for the new atlas relative to the existing atlas used for comparison ($P < 0.001$, two-sided two-sample *t*-test). Anatomical visualizations of the new and comparison atlases are shown above and below the box plots, respectively.

particular atlas regions. For all task conditions, we found that standard deviations were significantly lower for the new atlas than comparable random parcellations of the subcortex ($P < 0.01$; Supplementary Fig. 14). This cross-modality validation suggests that task-evoked activity is circumscribed to specific subcortical atlas regions more than expected by chance.

Parcellation replication. Ultrahigh field strength MRI can alleviate several technical challenges associated with subcortical imaging[24]. Hence, we sought to investigate whether the new atlas could be replicated using 7T fMRI. Qualitatively, gradient magnitude peaks were sharper and more circumscribed at 7T than at 3T (Fig. 6). Gradientography, model selection and boundary delineation applied to 7T data yielded a comparable but more detailed atlas of the subcortex that also comprised four scales (scale I: 16 regions; scale II: 34 regions; scale III: 54 regions; scale IV: 62 regions). The number of regions delineated at scale I was identical between 3T and 7T data (Fig. 6). However, additional boundaries could be identified with 7T data at scales II–IV (Supplementary Figs. 15 and 16). In particular, the amygdala was partitioned into medial and lateral subregions at 3T, whereas its medial extent was further divided into centromedial and superomedial components at 7T, and the body of the hippocampus was further divided mediolaterally. Based on normalized mutual information (NMI), the spatial correspondence between the 3T and 7T atlases was excellent (NMI = 0.93, 0.88, 0.84 and 0.83 for scale I, scale II, scale III and scale IV, respectively; Supplementary Fig. 16). Correspondence between 3T and 7T atlases was significantly higher than expected due to chance at all parcellation scales ($P < 0.01$; Supplementary Fig. 16).

Personalized parcellation. Having delineated a group-consensus atlas of the subcortex, we next sought to personalize the new atlas to account for individual variation in connectivity architecture[25,26]. Following previous work on cortical atlas individualization[2],

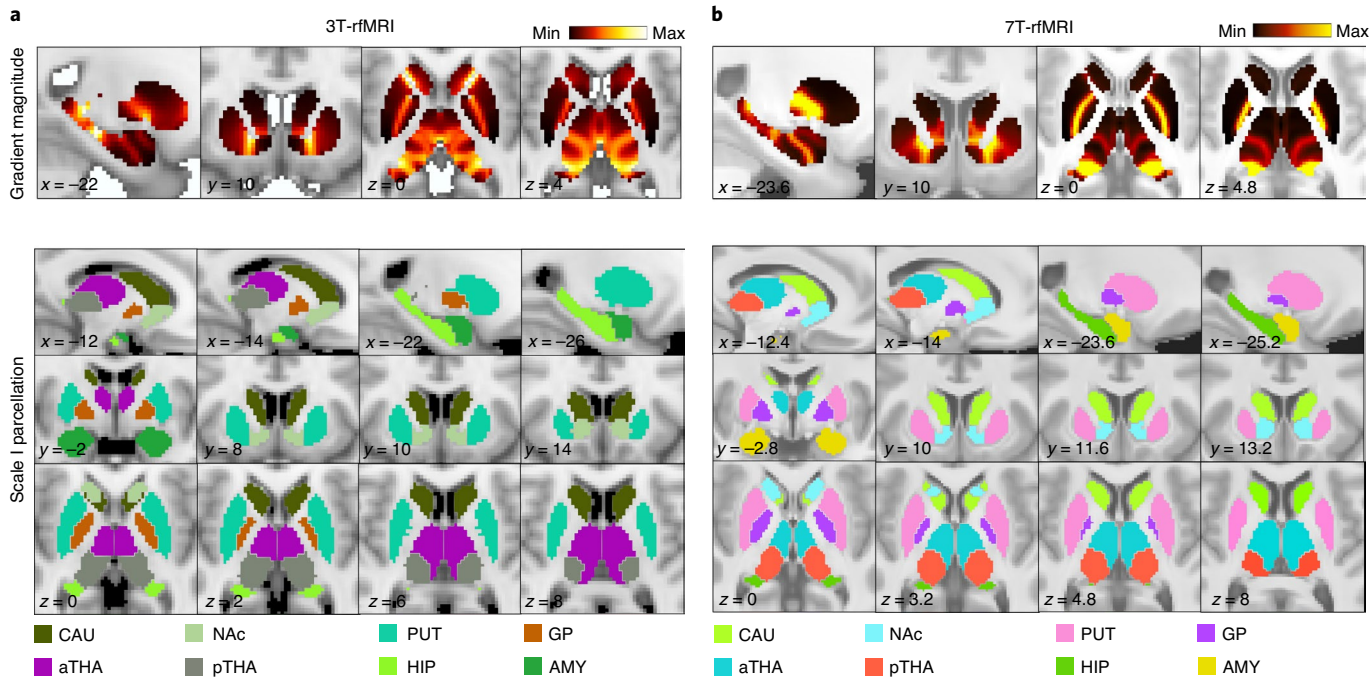

Fig. 6 | Replication of 3T parcellation using 7T-rfMRI. a, Group-consensus gradient magnitude image (top row) and scale I atlas (bottom three rows) delineated using 3T-rfMRI ($n=1{,}080$). Bands colored in yellow–orange indicate local maxima in the gradient magnitude image. Homologous regions are identically colored. **b**, Same as **a**, but delineated using 7T-rfMRI ($n=183$). The eight regions comprising scale I were replicated with 7T imaging. All slice coordinates are indicated in MNI space (mm). A reference structural MRI image is used as the background.

a machine-learning classifier was trained ($n=100$, REST2) to recognize putative connectivity fingerprints characteristic of each scale IV region. The trained classifier was then used to delineate personalized boundaries for each region in an independent group ($n=921$, REST2). In particular, for each individual, voxels were assigned a classification probability of belonging to each region and a winner-takes-all rule was used to delineate personalized parcellations (Supplementary Fig. 17).

Quantification using the Dice coefficient showed that the extent to which the personalized atlases recapitulated the group-consensus atlas is markedly varied among individuals and regions. This was most evident within the striatum (Fig. 7a) and the medial subdivisions of the hippocampal head. Group-averaged Dice coefficients (Fig. 7b) varied from 0.22 (subdivision of the medial hippocampal head (HIP-head-m2-lh)) to 0.85 (dorsoposterior putamen (PUT-DP-rh)) for a classification probability threshold of 50%. Hotspots in the group-averaged classification probability maps were encircled by the group-consensus boundaries (Fig. 7c), which indicates self-consistency in classification performance. Individual variation in Dice coefficients was highly skewed, as evidenced for the dorsoposterior putamen and the hippocampal body (Fig. 7d). Specifically, a majority of individuals recapitulated the group-consensus (Dice>0.8), but 2–7% showed marked divergence (Dice<0.5). Boundary contraction was the most common source of divergence among individuals and the group-consensus atlas (Fig. 7c). The same individuals consistently deviated from the group-consensus atlas for all regions, which suggests that this deviation was specific to individuals rather than regions (Supplementary Fig. 18).

It was possible for the classifier to fail to detect particular regions in some individuals. For each region, the proportion of individuals for whom at least one voxel was assigned to that region, called the regional detection rate, was analyzed as a function of the classification probability threshold (Fig. 7e). The regional detection rate for a classification probability threshold of 50% was rendered on the group-consensus atlas (Fig. 7e, inset). The averaged regional detection rate was 97.8%, with detection rates falling below 95% for only two bilateral regions (that is, the HIP-head-m1 and the HIP-head-m2). This suggests that the group-consensus atlas is representative of most individuals, although a minority of individuals may benefit from delineation of individual-specific atlases to account for inter-individual variation in regional boundaries.

Rest compared to task-evoked conditions. We have thus far focused on rfMRI data. Using tfMRI, we next investigated the extent to which the functional topography of the subcortex reconfigures in response to changing cognitive demands[27]. For each of the seven tasks and the two rest sessions, gradient magnitudes (gradient I) were projected onto the previously mapped ventral and dorsal streamlines, which yielded task-evoked diversity curves.

Qualitatively, diversity curve peaks were consistent among the seven task conditions, but some peak locations and magnitudes differed relative to the rest condition (Supplementary Fig. 19a). The most prominent task–rest variation was bifurcation of the boundary separating the putamen–NAc and the NAc–caudate during the task conditions. Functional separation between the amygdala and the anterior extreme of the hippocampus, the posterior hippocampus and the thalamus, as well as anterior–posterior partition within the thalamus were also weaker during task compared to rest conditions, whereas separation between the anterior and posterior hippocampus was stronger and anteriorly extended (Supplementary Fig. 19a). To investigate this further, the Pearson correlation coefficient was used to quantify the similarity in diversity curves between all pairs of conditions, which yielded correlation matrices (Supplementary Fig. 19b) that were visualized as correlation graphs (Supplementary Fig. 19c). Diversity curves were highly consistent between the two rest sessions (dorsal: $r=0.91$, $P<0.0001$; ventral: $r=0.98$, $P<0.0001$), but were distinct from the seven task conditions, thereby resulting in distinct task and rest modules (Supplementary

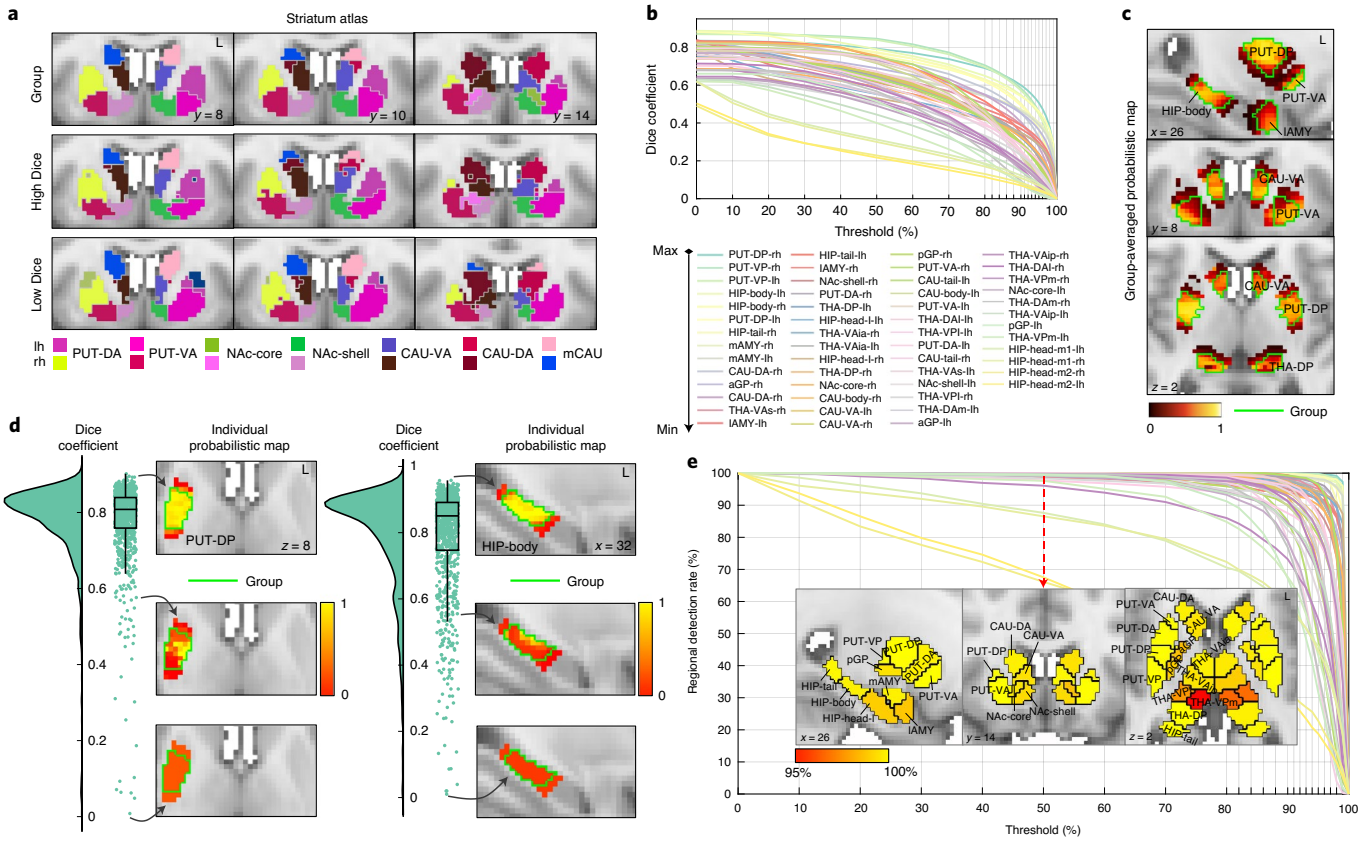

Fig. 7 | Personalization of the scale IV atlas. a, Group-consensus atlas for striatal regions (top row) and personalized atlases for representative individuals with relatively high (middle row, Dice = 0.81) and low (bottom row, Dice = 0.15) similarity with the group-consensus atlas. **b**, Curves showing group-averaged Dice coefficients as a function of the classification probability threshold (0–100%). A separate curve is shown for each of the 54 regions constituting the scale IV atlas. The classification probability is the probability that a subcortical voxel is classified as belonging to a specific region in a given individual. Parcellations were realized for each classification probability threshold and compared to the group-consensus atlas using the Dice coefficient. The area under the curve was computed for each region, and the legend is shown in descending order from top to bottom and left to right. **c**, Group-averaged classification probability maps exemplified in seven regions. The voxel color is scaled by the classification probability of the voxel belonging to the given region. The group-consensus boundary is shown in green. **d**, Box plots showing the distribution of individual variation in Dice coefficients of the dorsoposterior putamen (PUT-DP, left) and the hippocampal body (HIP-body, right). Each data point represents the Dice coefficient of one of the 921 individuals used to evaluate classification performance. The bottom and top edges of the boxes indicate the 25th and 75th percentiles of the distribution, respectively. The central mark indicates the median. The whiskers extend to the most extreme data points that are not considered outliers (1.5-times the interquartile range). Classification probability maps are shown in anatomical space for representative individuals with relatively high (top), moderate (middle) and low (bottom) Dice coefficients. The group-consensus boundary is shown in green. **e**, Curves showing the regional detection rates as a function of the classification probability threshold. The inset shows anatomical renderings of regional detection rates for a nominal classification probability threshold of 50%. The two hippocampal regions (bilateral HIP-head-m1 and HIP-head-m2) with the lowest regional detection rates are obscured in the slices shown. The dashed red arrow indicates the threshold (50%) used for the images shown in the inset. All slice coordinates are indicated relative to MNI space (mm). A reference structural MRI image is used as the background in anatomical visualizations.

Fig. 19d). Task-evoked reconfiguration of functional organization was not due to differences in the number of fMRI volumes acquired between task and rest conditions (Methods).

Links between subcortical topography and cortical networks. Next, we investigated functional connectivity between canonical cortical networks and subcortical regions (scale IV). Similarities in patterns of functional connectivity with the subcortex were quantified between pairs of 12 established cortical networks[8] (Supplementary Fig. 20a). Based on the similarity in the subcortical areas to which they most strongly connected, the 12 cortical networks could be parsed into three distinct groups and positioned on a task-positive to task-negative[17] organizational axis (Supplementary Fig. 20b). One extreme of the axis comprised cortical networks that typically showed increased activity during attention-demanding cognitive tasks (that is, task-positive), including the salience, cingulo–parietal, cingulo–opercular, frontoparietal and dorsal attention network. In contrast, the default mode network, which is often characterized by task-related deactivation (that is, task-negative) occupied the other extreme. Between them, sensorimotor-related networks (visual, auditory, dorsal and ventral somatomotor) and ventral attention and retrosplenial–temporal networks formed a third group, in which the latter two also linked to the default mode network. This result suggests that there is a transition from sensorimotor functions toward task-negative functions.

Using the new atlas to relate human behavior to subcortical connectivity. Finally, we used the new atlas to unveil new relationships between subcortical functional connectivity and individual variation in human behaviors. Five behavioral dimensions characterizing (1) cognition, (2) illicit substance use, (3) tobacco use, (4) personality and emotional traits and (5) mental health were

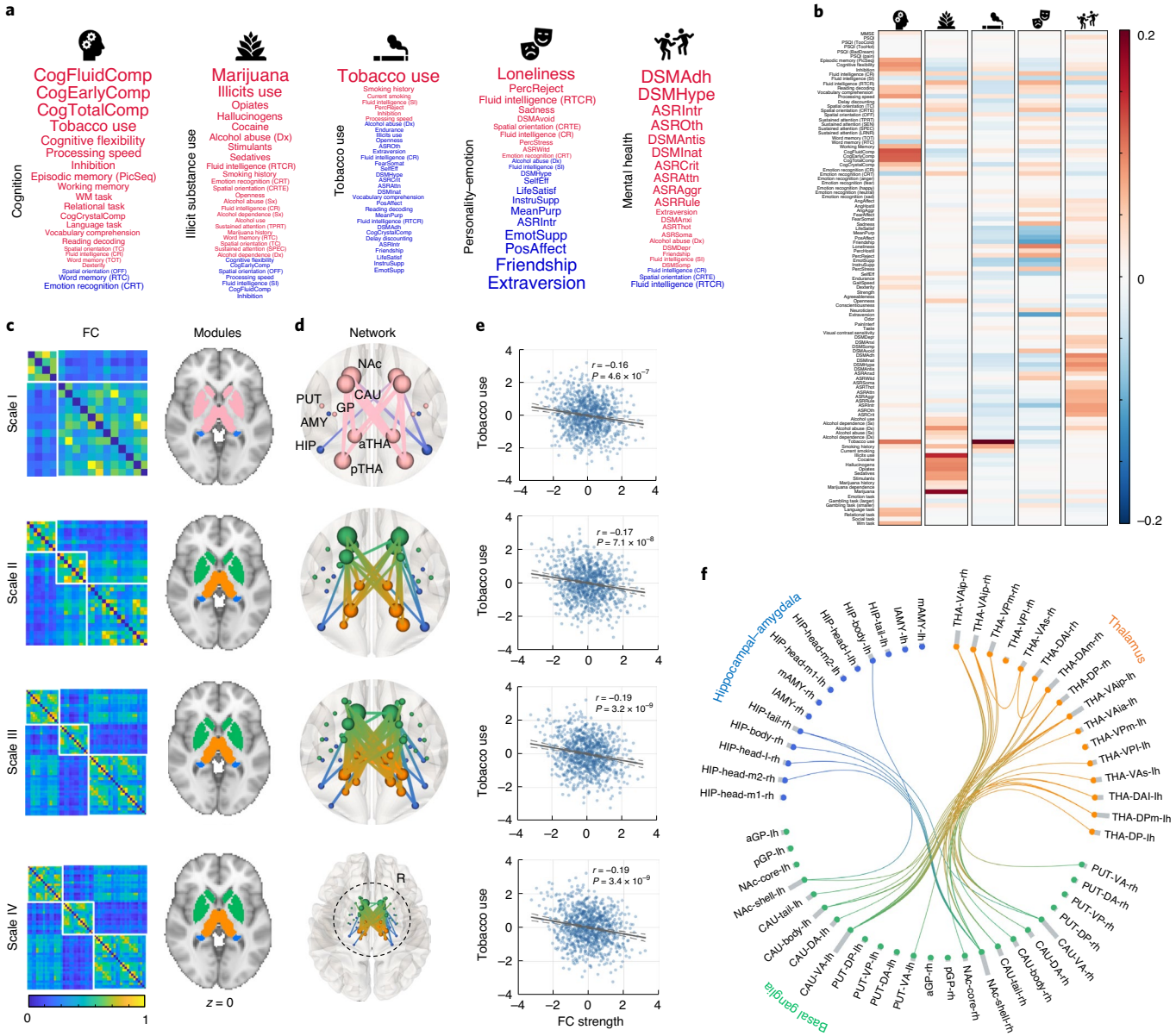

Fig. 8 | Associations between individual variation in behavioral dimensions and subcortical functional connectivity. a, Word clouds characterizing five behavioral dimensions derived from 109 behavioral items using ICA. Only the most highly weighted behavioral items are shown. The font size is scaled according to the absolute weight of the item in the ICA de-mixing matrix. The font color indicates weight polarity (red, positive; blue, negative). **b**, A consensus de-mixing matrix estimated using ICA. Rows correspond to behavioral items. Columns correspond to the five latent dimensions. Cell color is modulated by de-mixing weight. **c**, Left: group-averaged functional connectivity (FC) matrices for scale I–IV atlases. Rows and columns are reordered to accentuate the modular structure (white blocks). Right: modules are also rendered in anatomical space, where regions constituting the same module are identically colored. Scale I comprises two modules (blue, the hippocampus and the amygdala; pink, the basal ganglia and the thalamus). The latter module was divided into two submodules that separated the basal ganglia (green) and the thalamus (orange) at scales II–IV. **d**, The subcortical network for which functional connectivity significantly associated with individual variation in the behavioral dimension characterizing tobacco use ($n = 958$). The linear association was significant for all atlas scales (scales I–IV, $P < 0.0001$, permutation test, $n = 10{,}000$, one-sided, as implemented in NBS). Nodes and edges are colored according to module allegiance. Thicker edges are associated with stronger effects. Each row shows results for a distinct atlas scale. **e**, Scatter plots and lines of best fit for the linear association between subcortical functional connectivity and individual variation in the dimension characterizing tobacco use. Each data point represents one of the 958 individuals. Lines of best fit are indicated with solid gray lines and dashed gray lines show 95% CI. *R* values are the Pearson correlation coefficients and *P* values assess the null hypothesis of zero correlation. Each row shows results for a distinct atlas scale. **f**, The subcortical network associated with tobacco use shown as a circular network for the scale IV atlas, highlighting the role of thalamus–striatum–hippocampal circuits. Nodes and connections are colored according to module allegiance. The histogram (gray bar) shown for each node indicates the nodal degree.

considered (Fig. 8a,b and Supplementary Fig. 21). These dimensions were derived from 109 behavioral items (Supplementary Table 4) using independent component analysis (Methods). Using the network-based statistic (NBS)[28], higher tobacco use was significantly associated with lower functional connectivity in the thalamus–caudate, the thalamus–NAc and the hippocampus–NAc

circuits (Fig. 8c–f). This is consistent with the involvement of the cortical–basal-ganglia–thalamus–cortical circuit and mesolimbic dopamine pathways in reward and goal-directed behaviors[29] such as smoking (Discussion). Although connectivity explained 4% of the individual variation in the behavioral dimension characterizing tobacco use, the relationship was highly consistent across all four parcellation scales (scale I: $r=-0.16$, $P=4.6\times10^{-7}$; scale II: $r=-0.17$, $P=7.1\times10^{-8}$; scale III: $r=-0.19$, $P=3.2\times10^{-9}$; scale IV: $r=-0.19$, $P=3.4\times10^{-9}$; Fig. 8e) and reproducible across fMRI sessions (Supplementary Fig. 22). The remaining four dimensions did not significantly relate to subcortical connectivity after correction for multiple testing.

Discussion

The preponderance of the research field toward the cerebral cortex, coupled with technical challenges inherent to subcortical imaging, has led to the subcortex remaining a terra incognita[30]. Through the provision of high-quality, ultrahigh field strength MRI data, collaborative initiatives such as the HCP[18] herald new opportunities to overcome some of the challenges associated with reliably imaging this deep and anatomically complex brain structure.

To address the incomplete representation of subcortical regions in many leading MRI atlases, we used multi-session MRI data of brain function acquired in more than 1,000 healthy adults participating in the HCP to provide fundamental new insight into the organizational architecture and connectivity of the human subcortex. Putative functional boundaries were recursively delineated at locations of sharp spatial change in subcortical-to-whole-brain functional connectivity fingerprints, which led to the semiautomated delineation of a new multiscale parcellation atlas. The new atlas is openly available as a standalone parcellation and has been integrated into established cortex-only[2,8] and cerebellar[14] atlases. Crucially, the new subcortical atlas was replicated using ultrahigh field strength MRI, and the delineated regions were functionally homogeneous and identifiable within individual participants, thus enabling the delineation of personalized atlases that accounted for individual variation in connectivity. We used the new atlas to link cortical networks with the subcortex and to uncover a robust and reproducible relationship between individual variation in behavioral dimensions and subcortical functional connectivity.

Reconciling parcellations and gradients. Topographic representations of variation in brain structure, function and connectivity usually follows one of two paradigms: (1) a mosaic of discrete parcels separated by hard boundaries[2,3,8,14] and, more recently, (2) gradients of continuous topographic variation[11–13,15,23,31]. We introduced a principled way forward to reconcile these two seemingly incompatible paradigms that resulted in a hybrid representation that parsimoniously accommodated both discrete boundaries and continuous gradients. Boundaries were only delineated at locations where gradient magnitudes were significantly larger than what would be expected under an appropriate null hypothesis. This was formalized with a model selection process and an accompanying null model to test the null hypothesis of an absence of discrete spatial boundaries at candidate locations. Without model selection, boundaries would be delineated for all gradient magnitude peaks, which could potentially result in spurious boundaries that represent geometric constrictions rather than sharp changes in connectional topography. The bow-tie constriction shown in Supplementary Fig. 5b provides an example of spurious boundary delineation, whereby a gradient peak appears at the bow knot, despite uniform spatial variation in connectivity. While the sheet-like geometry of the cerebral cortex is less likely to be affected by this phenomenon, complex subcortical geometry is more vulnerable to false-positive boundaries and over-parcellation.

Gradientography was developed to enable gradient visualization and accurate localization of candidate boundaries. Gradientography is a fMRI analog of diffusion MRI tractography and can be considered a form of fMRI tractography. Whereas tractography is conventionally used to map white matter fiber trajectories, gradientography, as introduced here, was used to map trajectories of maximum change in connectopic gradients. Future applications of gradientography are not limited to the subcortex, and the method could provide a bridge between cortical gradients and parcellations.

A new multiscale subcortical atlas. The new subcortical atlas comprises four scales, labeled scale I to scale IV, each providing a self-contained parcellation atlas. Additional scales were not warranted because the model selection procedure decided against boundary delineation for all regions constituting scale IV. Researchers seeking to use the new atlas should select the atlas scale that provides adequate resolution to address the research question at hand and approximately matches the scale of any cortical atlas that is used in terms of the total number of regions. While the new atlas is unimodal, rfMRI connectivity is argued to be one the most useful modalities in predicting areal boundaries[2].

Hierarchy is a hallmark of brain organization[32], and hierarchical organization is not necessarily a fait accompli of the recursive application of the model selection procedure used in this study (see "Further methodological considerations" in the Methods). Scale I of the atlas recapitulated subcortical regions delineated in existing MRI atlases, such as those available as part of Lead-DBS, Freesurfer and FSL. The position of gradient peaks enabled automated boundary demarcation among the amygdala, the hippocampus, the thalamus, the caudate, the putamen, the NAc and the GP. Discrete partition of the basal ganglia into the latter four regions and separation between the amygdala and the hippocampus were clearly supported by model selection. Contrastingly, the few existing connectivity-based subcortical atlases[6,33] generally do not preserve established anatomical boundaries[34], which suggests that there is structure–function divergence or methodological variation. Given that these broad subcortical regions are anatomically distinguishable, rudimentary alignment with functional architecture might be expected[1]. Indeed, this assumption motivates the parcellation of regions in isolation, such as the hippocampus[31,35], the striatum[23,36] and the thalamus[37], without first establishing functional boundaries between them. Despite correspondence between the principal connectional topography (gradient I) and anatomy at scale I, fine variation within specific structures was captured by the secondary connectional gradients. As shown in Supplementary Fig. 10b, the dominant striatal gradient separates the putamen, the NAc and the caudate, whereas the second and third gradients are organized along a dorsal-to-ventral axis, thus reflecting connectional variation between the dorsal (putamen and caudate) and ventral (NAc) striatum[38]. Similarly, the dominant thalamic gradient follows an anteroposterior trajectory, whereas the second and third gradients capture connectional variation between the mediodorsal and ventrolateral thalamus (Supplementary Fig. 10c), which may relate to organizational differences between subgroups of thalamic nuclei[39].

Scale II delineates functional divisions within each of the scale I regions. The NAc was partitioned into regions potentially representing its shell and core, and anterior (head) and posterior (body and tail) caudate subdivisions were identified, and the amygdala was divided into lateral and medial subregions. A striking divergence between anatomy and function was evident for the GP. The GP is anatomically divided along a mediolateral axis into the internal and external GP[34], whereas functional connectivity gradients indicated an anteroposterior partition. Scales III and IV differed only in the parcellation of the thalamus and the hippocampus.

Parcellation validation. While histological validation is the gold standard, functional parcellations do not necessarily converge toward anatomical or cytoarchitectural divisions to be deemed

useful[1]. To establish confidence in the new atlas, we first demonstrated that all atlas scales, except scale I, were significantly more homogenous[8] than expected due to chance. While scale I regions converged with subcortical anatomy, they encompassed multiple functional systems, thereby rendering scale I too coarse to achieve homogeneity. Homogeneity of functional signals within delineated subregions is a fundamental principle of functional parcellation[1]. Interestingly, existing structural connectivity and histological parcellations of the thalamus[37] and the hippocampus[22], respectively, were not functionally homogenous (Supplementary Fig. 12), which indicates that there is a need for modality-specific atlases.

Second, we reproduced the majority of regions using 7T fMRI. The improved contrast-to-noise ratio and spatial resolution of 7T compared with 3T MRI is particularly important for subcortical imaging[24]. Despite remarkable consistency between the 3T and 7T atlases, 7T fMRI enabled the delineation of finer structures. For example, a bipartite partition of the amygdala (medial and lateral) was delineated at 3T, whereas 7T fMRI aided further division of the medial amygdala into central and superior components, which broadly recapitulated the three cytoarchitectural zones of the amygdala[22]. Conversely, the NAc was consistently parcellated into two subdivisions at 3T and 7T, thus possibly representing the immunohistochemically distinct shell and core subdivision that is common to primates[40]. Therefore, while subcortical cartography is viable with functional connectivity inferred from high-quality 3T fMRI data, the level of detail that can be resolved at 7T is superior for several regions.

The subcortex atlas was personalized to account for individual connectivity differences. We showed that while the group-consensus atlas was sufficient to represent the majority of individuals, personalized subcortex atlases were warranted for a small proportion of others. Deviations from the group-consensus atlas were specific to particular individuals rather than specific regions (Supplementary Fig. 18), which suggests that deviations may capture inherent inter-individual differences in functional architecture. Further work is needed to understand whether individual variation in subcortical connectional topography relates to anatomy, task-evoked activation and behavior. Individual-specific atlases of the subcortex can potentially facilitate personalization of targeted therapies in psychiatry[41] and to improve prediction of behavior[26].

Dynamic and intrinsic subcortical topography. Using tfMRI, we identified subtle reconfigurations in the functional topography of the subcortex in response to changing cognitive demands. While a previous study[27] examined variation in parcel size in response to task stimuli, diversity curves enabled the characterization of changes in functional boundaries. During task conditions, the anterior putamen integrated more extensively with higher-order association networks, whereas the hippocampus segregated along its long axis, yielding a more integrated anterior hippocampal–amygdala system that may have served cognitive–emotional processing, and a more integrated posterior hippocampal–thalamus that potentially benefited learning and spatial memory[21]. These task-induced dynamic reconfigurations may reflect a process of neuronal recruitment into specific functional processing systems to cope with cognitive demands. The flexible reconfiguration of neural networks may largely depend on the intrinsic functional organization within local cortical circuits given that a task-positive to task-negative organizational axis was evident in the subcortex (Supplementary Fig. 20). However, context-dependent recruitment across intrinsic parcels, leading to the adjustment of functional boundaries, may additionally assist adaptive task responses. Further work is needed to understand task-evoked reconfigurations of functional boundaries, not only in the subcortex but also across the brain.

Tobacco and subcortical connectivity. We performed a hypothesis-free decomposition of HCP cognitive and behavioral measures into five latent dimensions to facilitate an investigation of relationships between brain connectivity and behavior. Previous HCP investigations typically investigate hundreds of individual measures, which leads to a formidable multiple testing problem, or preselect a limited number of specific behavioral domains or measures. The five latent dimensions identified here condense information from more than 100 individual measures.

The new atlas unveiled a novel relationship between individual variation in functional connectivity and the behavioral dimension characterizing tobacco use. Higher tobacco use was associated with lower thalamus–caudate, thalamus–NAc and hippocampus–NAc functional connectivity. The shell of the NAc and the adjacent ventroanterior caudate (CAU–VA) were most implicated (Fig. 8f). This accords with the primary role of the shell of the NAc and the anterior caudate in reward anticipation, incentive motivation and action[42,43]. The NAc integrates glutamatergic (excitatory) inputs from the hippocampus and the thalamus[44,45], which enables the regulation of dopamine neuron activity in the ventral tegmental area[45]. While nicotine, the main neuroactive compound of tobacco, stimulates dopamine release and transmission in the NAc, especially the medial shell[46], chronic nicotine administration yields neuronal damage in the hippocampus and the striatum[47,48], which potentially manifests in lower functional connectivity in the thalamus–striatum–hippocampus system. Therefore, the putative downregulation of the mesolimbic circuit by nicotine may provide insight into the increased prevalence of tobacco smoking in individuals with schizophrenia[49]. In particular, this downregulation effect of nicotine on the hyperdopaminergic state in the mesolimbic system may correspond with the upregulation effect of nicotine in the prefrontal cortex[50], which alleviates symptoms in the disorder.

Conclusions

We unveiled the complex topographic organization of the human subcortex and parsed this complexity into four scales constituting a new atlas for the brain's remaining terra incognita. The new atlas is scalable, personalizable, reproducible at ultrahigh field strengths and openly available. Future work will focus on the addition of structures such as subthalamic regions, the red nucleus and the hypothalamus. The new atlas can be incorporated with existing cortex-only atlases to aid the investigation of cortico–subcortical interactions and the mapping of holistic connectomes using MRI.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-020-00711-6.

Received: 29 November 2019; Accepted: 21 August 2020;
Published online: 28 September 2020

References

1. Eickhoff, S. B., Yeo, B. T. T. & Genon, S. Imaging-based parcellations of the human brain. *Nat. Rev. Neurosci.* **19**, 672–686 (2018).
2. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
3. Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
4. Shepherd, G. M. Corticostriatal connectivity and its role in disease. *Nat. Rev. Neurosci.* **14**, 278–291 (2013).
5. Choi, E. Y., Yeo, B. T. & Buckner, R. L. The organization of the human striatum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **108**, 2242–2263 (2012).
6. Ji, J. L. et al. Mapping the human brain's cortical–subcortical functional network organization. *NeuroImage* **185**, 35–57 (2019).
7. Cohen, A. L. et al. Defining functional areas in individual human brains using resting functional connectivity MRI. *NeuroImage* **41**, 45–57 (2008).

8. Gordon, E. M. et al. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* **26**, 288–303 (2016).
9. Schaefer, A. et al. Local–global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
10. Huntenburg, J. M., Bazin, P. L. & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).
11. Haak, K. V., Marquand, A. F. & Beckmann, C. F. Connectopic mapping with resting-state fMRI. *NeuroImage* **170**, 83–94 (2018).
12. Margulies, D. S. et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl Acad. Sci. USA* **113**, 12574–12579 (2016).
13. Guell, X., Schmahmann, J. D., Gabrieli, J. & Ghosh, S. S. Functional gradients of the cerebellum. *eLife* **7**, e36652 (2018).
14. King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. & Diedrichsen, J. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat. Neurosci.* **22**, 1371–1378 (2019).
15. Tian, Y. & Zalesky, A. Characterizing the functional connectivity diversity of the insula cortex: subregions, diversity curves and behavior. *NeuroImage* **183**, 716–733 (2018).
16. Coalson, T. S., Van Essen, D. C. & Glasser, M. F. The impact of traditional neuroimaging methods on the spatial localization of cortical areas. *Proc. Natl Acad. Sci. USA* **115**, E6356–E6365 (2018).
17. Fox, M. D. et al. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl Acad. Sci. USA* **102**, 9673–9678 (2005).
18. Van Essen, D. C. et al. The WU–Minn Human Connectome Project: an overview. *NeuroImage* **80**, 62–79 (2013).
19. Finn, E. S. et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664 (2015).
20. Müller, E. et al. Core and matrix thalamic sub-populations relate to spatio-temporal cortical connectivity gradients. *NeuroImage* **222**, 117224 (2020).
21. Poppenk, J., Evensmoen, H. R., Moscovitch, M. & Nadel, L. Long-axis specialization of the human hippocampus. *Trends Cogn. Sci.* **17**, 230–240 (2013).
22. Amunts, K. et al. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat. Embryol.* **210**, 343–352 (2005).
23. Marquand, A. F., Haak, K. V. & Beckmann, C. F. Functional corticostriatal connection topographies predict goal directed behaviour in humans. *Nat. Hum. Behav.* **1**, 0146 (2017).
24. Vu, A. et al. Tradeoffs in pushing the spatial resolution of fMRI for the 7T Human Connectome Project. *NeuroImage* **154**, 23–32 (2017).
25. Gordon, E. M. et al. Precision functional mapping of individual human brains. *Neuron* **95**, 791–807.e7 (2017).
26. Kong, R. et al. Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cereb. Cortex* **29**, 2533–2551 (2019).
27. Salehi, M. et al. There is no single functional atlas even for a single individual: functional parcel definitions change with task. *NeuroImage* **208**, 116366 (2020).
28. Zalesky, A., Fornito, A. & Bullmore, E. T. Network-based statistic: identifying differences in brain networks. *NeuroImage* **53**, 1197–1207 (2010).
29. Maia, T. V. & Frank, M. J. From reinforcement learning models to psychiatric and neurological disorders. *Nat. Neurosci.* **14**, 154–162 (2011).
30. Forstmann, B. U., de Hollander, G., van Maanen, L., Alkemade, A. & Keuken, M. C. Towards a mechanistic understanding of the human subcortex. *Nat. Rev. Neurosci.* **18**, 57–65 (2016).
31. Vos de Wael, R. et al. Anatomical and microstructural determinants of hippocampal subfield functional connectome embedding. *Proc. Natl Acad. Sci. USA* **115**, 10154–10159 (2018).
32. Churchland, P. S. & Sejnowski, T. J. Perspectives on cognitive neuroscience. *Science* **242**, 741–745 (1988).
33. Fan, L. et al. The Human Brainnetome Atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
34. Pauli, W. M., Nili, A. N. & Tyszka, J. M. A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Sci. Data* **5**, 180063 (2018).
35. Plachti, A. et al. Multimodal parcellations and extensive behavioral profiling tackling the hippocampus gradient. *Cereb. Cortex* **29**, 4595–4612 (2019).
36. Janssen, R. J., Jylänki, P., Kessels, R. P. C. & van Gerven, M. A. J. Probabilistic model-based functional parcellation reveals a robust, fine-grained subdivision of the striatum. *NeuroImage* **119**, 398–405 (2015).
37. Behrens, T. E. et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci.* **6**, 750–757 (2003).
38. Haber, S. N. & Knutson, B. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* **35**, 4–26 (2010).
39. Morel, A., Magnin, M. & Jeanmonod, D. Multiarchitectonic and stereotactic atlas of the human thalamus. *J. Comp. Neurol.* **387**, 588–630 (1997).
40. Meredith, G. E., Pattiselanno, A., Groenewegen, H. J. & Haber, S. N. Shell and core in monkey and human nucleus accumbens identified with antibodies to calbindin-D28k. *J. Comp. Neurol.* **365**, 628–639 (1996).
41. Cocchi, L. & Zalesky, A. Personalized transcranial magnetic stimulation in psychiatry. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 731–741 (2018).
42. Castro, D. C. & Bruchas, M. R. A motivational and neuropeptidergic hub: anatomical and functional diversity within the nucleus accumbens shell. *Neuron* **102**, 529–552 (2019).
43. Pauli, W. M., O'Reilly, R. C., Yarkoni, T. & Wager, T. D. Regional specialization within the human striatum for diverse psychological functions. *Proc. Natl Acad. Sci. USA* **113**, 1907–1912 (2016).
44. MacAskill, A. F., Little, J. P., Cassel, J. M. & Carter, A. G. Subcellular connectivity underlies pathway-specific signaling in the nucleus accumbens. *Nat. Neurosci.* **15**, 1624–1626 (2012).
45. Perez, S. M. & Lodge, D. J. Convergent inputs from the hippocampus and thalamus to the nucleus accumbens regulate dopamine neuron activity. *J. Neurosci.* **38**, 10607–10618 (2018).
46. Pontieri, F. E., Tanda, G., Orzi, F. & Di Chiara, G. Effects of nicotine on the nucleus accumbens and similarity to those of addictive drugs. *Nature* **382**, 255–257 (1996).
47. Gallinat, J. et al. Abnormal hippocampal neurochemistry in smokers: evidence from proton magnetic resonance spectroscopy at 3T. *J. Clin. Psychopharmacol.* **27**, 80–84 (2007).
48. Xu, Z., Seidler, F. J., Ali, S. F., Slikker, W. Jr. & Slotkin, T. A. Fetal and adolescent nicotine administration: effects on CNS serotonergic systems. *Brain Res.* **914**, 166–178 (2001).
49. de Leon, J. & Diaz, F. J. A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophrenia Res.* **76**, 135–157 (2005).
50. Koukouli, F. et al. Nicotine reverses hypofrontality in animal models of addiction and schizophrenia. *Nat. Med.* **23**, 347–354 (2017).

Methods

Data and preprocessing. *HCP dataset.* Human neuroimaging data acquired as part of the HCP S1200 release[18] were analyzed here. Participants were healthy young adults ranging between 22 and 37 years old. Some participants were genetically related. These relationships were controlled in statistical analyses by constraining permutations[51,52]. HCP datasets used in this study include three main imaging sessions. (1) Two sessions (REST1 and REST2) of rfMRI) acquired using multiband echo planar imaging (EPI) on a customized Siemens 3T MR scanner (Skyra system), where each session comprised two runs (left-to-right and right-to-left phase encoding) of 14 min and 33 s each (repetition time (TR) = 720 ms, echo time (TE) = 33.1 ms, voxel dimension: 2-mm isotropic). The two runs were temporally concatenated for each session, yielding 29 min and 6 s of data in each session. Concatenation of the two different phase-encoded data ensured that any potential effect of phase encoding on gradient direction was counterbalanced by the opposing phase encoding. (2) tfMRI data acquired using the identical multiband EPI sequence as the rfMRI session; however, the run duration ranged between 2 and 5 min depending on the specific task (7 tasks). (3) rfMRI acquired on a Siemens Magnetom 7T MR scanner using a multiband acquisition (TR = 1,000 ms, TE = 22.2 ms, voxel dimension: 1.6-mm isotropic), where participants completed 4 runs of approximately 16 min each. The first two runs (posterior-to-anterior and anterior-to-posterior phase encoding) were used in this study. We refer to this dataset as 7T-rfMRI. Details pertaining to the acquisition of rfMRI[24,53] and tfMRI[54] are described elsewhere. The seven tfMRI tasks were chosen to tap a broad range of cognitive and affective processes and activated a wide range of neural systems[54]. Participants that had completed at least two runs in each dataset were included. As a consequence, 33 of the original 1,113 individuals were excluded from the primary dataset (REST1) and 92 individuals were excluded from the replication dataset (REST2). Supplementary Table 1 shows the sample size and basic participant demographics for each dataset.

The second 3T-rfMRI session (REST2) was used for replication. Minimally preprocessed volumetric rfMRI and tfMRI data were sourced from the online HCP repository. Details of the minimal preprocessing pipeline can be found elsewhere[55]. The pipeline includes correction of head motion and spatially structured physiological noise with ICA+FIX[56,57]. ICA-based denoising methods outperform, or equal, the performance of alternative approaches on most benchmarks[58]. After eliminating spatial artifacts and distortions, the functional images were spatially aligned to the Montreal Neurological Institute (MNI) standard space using the FNIRT nonlinear registration algorithm. MNI standard space herein always refers to the MNI ICBM 152 nonlinear sixth generation stereotaxic registration model. Additional nuisance regression and temporal filtering were not performed as part of the pipeline. However, head motion registration parameters estimated as part of spatial alignment were curated, reduced to a single summary measure known as framewise displacement[59] and included as a confound in statistical analyses to control for individual variation in head motion[60,61].

The subcortical SNR and sensitivity of the blood-oxygenation-level-dependent (BOLD) signal was lower compared with the cortex because T2* signal decay times are shorter in subcortical regions compared with cortical regions due to iron enrichment of subcortical tissue[62]. Therefore, to enhance subcortical SNR and BOLD sensitivity, the minimally preprocessed data were subjected to two additional operations (Supplementary Fig. 8a). First, spatial smoothing was performed with a Gaussian smoothing kernel of 6-mm full width at half maximum (FWHM) for 3T images and 4-mm FWHM for 7T images. This choice of kernel width was guided by evaluations of gradient magnitude images derived from a range of smoothing kernels (4, 6 and 8 mm) and filtering options (that is, the Wishart filter, see below). Second, to further improve subcortical SNR and to suppress the effects of unstructured and autocorrelated noise that share similar temporal and spatial properties, the fMRI time series were reconstructed from a principal component analysis (PCA), whereby the components were filtered according to an empirically fitted Wishart distribution to dampen the effects of noise[2,63,64]. This PCA-based data reconstruction (that is, the Wishart filter) has been previously used in a multimodal cortical parcellation study[2] and comprehensively documented by Glasser and colleagues[63,64]. The Wishart filter was performed over all gray matter voxels for each individual separately in all datasets. In addition, task block regressors were created for each task condition based on the time of onset, the duration and the relative magnitude of each stimulus, and then convolved with hemodynamic response functions (spm_hrf.m function from SPM12, https://www.fil.ion.ucl.ac.uk/spm). The relevant regressors corresponding to each task were then regressed from the tfMRI data to ensure that functional connectivity was disambiguated from the potential confound of task co-activation effects[65]. The residuals remaining from this regression were used for all subsequent analyses.

Independent validation dataset. An independent fMRI dataset was acquired in Australia to test the reproducibility of the parcellation homogeneity results. Ten healthy adults were recruited and confirmed not to have any current or history of any neurological or psychiatric disorders at the time of MRI scanning (mean age 26 ± 2.1 years, 6 males). Imaging was performed in all participants using a Siemens Prisma 3T MRI scanner. Structural images of brain anatomy were acquired using an optimized magnetization-prepared rapid acquisition gradient echo (MPRAGE) T1-weighted sequence with 176 sagittal slices of 1-mm thickness, field of view of 240 × 256 mm^2, flip angle of 9°, TR of 1,900 ms, TE of 2.98 ms and voxel size of 1.0 × 1.0 × 1.0 mm^3. rfMRI was acquired using a T2*-weighted multiband gradient-echo EPI sequence of approximately 12 min, resulting in 880 volumes (TR = 810 ms, TE = 30 ms, voxel size = 2.0 × 2.0 × 2.0 mm^3).

rfMRI data were preprocessed using the fMRIPrep-1.5.9 pipeline[66]. In brief, each T1-weighted volume was corrected for intensity non-uniformity and skull-stripped. The extracted brain was spatially normalized to MNI standard space through nonlinear registration using ANTs. Cortical surfaces were reconstructed using FreeSurfer. Functional data were motion-corrected using mcflirt (FSL) followed by co-registration to the corresponding T1-weighted image using boundary-based registration with six degrees of freedom. Motion-correcting transformations, BOLD-to-T1w transformation and T1w-to-MNI warp were concatenated and applied in a single step using antsApplyTransforms (ANTs).

Confounds including the 24 head-motion parameters (6 basic motion parameters + 6 temporal derivatives + 12 quadratic terms and their 6 temporal derivatives), mean white matter and cerebrospinal fluid signals were computed and regressed from the preprocessed fMRI data for each individual. The residuals of this regression were then subjected to spatial smoothing with a Gaussian-smoothing kernel of 6-mm FWHM and Wishart filtering[63,64].

For all datasets, statistical methods were not used to predetermine sample sizes. All relevant and appropriate data provisioned by the HCP were included to maximize samples. Our use of HCP datasets and our recruitment of the independent dataset complied with the Australian NHMRC national framework for the conduct of human research. Informed consent was obtained from all the participants. Data collection and analyses were not performed blind to the conditions of the experiments.

Mask delineation. A binary subcortex mask comprising the left and right basal ganglia, the thalamus, the hippocampus and the amygdala was delineated using three steps. First, the probabilistic Harvard–Oxford Subcortical Structural Atlas (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases) was used to delineate a binary mask for each of the following subcortical regions: the thalamus, the caudate, the putamen, the NAc, the GP, the hippocampus and the amygdala. Voxels with a probability of 50% or more of belonging to one of these regions were included in the relevant mask, except for the GP. A higher threshold (60%) was applied to the GP to minimize potential white matter contamination of the fMRI signal due to the numerous myelinated axons of the striato–pallido–nigral bundle abutting the periphery of this region[67]. Second, the masks for each individual subcortical region were merged to yield a binary mask for the entire subcortex. Finally, the subcortex mask was left–right symmetrized. Left–right symmetrization accords with previous subcortical atlas literature[34,68] and enabled symmetrization of subsequent computations to improve the SNR. Symmetrization is not an unprecedented step given that intra-hemispheric asymmetries in the volume of subcortical nuclei are less evident in healthy young adults compared with cortical asymmetries[69]. Mask symmetrization was achieved by reflecting the left hemisphere of the mask around the mid-sagittal plane and computing the intersection between the right hemisphere of the mask and this reflection. Voxels constituting the intersection yielded a left subcortex mask, which was merged with its reflection to yield the final left–right symmetric subcortex mask. The final mask comprised a total of $N = 7{,}984$ voxels (63.87 cm^3). The subthalamic nuclei, the substantia nigra and comparably small subcortical nuclei were not included because the spatial resolution of the fMRI data was deemed insufficient to accurately resolve putative substructures within these nuclei (for example, the pars reticulata and compacta). Submillimeter fMRI and multi-echo EPI would benefit the investigation of the functional architecture of these exceedingly small nuclei[70]. A binary gray matter mask was delineated using the MNI152 probabilistic gray matter atlas. Voxels belonging to cortical gray matter were included in the gray matter mask as well as any voxels constituting the final subcortex mask. fMRI data were consistently absent for a majority of individuals in a small proportion of voxels constituting the gray matter mask. After eliminating these voxels, the final gray matter mask comprised $M = 164{,}360$ voxels (1,314.9 cm^3), which included the midbrain, the pons and the cerebellum. The subcortex and gray matter masks were delineated in 2-mm isotropic space and were used for the 3T fMRI data. To suit the 7T-rfMRI data, which was in 1.6-mm isotropic space, the masks were resliced to this resolution and median filtered to smooth the mask edges.

Functional connectivity and eigenmaps. Whole-brain functional connectivity was mapped for each subcortex voxel. Spatial gradients in the resulting maps were then computed to yield a continuous representation of functional connectivity variation across the subcortex. Connectopic mapping[11] was used to map spatial gradients for each individual, which involved computing a sequence of eigendecompositions to yield a Laplacian eigenmap[71]. Specifically, following the procedure described by Haak and colleagues[11], the temporally concatenated fMRI signals were represented for each individual in a matrix of dimension $T \times M$, where T denotes the number of time frames and M denotes the number of gray matter voxels. PCA was used to reduce the dimensionality of this matrix to $T \times (T-1)$. The fMRI signal at each subcortical voxel was then correlated (Pearson correlation) with each column of the PCA-transformed matrix, resulting in a connectivity matrix of dimension $N \times (T-1)$, where N denotes the number of subcortex voxels. Correlation

coefficients were r-to-z transformed using the Fisher transformation. Each row of this matrix provided a connectional fingerprint[19] for a particular subcortical voxel in the PCA-transformed space. Similarity in the connectional fingerprints between each pair of subcortical voxels was then quantified with the η^2 coefficient, which results in a symmetric matrix of dimension $N \times N$ for every individual. The similarity matrix was transformed into a sparse graph using the weighted adjacency matrix, whereby the weights of all connections with a Euclidean distance less than ε were set to zero. The connection density was determined by the smallest value of ε that ensured a connected graph[11,72,73]. This threshold varied between individuals, yielding graphs with a connection density of 0.4% on average. The Laplacian matrix, L, was then computed according to $L = D - W$, where D denotes the diagonal matrix of node strengths and W denotes the sparse adjacency matrix. In particular, $D_{ij} = \sum W_{ij}$ if $i = j$, otherwise $D_{ij} = 0$. Finally, the eigenvectors and eigenvalues of the Laplacian matrix were computed. The smallest eigenvalue was necessarily zero, while all other eigenvalues were positive due to the connectedness of the graph. The eigenvector with a zero eigenvalue was a constant and discarded. The eigenvectors with the second, third and fourth smallest eigenvalues were referred to as gradients I, II and III respectively. These steps are shown in Fig. 1a. Each of the three gradients, or eigenmaps, characterized a continuous mode of spatial variation in functional connectivity across the spatial extent of the subcortex. We use the terms eigenmap and gradient interchangeably. Eigenvectors corresponding to the next largest eigenvalues typically explained substantially less variance in the Laplacian matrix than the first four eigenvectors (Supplementary Fig. 2) and were thus given no further consideration here. The N values defining each gradient were projected onto the three-dimensional (3D) anatomy of the subcortex in MNI standard space for visualization and further analyses. Unless otherwise specified, the same process described above was used to estimate gradients for each individual and for each imaging modality, including 3T-rfMRI, 7T-rfMRI and tfMRI.

Gradientography. Tractography is an established 3D modeling technique to visually represent axonal fiber bundles using diffusion MRI[74]. We developed an analogous technique for fMRI called gradientography to characterize spatial gradients in functional connectivity. Gradientography was performed at the group level using a group-consensus representation of each eigenmap. In particular, similarity matrices were averaged across all individuals, and eigenmaps were computed for the group-averaged similarity matrix using the same process described above. The resulting group-consensus eigenmaps were projected onto the 3D anatomy of the subcortex in MNI standard space (Fig. 1b, column I and Supplementary Fig. 1, column I). Using the Sobel gradient operator, directional gradients were computed for each subcortical voxel constituting the 3D image representing each eigenmap. Each subcortical voxel was therefore endowed with a gradient directed along each of the three canonical axes as well as a gradient magnitude (Fig. 1b, column II and Supplementary Fig. 1, column II). We denote the gradient direction and magnitude at a given voxel with $g = \left[g_x, g_y, g_z\right]^T$ and $|g|$, respectively. The gradient magnitude indicated how rapidly the functional connectivity of a local area of subcortex changed per unit length. Before application of the Sobel operator, the peripheral boundary of the subcortex in each eigenmap image was dilated by approximately one voxel to diminish spurious gradients from appearing along the direction that is perpendicular to the boundary. Erosion was performed after application of the Sobel operator to reverse this dilation and restore the original volume of the subcortex. The gradient directions and magnitudes were approximately mirror symmetric in the sagittal plane (Supplementary Fig. 3a). Therefore, to enhance the SNR, the gradient directions were left–right symmetrized (to be consistent with the symmetrization of the subcortex mask). In particular, gradient directions in the left hemisphere were reflected into the right hemisphere around the mid-sagittal plane and averaged with the corresponding directions in the right hemisphere. These averaged gradient directions were then reflected back into the left hemisphere, and gradient magnitudes were recomputed (Supplementary Fig. 3b). This was performed separately for each of the three eigenmaps, and the difference in the gradient direction before and after symmetrization was quantified (gradient I: $\theta = 5.42 \pm 4.98°$; gradient II: $\theta = 5.04 \pm 5.0°$; gradient III: $\theta = 4.95 \pm 5.36°$).

We developed the new gradientography procedure to provide a principled framework to visualize and analyze spatial gradients in functional connectivity. First, 3D tensors were fitted to each subcortical voxel using the constrained tensor model[75,76], which is synonymous with diffusion tensor imaging. The constrained tensor model is given by

$$D = \beta g g^T + \alpha I$$

where D denotes a 3×3 symmetric tensor, g is a column vector denoting the normalized gradient direction computed with the Sobel operator, I is the identity matrix, and β and α are parameters that quantify the strength of anisotropic and isotropic diffusion, respectively. Here, we set $\beta = Kc|g|$ and $\alpha = K(1 - c)$, where $c = 0.995$ is a weighting factor that determines the extent of tensor anisotropy and K is a scale factor that governs the tensor size for visualization purposes. In this way, cigar-shaped tensors indicated rapid change in subcortical functional connectivity along the orientation of the longest axis of the tensor. In contrast, spherical tensors indicated functional connectivity that was constant along all spatial directions. It is important to note that tractography was only dependent on the gradient direction, g. The choice of β, α and c only influenced visualization of the tensor field (Fig. 1b, column III and Supplementary Fig. 1, column III). While tensors were separately fitted for each of the three eigenmaps, an alternative methodology would have been to fit a bi-tensor or multi-tensor model to each voxel[77], enabling representation of multiple eigenmaps within a single tensor field. This is conceptually analogous to modeling crossing fibers in diffusion MRI tractography[78]. However, the potential utility of modeling interactions between crossing gradients in this way remains unclear; thus, each eigenmap was independently analyzed. We relegate multidirectional gradientography to future work. Streamline propagation was performed using the interpolated streamline method with a step length of 1 mm, angle threshold of 30° and seeding 20 streamlines from randomly chosen coordinates within each subcortical voxel, yielding a total of approximately 15,000 streamlines. Streamline propagation was terminated when reaching the subcortex mask boundary. Gradientography was performed using Diffusion Toolkit software and visualized with TrackVis[79]. Streamlines were virtually unchanged when left–right symmetrization of the gradient directions was not performed. All streamlines were permitted to follow curvilinear trajectories of arbitrary complexity. The gradient magnitude image was projected onto the streamlines to parameterize the magnitude of change in functional connectivity and to visualize transitions in the gradient field with respect to the curvature of streamlines. In particular, each spatial coordinate constituting a streamline was assigned a gradient magnitude based on trilinear interpolation of the gradient magnitude image (Fig. 1b, right most). Gradientography was also performed for gradients II and III using the same procedure (Supplementary Fig. 1). While streamline trajectories varied among gradients I, II and III, the locations of local maxima in the gradient magnitude images of the three gradients were consistent (Supplementary Figs. 1 and 3). The gradient magnitude images for gradients II and III were spatially correlated (Spearman correlation) to gradient I (gradient I/II: $r = 0.44$, gradient I/III: $r = 0.73$); thus, only gradient I was used to delineate functional boundaries.

We also observed that head motion (framewise displacement (FD)) showed minimal effect on the group-consensus connectivity gradients. Specifically, individuals were ranked according to head motion, and individuals in the top 20% (mean FD: 0.26 ± 0.06 mm) and bottom 20% (mean FD: 0.10 ± 0.01 mm) defined two distinct subgroups. Group-consensus gradient magnitude images were computed for each of two subgroups and compared with the group-consensus gradient magnitude image generated within all individuals. We found that the group-consensus gradient magnitude images were strongly correlated between each pair of the three groups (Spearman correlation, $r > 0.9$).

To preclude streamlines from traversing the slender anatomical gap between the tail of the caudate nucleus and the most anterior extent of the thalamus, the subcortex mask was separated at this precise location, giving rise to three spatially contiguous subcortical components: (1) the left basal ganglia, which comprised the striatum and the GP; (2) the right basal ganglia; and (3) the left and right thalamus, the hippocampus and the amygdala. Because these three components were spatially discontiguous, tractography was performed separately for each component. A boundary between the caudate and the thalamus was therefore imposed a priori, consistent with established anatomical knowledge. Due to left–right symmetrization of gradient directions, streamlines for the left and right basal ganglia were mirror symmetric; thus, it was sufficient to analyze only one of the hemispheres of the basal ganglia. We refer to the streamlines traversing the basal ganglia as the dorsal streamlines and the dorsal connectional topography (Fig. 2a), while collectively referring to the streamlines traversing the left and right thalamus, the hippocampus and the amygdala as the ventral streamlines and ventral connectional topography (Fig. 2b). Permitting streamlines to traverse the caudate–thalamus gap did not alter the location of local maxima in gradient magnitudes, although they were less prominent in this case (Fig. 1b).

Diversity curves. The dorsal and ventral group of streamlines each comprised hundreds of distinct streamlines that characterized the connectional topographies associated with these regions. Streamlines shorter than 60 mm and 160 mm were discarded from the dorsal and ventral group, respectively. The remaining streamlines within each group traversed trajectories that shared similar profiles; thus, it was feasible to determine a single representative streamline for each group. To determine a representative streamline, the mean closest point distance[80] was computed between all pairs of streamlines constituting a group. In particular, the distance from one streamline to another was determined by first computing the distance between each point on the first streamline to all points on the second streamline. The minimum distance was then determined for each point on the first streamline. The distance from the first to second streamline was defined as the mean of these minimum distances. This was repeated for all pairs of streamline to generate a $J \times J$ distance matrix, where J denotes the number of streamlines. The distance matrix was symmetrized in a way that the mean of the two distances between a pair of streamlines was used[80]. The representative streamline was chosen as the streamline with the shortest distance, on average, to all other streamlines. Having defined a representative streamline for the dorsal and ventral group of streamlines, it was then possible to compute a diversity curve with respect to each of the representative streamlines. This yielded a curvilinear parameterization of

the group-consensus eigenmaps and their gradient magnitudes, thereby enabling statistical inference. The diversity curve concept is described in detail elsewhere[15].

Diversity curves were mapped in three steps: (1) projection of the gradient magnitude image onto each streamline; (2) spatial alignment of streamlines to the representative streamline using dynamic time warping[81]; and (3) averaging across the aligned streamlines to yield a group-representative diversity curve to provide a basis for statistical inference. Specifically, each spatial coordinate constituting a streamline was assigned a gradient magnitude based on trilinear interpolation of the gradient magnitude image. While each streamline comprised the same number of equidistant coordinates, the exact lengths of streamlines varied, resulting in some degree of misalignment between streamlines in terms of coordinate-to-anatomy correspondence. For example, the tenth coordinate on a given streamline was not necessarily closest to the tenth coordinate on a neighboring streamline. Thus, averaging across a set of streamlines could not be performed based on correspondence in the index of coordinates. Dynamic time warping was used to spatially align individual streamlines to the representative streamline and to facilitate averaging of any measure parameterized along the length of each individual streamline.

Dynamic time warping is typically used to align signals in time, whereas the method was used here to achieve spatial alignment. A warping path was computed for each individual streamline such that the sum of the Euclidean distance was minimal between the warped and representative streamline. This involved either stretching or compressing coordinates along the trajectory of each individual streamline to achieve alignment with the representative streamline. Alignment quality was visually assessed, and streamlines with a warp distance exceeding the averaged warping distance over all streamlines (dorsal: 1.72 mm, ventral: 0.54 mm) were omitted. This yielded a total of 4,693 and 3,862 aligned streamlines for the dorsal and ventral group, respectively. The gradient (eigenmap) and gradient magnitude images projected onto each of these streamlines were then averaged for each point along the streamline trajectory, yielding a single gradient and gradient magnitude diversity curve for the dorsal (Fig. 2c) and ventral group (Fig. 2d). Gradient magnitude peaks in the diversity curves indicated putative functional boundaries between regions. Consistent with the gradient magnitude image across the entire subcortex, the gradient magnitude images for gradients II and III were also spatially correlated (Spearman correlation) to gradient I for both dorsal (gradient I/II: $r=0.79$, gradient I/III: $r=0.79$) and ventral (gradient I/II: $r=0.80$, gradient I/III: $r=0.54$) components. Statistical inference was therefore confined to the dominant gradient (gradient I).

Focusing on the dominant gradient is not without precedent, and this gradient shows good correspondence with underlying neuroanatomy[23], although non-dominant gradients can reveal complementary connectivity topographies[11,23]. Three local maxima were evident in the dorsal streamline group, corresponding to established neuroanatomical boundaries among the GP–putamen, the putamen–NAc and the NAc–caudate (Fig. 2c). For the ventral streamline group, nine local maxima were evident in the diversity curve representing the gradient magnitude, which suggests that there are putative functional boundaries among the amygdala–anterior hippocampus, the anterior–posterior hippocampus, the posterior hippocampus–posterior thalamus, the anterior–posterior thalamus and the left–right thalamus (Fig. 2d).

Null model. Although local maxima in the gradient magnitude images or the diversity curves representing these images indicated putative functional boundaries, these maxima could have also potentially arisen as a matter of random fluctuations reflecting inter-individual variation and finite sample effects or due to the geometry of the subcortex (Supplementary Fig. 5). Therefore, null data were generated to statistically determine whether local maxima in the gradient images were larger than expected due to these random (sample) effects and/or confounds. The null hypothesis was of a spatially homogeneous gradient magnitude such that the rate at which whole-brain connectivity changed per unit length was constant at all points in the subcortex. Under this null hypothesis, any spatial heterogeneity was assumed to be due to sample effects or geometry. Functional boundaries were deemed to present at local maxima in the empirical gradient magnitude images when this null hypothesis could be locally rejected. Null data were generated by rewiring edges in the graph defined by the sparse adjacency matrix W. As detailed above (see "Functional connectivity and eigenmaps"), W defines a graph in which each node corresponds to a subcortical voxel, and the edge weights quantify similarity in functional connectivity profiles. A rewiring algorithm was developed to randomly reposition each edge, subject to the constraint that edges were more likely to be placed between pairs of voxels in close spatial proximity. This constraint was imposed to preserve the effect of smoothing in the null data. The following methodology was used to determine where edges should be repositioned to respect fMRI smoothing and subcortex geometry.

First, the fMRI data for each subcortical voxel was replaced with independent and randomly sampled Gaussian data that were matched in length and amplitude distribution. Second, the Gaussian data were smoothed using the same FWHM as the fMRI data. Third, the Pearson correlation coefficient was computed using the smoothed Gaussian data between all pairs of subcortical voxels. Finally, all voxel pairs were ranked from highest to lowest on the basis of the correlation coefficient. At each iteration of the rewiring algorithm, the edge being rewired was repositioned to the highest-ranking pair of voxels that had not yet received a rewired edge. To ensure that the rewired graphs were not fragmented, the first $N-1$ edges were repositioned according to a minimum spanning tree (MST), where the MST was computed for a lattice graph in which each node corresponded to a subcortical voxel. Edges in the rewired graphs inherited their edge weights from W. An ensemble of 100 rewired graphs and corresponding sparse adjacency matrices $\bar{W}_1, \bar{W}_2, \ldots, \bar{W}_{100}$ were generated by repeatedly initializing this rewiring procedure with randomly sampled Gaussian data. We found that repositioning edges to randomly chosen pairs of spatially neighboring voxels yielded virtually identical null data.

Supplementary Fig. 5a shows the graph corresponding to W and $\bar{W}_1$ for an axial slice that exemplifies the functional boundary between the GP and the putamen. Very few edges in the graph for W intersect this boundary, whereas the graph for $\bar{W}_1$ comprises numerous rewired edges that interconnect the GP and the putamen. Consistent with anatomy, this suggests that the strong gradient in functional connectivity that yields the boundary between these two regions is most likely not due to chance or the effect of smoothing and/or subcortex geometry. Supplementary Fig. 5b shows an example that demonstrates the importance of preserving geometry in the null data. In the bow-tie-shaped region constituting this example, binary edges are randomly positioned between neighboring pixels constituting the bow tie. This positioning of edges is consistent with the null hypothesis of a spatially homogenous gradient magnitude. However, the bow knot is the location of a geometric constriction, and if this geometry is not preserved in the null data, a specious boundary between the left and right sides of the bow is suggested.

Laplacian eigenmaps and gradient magnitude images were computed for each of $\bar{W}_1, \bar{W}_2, \ldots, \bar{W}_{100}$ by following the same procedure described above for W (see "Functional connectivity and eigenmaps"). An ensemble of 100 diversity curves were then computed by projecting each of these gradient magnitude images onto the original streamlines that were computed with respect to W. These diversity curves were consistent with the null hypothesis and we refer to them as null diversity curves (Fig. 2c,d).

Boundary delineation. Statistical inference was performed to determine whether each local maximum in the gradient magnitude images was sufficiently large to warrant delineation of a discrete functional boundary. If the null hypothesis could not be rejected and thus delineation was not warranted, connectional topography was deemed to be most parsimoniously represented as a continuum. The decision to delineate a boundary or not was informed by statistical testing of the null hypothesis performed on both the gradient magnitude images and the diversity curves onto which these images were projected. All local maxima in the gradient magnitude images were considered as candidate boundaries, but candidate boundaries were delineated only if the null hypothesis was rejected; that is, the gradient magnitude was larger than expected due to chance and the effects of smoothing and subcortex geometry. In particular, a P value for each local maximum on a diversity curve was estimated as the proportion of null diversity curves with local maxima that (1) exceeded or equaled the observed maximum and (2) were located within the same vicinity as the observed maximum. The FDR was controlled at a threshold of 5% across the set of all maxima. The decision to delineate a boundary was supported for maxima with P values that survived FDR correction. Figure 2c,d shows examples of diversity curve maxima for which the null hypothesis was rejected (denoted by blue asterisks).

The null hypothesis could also be tested with respect to the gradient magnitude images. In this case, the Kolmogorov–Smirnov (KS) test was used to assess the null hypothesis of equality in the distribution across voxels between the observed gradient magnitudes and the null gradient magnitudes. The null hypothesis was rejected if the tail of the distribution of gradient magnitudes was longer in the observed data compared with the null data. The main reasons for reverting to the KS test at finer scales were that (1) generally, only one candidate boundary was evident within each region constituting scales II–IV (Supplementary Fig. 23) and (2) we found that mapping diversity curves for relatively small regions did not provide substantial additional insight. Diversity curves and gradientography yield most insight when applied across a relatively large expanse that comprises multiple candidate boundaries (that is, scale I). Supplementary Fig. 23 shows examples of using the KS test to determine whether or not to delineate a candidate boundary in two specific subcortical regions. A strong gradient is evident in the anterior putamen (Supplementary Fig. 23a, yellow strip), and the KS test indicates that the gradient magnitude distribution across the voxels of this region (green curve) is significantly longer tailed than the null data (gray curves, $P<0.01$). Therefore, a boundary was delineated that resulted in the parcellation of the anterior putamen into one dorsal and one ventral component. In contrast, the null hypothesis could not be rejected with the KS test for the NAc-shell (Supplementary Fig. 23b, $P=0.22$); thus, subdivision of this region was not warranted despite the presence of a candidate boundary (yellow strip).

The final decision to delineate a boundary was not exclusively determined by whether the null hypothesis could be rejected or not. Other criteria that informed this decision were as follows:

1. Size criterion: a candidate boundary was not delineated if at least one of the regions resulting from boundary delineation would have comprised fewer than 100 voxels, irrespective of whether the null hypothesis was rejected.

2. Prior knowledge: a candidate boundary that corresponded with established subcortex anatomy or cytoarchitecture was delineated, irrespective of whether the null hypothesis was rejected.
3. Inter-hemispheric homolog: a candidate boundary was not delineated if the null hypothesis could not be rejected for the corresponding contralateral boundary. Edges in the rewired graphs (null data) inherited their edge weights from W, which was not necessarily symmetric. Therefore, the null data were not symmetrized; thus, discrepancies in statistical inference were possible between contralateral boundaries.

These criteria were only ever used for three boundaries of the more than 50 decisions pertaining to boundary delineation (Supplementary Fig. 6). In particular, the size criterion prevented further segmentation of a region of the caudate (CAU-DA), the prior knowledge criterion was used to delineate the thalamus–hippocampus boundary at scale I, despite failure to reject the null hypothesis, and the homolog criterion was used for the medial amygdala (mAMY) at scale II. These criteria were not used for any other cases, and all other boundary delineations were supported by the model selection procedure.

Hierarchical parcellation. The subcortex was parcellated into distinct functional regions in a hierarchical manner. Each successive scale of the hierarchy involved delineation of additional boundaries that conveyed finer details in connectional topography. Foremost, boundaries were delineated on the basis of maxima in the dorsal and ventral diversity curves for which the null hypothesis was rejected. This resulted in a subcortex parcellation comprising eight bilateral regions that recapitulated well-known subcortical regions and established anatomical boundaries (Fig. 3a). These eight bilateral regions defined scale I of the hierarchical parcellation (16 regions in total). Scale I is not intended to provide a novel parcellation, but rather establishes the validity of the model selection procedure and represents a necessary first step to delineate the novel scale II–IV atlases at finer scales. The next scale of the hierarchy (scale II) was delineated by separately computing Laplacian eigenmaps and gradient magnitudes for each of the 16 regions using the same procedure that was initially applied to the entire subcortex. This process was repeated in a recursive manner until the null hypothesis could not be rejected for any candidate boundaries, a condition that was satisfied at scale IV of the hierarchy (Figs. 3b and 4). A given region in scale k appeared as either two or more subregions in scale $k+1$, if the null hypothesis was rejected, otherwise it appeared unaltered. For each scale, the FDR was controlled at 5% across the family of null hypothesis tests performed for all candidate boundaries. Unlike alternative multiple testing corrections, the FDR is scalable and adaptive[82], meaning that the greater number of candidate boundaries tested at scale IV, compared with scale I, did not make FDR correction at scale IV inherently more stringent. Note that the FDR test was performed independently for each scale. Successive scales of the parcellation hierarchy conveyed finer boundaries and more regions (scale I: 16 regions; scale II: 32 regions; scale III: 50 regions; scale IV: 54 regions). The diameter of all regions constituting scale IV substantially exceeded the FWHM of the smoothing kernel width (minimum: 10.2 mm, maximum: 24.4 mm, median: 17.2 mm). While exceedance of the FWHM is a heuristic criterion, this suggests that the extent of smoothing was sufficiently constrained to enable delineation of the smallest region.

For scale I, the null hypothesis was tested by performing inference on maxima in the dorsal and ventral diversity curves. In particular, inference was performed to determine whether each maximum in the gradient magnitude was significantly larger than expected due to chance, geometry and smoothing (see "Boundary delineation"). For scale II and beyond, as described above, the KS statistic was used to test the null hypothesis due to challenges in mapping diversity curves for the relatively small regions constituting these scales. Delineation of small regions was also more susceptible to the effects of noise, especially signal contamination from surrounding white matter and cerebrospinal fluid. To reduce these effects, for scale II and beyond, the adjacency matrix (W) used to compute Laplacian eigenmaps was thresholded using the disparity filter[83]. The significance level for the disparity filter was set to the minimum value required to yield a graph that was not fragmented.

Watershed algorithm. The watershed transform algorithm[84] was used to cluster voxels and thereby identify spatially contiguous regions that resulted from delineation of a boundary, as applied elsewhere to parcellate the cortex[8]. The watershed transform was applied to the 3D gradient magnitude images. To initiate the watershed, seed voxels were first chosen such that exactly one seed voxel resided on each side of a boundary. Seed voxels were approximately positioned to reside within a regional minimum of the gradient magnitude image. A regional minimum that was distant from the boundary was preferred, although the parcellation was robust against the precise seed voxel location. Clusters were then iteratively grown for each seed region until locations were reached where voxels could no longer be unequivocally assigned to one of the clusters. Clusters were therefore analogous to water catchment basins (that is, minima in the gradient magnitude images), while boundaries represented watershed ridge lines (that is, maxima in these images). Before applying the watershed transform, the gradient magnitude images were (1) normalized to ensure that all values resided on the unit interval and (2) morphologically reconstructed to ensure that regional minima were only present at the location of the chosen seed voxels. The latter step aimed to enhance the contrast between voxels within the expected cluster and boundary voxels. The output of the watershed transform was a cluster of voxels for each seed voxel. Voxels that resided at boundary locations where not necessarily assigned to any cluster, yielding empty space between parcellated subregions. To fill the empty space, for each unallocated voxel, the shortest path was computed to each seed voxel. Shortest paths were computed with respect to the sparse adjacency matrix (W) following a strength-to-distance remapping of the edge weights. Unallocated voxels were then assigned to subregions with the minimum shortest path distance.

Parcellation validation and replication. *Homogeneity estimation.* A valid functional parcellation should comprise regions that are internally homogenous with respect to the functional connectivity profile of each constituent voxel. The regional homogeneity of scale I–IV parcellations was compared with the extent of homogeneity achieved with alternative subcortex parcellations as well as the homogeneity that could be expected due to chance.

The fMRI data for a region was represented as a matrix of dimension $N \times T$, where N and T denote the number of voxels constituting the region and the number of time frames, respectively. Regional homogeneity was defined as the variance explained by the first principal component of this matrix, which was estimated using PCA. Homogeneity therefore quantified the extent of synchrony in fMRI activity across voxels constituting a region. While homogeneity can be computed with respect to functional connectivity[8], we performed estimation directly on the fMRI time series acquired in the second session (REST2). This ensured that validation was performed using independent data and an independent quantity. Homogeneity was separately computed for each region and each individual. For each individual, homogeneity was then averaged across all regions to yield an overall estimate of homogeneity for the parcellation as a whole, referred to as parcellation homogeneity (Supplementary Fig. 11a).

Parcellation homogeneity increased from scale I to scale IV (Fig. 5a, red dotted line). This was because small regions were inherently more homogeneous than large regions[8]. To determine whether the observed homogeneity values were larger than expected due to chance, the subcortex was randomly parcellated such that the total number of regions and the distribution of region size were matched between the random and observed parcellations (Supplementary Fig. 11b). Parcellation homogeneity was then estimated for 100 such random parcellations to yield a null distribution for homogeneity under the null hypothesis of randomly delineated regions. A P value for this null hypothesis was estimated as the proportion of random parcellations with homogeneity exceeding or equaling the observed homogeneity. Parcellations were deemed homogeneous if the null hypothesis was rejected ($P < 0.05$). Random parcellations were generated by (1) randomly placing seed voxels throughout the subcortex mask in such a way that the Euclidean distance between all pairs of seed voxels was approximately maximized and (2) assigning all non-seed voxels to their closest seed voxel in terms of the Euclidean distance. This assignment process defined a set of randomly delineated regions. The number of seed voxels was set to the number of desired regions. Parcellations were discarded if they contained any regions that were not spatially contiguous or not matched with respect to the distribution of region sizes. Discarded parcellations were replaced until 100 valid random parcellations were generated.

Parcellation homogeneity was also computed for existing subcortex parcellation atlases (Supplementary Table 3). The scale (I–IV) that comprised the same or fewer parcels than the existing parcellation was chosen for comparison. Choosing a scale with fewer parcels provided an advantage to the existing parcellations. We specifically evaluated the homogeneity of two existing parcellation atlases for the entire subcortex[6,33] as well as parcellation atlases for the hippocampus[22,35,85], the thalamus[37] and the striatum[36]. Cerebellar and brainstem regions were omitted from the parcellation delineated by Ji et al.[6] and Fan et al.[33] to ensure homogeneity estimates were specific to the subcortex and therefore comparable to the other parcellations. To facilitate a fair comparison across parcellations containing a disparate number of regions, observed parcellation homogeneity values were normalized with respect to the mean homogeneity across the set of random parcellations. Ratios exceeding one indicated that parcellations were more homogeneous than expected due to chance.

7T replication. High-resolution fMRI acquired at ultrahigh magnetic field strength (7T) can enhance functional contrast-to-noise ratios, improve spatial localization and alleviate partial volume effects[24]. We sought to test the extent to which parcellations delineated using 3T-rfMRI could be replicated using 7T-rfMRI. To this end, the subcortex was hierarchically parcellated by applying the same procedures described above to the 7T-rfMRI data. This involved mapping Laplacian eigenmaps and gradient magnitudes, null hypothesis testing for each maximum comprising the gradient magnitude images and application of the watershed transform to delineate subregions. This yielded a 7T parcellation atlas of the subcortex that also comprised four scales (scale I: 16 regions; scale II: 34 regions; scale III: 54 regions; scale IV: 62 regions).

To enable comparisons between 3T and 7T parcellations, the 3T parcellations were up-sampled to the same spatial resolution as 7T (1.6-mm isotropic), and the spatial similarity between each scale of 3T and 7T parcellations was estimated using normalized mutual information (NMI)[86]. NMI is an information

theoretic measure that quantifies the extent to which knowledge of a parcellation reduces the uncertainty in the location of regions in a second parcellation. NMI ranges between zero and one, with higher NMI values indicating greater spatial correspondence between parcellations.

Personalized parcellation. Following Glasser and colleagues[2], a machine-learning approach was used to personalize the group-consensus parcellations and to distinguish individual differences in functional organization, thereby yielding a unique parcellation for each individual. This enabled quantification of the extent of variation among individuals for each region. The subcortex mask remained fixed for all individuals; thus, only variation in functional boundaries within the subcortex was permitted, without variation in the overall subcortical volume. The periphery of each region comprising the scale IV parcellation was first dilated by approximately 4–6 mm to define a zone of uncertainty around each region. The uncertainty zone enabled flexibility in the location of boundaries for individuals who exceeded the group-consensus boundary. The extent of dilation was such that each region in the group-consensus atlas comprised approximately the same number of voxels as its uncertainty zone, meaning that a voxel was equally likely to be randomly classified as belonging to the region compared to its uncertainty zone. Enlarging the uncertainty zone would have permitted individual boundary locations to deviate to a greater extent from the group-consensus boundaries, although major deviations were not expected in young and healthy adults, which is consistent with previous work on the cortex[2]. A binary support vector machine (SVM) classifier was trained to classify whether a voxel resided in the region or its uncertainty zone. We use M to denote the combined number of voxels constituting the region and its uncertainty zone.

Combining REST1 and REST2 would potentially yield more accurate estimates of functional connectivity compared with any of the individual sessions. However, given that parcellation was performed using data from the first 3T-rfMRI session (REST1), to assess inter-session reproducibility and to minimize circularity, the SVM was trained and evaluated using data from the second session (REST2). One hundred individuals were randomly selected to train the SVM, and the trained SVM was then applied to delineate a personalized parcellation for the remaining 921 individuals. It is important to remark that each individual contributed M training samples because classification was undertaken on a per-voxel basis. The SVM feature space was defined over the symmetric $N \times N$ similarity matrix, whereby each cell in this matrix stored the similarity in functional connectivity fingerprints between a pair of subcortex voxels (see "Functional connectivity and eigenmaps"). The feature vector for a particular voxel was selected as the relevant row in this matrix. The feature space was separately mapped for each individual and required computation of individual similarity matrices. A radial basis function kernel was used, and the feature space was standardized.

For each of the 921 individuals in the test set, the SVM was used to predict the likelihood (posterior probability) of each voxel belonging to the region under consideration, as opposed to belonging to the uncertainty zone of the region. See Supplementary Fig. 17 for a schematic of this binary classification problem. The threshold for the posterior probability of a voxel belonging to the region was varied between 0 and 100%, whereby a threshold of 0% led to all voxels, including those in the uncertainty zone, being classified as belonging to the region. Increasing the threshold eventually resulted in loss of spatial contiguity. The extent to which the threshold could be increased while preserving these properties provided insight into parcellation robustness. This process was repeated separately for each region comprising the scale IV parcellation, requiring an SVM to be independently trained for each region. It was possible for voxels to be included in the uncertainty zones of multiple regions. A voxel could therefore be assigned a probability of belonging to each of many regions. To delineate a personalized parcellation for each individual, a winner-takes-all rule was used to break ties, such that voxels were allocated to regions with the highest probability. Crucially, if the highest probability region did not exceed the posterior probability threshold, the voxel remained unassigned. This introduced the possibility of certain regions in the group-consensus parcellation being absent from the personalized parcellation of an individual. The regional detection rate was the proportion of individuals for which a given region was allocated at least one voxel. The detection rate of each region was evaluated as a function of the posterior probability threshold. The group-consensus parcellation was assumed to provide a valid representation of individuals for regions with high detection rates.

The spatial correspondence between the group-consensus and personalized parcellations was also evaluated with the Dice coefficient. In particular, the Dice coefficient was separately computed for each individual and each region as a function of the posterior probability threshold. If X and Y denote the indices of the voxels constituting the personalized region of an individual and the group-consensus region, respectively, the Dice coefficient was computed as $2|X \cap Y|/(|X| + |Y|)$. An average Dice coefficient was then computed for each individual by averaging the regional Dice coefficients.

Task-evoked connectional gradients. We aimed to investigate whether the connectional topography of the subcortex would exhibit significant variation between task conditions and rest[27,87–89]. The tasks considered were emotion processing, gambling, language processing, motor execution, relational matching, social inference and working memory. Details pertaining to each task condition are described elsewhere[54].

Following the same procedure described above (see "Functional connectivity and eigenmaps"), group-consensus eigenmaps and gradient magnitude images were computed for each of the seven tasks. For each task, the gradient magnitude image for gradient I was projected onto the dorsal and ventral streamlines. This yielded diversity curves for each task (Supplementary Fig. 19a). Tractography of the task-derived gradients was not performed, and streamlines previously determined at rest (see "Diversity curves") were used as references. To facilitate comparisons of diversity curves between different tasks, only the 725 individuals (360 males) that completed all nine conditions (two rest and seven tasks) were included. The mean diversity curve across the nine conditions was first computed and regressed from the diversity curves for each of the nine conditions. This eliminated the common variance across the conditions. The Pearson correlation coefficient was computed between the resulting residuals to generate a correlation matrix of dimension 9×9, whereby each element corresponded to a pair of task conditions. Permutation testing was used to determine a P value for each correlation coefficient. Specifically, for each pair of comparisons, the points on one of the diversity curves were randomly permutated and the correlation was recomputed on the permuted diversity curves. This was repeated for 10,000 permutations, yielding a null distribution for the correlation coefficient. The P value was given by the proportion of permutations with the correlation coefficient that exceeded or equalled the correlation coefficient in the unpermuted data. Bonferroni correction was controlled at $0.05/36 = 0.0014$ across the set of all unique pairs of conditions. The pairs of conditions that survived correction are shown in Supplementary Fig. 19b and visualized as a graph to show similarities in task conditions (Supplementary Fig. 19c). Furthermore, a matrix of dimension $H \times 9$ was then constructed, where H denotes the number of points along the diversity curve and each column corresponds to one of the nine conditions. Each cell in the matrix stored the residual value of each diversity curve point. Rows and columns were reordered according to average-linkage clustering based on the Pearson correlation coefficient (Supplementary Fig. 19d).

Given that substantially fewer fMRI volumes were acquired during task compared to rest (see "Data and pre-processing"), supplementary analyses were undertaken to test whether task-related changes in connectional gradients could be attributed to the acquisition of fewer volumes. Specifically, the group-consensus eigenmaps and the gradient magnitude images were recomputed using only the first quarter of time frames ($T = 278$) from each run of the rfMRI data (278×2 in total), thereby matching the average number of time frames across the seven tasks. We found that the gradient magnitude images computed using the shortened and full-length rfMRI acquisitions were highly correlated (Spearman correlation: $r > 0.98$ for both REST1 and REST2). This suggests that differences in the number of volumes between the tfMRI and rfMRI acquisitions do not explain the task-evoked reconfiguration of connectional gradients and diversity curves.

Relating subcortical functional network to behaviors. Individual variation in functional connectivity associates with human behaviors that relate to cognition, emotion and psychopathology in both health and mental illness[19,90–92]. While cortico–cortical connections have been well studied and predict individual variation in human behavior, less is known about the predictive utility of functional connectivity within the subcortex. Our multiscale subcortex parcellation facilitated investigations of associations between intrasubcortical functional connectivity and behavioral phenotypes for each atlas scale.

Functional connectivity and modular structure. Subcortical functional connectivity matrices were mapped for each individual and each scale of the parcellation. To this end, representative fMRI time series were determined for each region by averaging across all relevant voxels. Pearson correlation coefficients in regional time series were computed for all pairs of regions and represented in the form of a symmetric functional connectivity matrix. The connectivity matrices were then r-to-z transformed (Fisher transformation). The rows and columns of the connectivity matrices were reordered to accentuate modular structure, whereby the modular structure was determined using the Louvain community detection algorithm[93]. Separate connectivity matrices were mapped for the first (REST1) and second (REST2) 3T-rfMRI sessions to provide a discovery and within-sample replication set.

Behavioral phenotypes. Behavioral measurements and procedures for the HCP are described in detail elsewhere[54]. Items that tapped alertness, cognition, emotion, sensory–motor function, personality, psychiatric symptoms, substance use and life function were selected for further analyses, subject to the following conditions:

1. Only total or subtotal scores were selected for psychological assessments.
2. Only raw scores were selected for items with both raw and age-adjusted scores.
3. Items based on strictly binary responses were excluded.
4. Items with missing responses in more than 10% of individuals were excluded.

Items relating to task performance in emotional processing, language processing, social recognition, motor execution, working memory, relational

matching and reward/decision-making processing were also included. This resulted in a total of 109 behavioral measures for each individual (Supplementary Table 4). Participants with missing responses for any of these 109 items were then excluded, resulting in a final sample comprising 958 individuals (453 males, mean age of 28.7 ± 3.7 years). Two 3T-rfMRI sessions were acquired in all these individuals.

Independent component analysis (ICA) was used to reduce the set of behavioral measures to a set of latent dimensions with the aim of performing inference on the dimensions rather than the full set of behaviors. Before performing ICA, behavioral items with continuously distributed scores (87 out of 109) were normalized with a rank-based inverse Gaussian transformation[94]. Age and sex were then regressed from all items and ICA was performed on the residuals that resulted from this regression. Other confounding factors such as blood pressure, hemoglobin, height and weight were not included due to missing responses for several individuals. ICA was performed using the Icasso package[95] for Matlab, which provides an implementation of the FastICA algorithm[96]. Under ICA, the matrix of residuals (behaviors × individuals), denoted with Y, was factorized into the product $Y = A^{-1}S$, where A^{-1} denotes the mixing matrix (behaviors × components) and S denotes the estimated independent sources (components × individuals). The inverse of A^{-1} is denoted A (components × behaviors) and referred to as the de-mixing matrix.

To estimate the de-mixing matrix, FastICA was performed using default parameter settings, including a cubic nonlinearity and deflationary orthogonalization. Bootstrapping was used to sample individuals with replacement, and ICA was performed independently on each bootstrapped sample using randomly chosen initial conditions. A total of 500 bootstrapped samples were generated. Independent components that represented a consensus across the 500 samples were then derived using a clustering procedure. In particular, agglomerative clustering with average linkage was used to partition the de-mixing matrix into putative clusters. The number of dimensions in the feature space was the number of behaviors, and each cluster comprised groups of components (that is, rows from A). Clearly separated clusters indicated that independent components were consistently and reliably estimated despite randomization of initial conditions and bootstrapping. A cluster quality index was used to quantify cluster separation and provided a metric to select the number of independent components (see below). Cluster separation was quantified by the difference in average intra-cluster similarity and average inter-similarity, denoted by Iq. A higher Iq suggests compact clusters, whereas a lower Iq suggests that clusters are not well separated. These operations were performed internally in the Icasso package, and further details are available in the documentation of the package.

In addition to these internal procedures, we performed a sampling and matching process to ensure stability and reliability of the estimated independent components. Specifically, the ICA as described above was repeated for ten trials, yielding an ensemble of de-mixing matrices, $A_1, A_2, \ldots, A_{10}$ and corresponding independent components, $S_1, S_2, \ldots, S_{10}$. Variability in A_u and S_u across trials was due to randomization of initial conditions and random sampling inherent to bootstrapping. Each row of A corresponded to an independent component, and each column quantified the weighting of a behavioral item in the given component. We sought to evaluate the similarity of the resulting estimates across trials to investigate whether the Icasso procedure yielded reproducible independent components across the ten trials. For each pair of trials u and v, a correlation matrix of dimension $I \times I$ was computed, where I denotes the number of independent components. Element (i,j) of this correlation matrix stored the Pearson correlation coefficient between row i of A_u and row j of A_v. We denote the correlation matrix between trial u and v with C_{uv}, $u,v = 1,2\ldots,10$. Given that independent components are resolved up to an arbitrary sign flip, the absolute value of all correlation coefficients was considered. The nth component in trial u did not necessarily correspond to the nth component in trial v; thus, a one-to-one matching of components was required for each pair of trials. The Hungarian algorithm[97] was applied to C_{uv} to determine a one-to-one matching that maximized the correlation coefficients between the matched pairs of components. If ICA estimates were reproducible between trials u and v, each column of C_{uv} should comprise a correlation coefficient that is substantially larger than all others in the column, indicating an obvious matching. The trial that was best matched to all other trials was chosen as the reference trial, whereby matching quality was quantified as the average correlation coefficient computed over all matched pairs of components in C_{uv}.

Once the reference trial was determined, the components in each of the other trials were reordered to match the reference, as dictated by the optimal one-to-one matching computed using the Hungarian algorithm. The polarity of the component was potentially flipped to ensure consistency with the reference trial. The reordered de-mixing matrices $A_1, A_2, \ldots A_{10}$ were then averaged across trials to yield a single de-mixing matrix that represented a consensus across the ten trials. The independent components $S_1, S_2, \ldots, S_{10}$ were reordered and then averaged in the same way.

Due to the effects of bootstrapping and randomization, the de-mixing matrix was not necessarily perfectly orthogonal, and some components were minimally correlated. The Pearson correlation coefficient was computed between all pairs of components in the de-mixing matrix. The reciprocal of the largest correlation coefficient was used to quantify the extent of interdependence between components such that higher values of the reciprocal were desired. This was performed for each of the reordered de-mixing matrices $A_1, A_2, \ldots, A_{10}$, and the average across trials resulted in a consensus index of the independence.

Three criteria were therefore used to evaluate ICA performance and to assist in determining the best number of components to estimate:

1. Cluster separation: the extent of separation between clusters among independent components found across bootstrapped samples. This was estimated internally as part of the Icasso package and provides an estimate of reliability and stability of independent components.
2. Trial matching quality: the quality of the one-to-one matching of independent components across ICA trials. This quantified the reproducibility of independent components and was measured as the average distance between matching independent components of each trial and the reference trial.
3. Independence index: the independence of components in the de-mixing matrix. This was quantified as the reciprocal of the largest correlation coefficient between all pairs of components.

The ICA procedure described above, including bootstrap resampling and trial matching, was separately repeated for candidate models ranging from 3 to 30 independent components. The optimal model (that is, the number of components) was informed by plotting the above measures of cluster separation, trial matching quality and independence index as a function of the number of components and identifying abrupt changes based on the elbow criterion (Supplementary Fig. 24). While both cluster separation (Supplementary Fig. 24a) and trial matching quality (Supplementary Fig. 24b) showed abrupt reductions when moving from six to seven components, which suggests that six components was optimal, the independence index (Supplementary Fig. 24c) suggested that five components was most suitable. We could therefore narrow down model selection from 28 candidates to only 2 candidate models: 5 or 6 independent components. To break the tie between these two models, the independent components constituting both models were visualized as word clouds, with the font size of each behavior scaled according to its de-mixing weight. Although, the five-component and six-component models were highly comparable, we selected the five-component model because it was the simpler model and enabled clear labeling of each independent component. The five independent components characterized cognitive performance, illicit substance use, tobacco use, personality and emotional traits, as well as mental health. Hence, the 109 behavioral items initially selected were decomposed into five independent components, which we refer to as behavioral dimensions (Fig. 8a,b and Supplementary Fig. 21).

NBS analysis. The NBS[28] was used to test whether functional connectivity mapped within each parcellation scale (scales I–IV) significantly associated with any of the five behavioral dimensions. The NBS was applied independently to each dimension and parcellation scale. For each pair of regions, a general linear model was formulated in which functional connectivity was the dependent variable and the independent variables included an intercept term, the behavioral dimension (that is, the relevant column from the estimated independent sources) and confounds including head motion parameters summarized by the framewise displacement[59] and total subcortical volume estimated by FreeSurfer (https://surfer.nmr.mgh.harvard.edu/), as provided by the HCP. Given that age and sex were initially regressed from all behavioral scores before performing ICA, these potential confounds were not included in the statistical model. Each connection was endowed a t-statistic and a corresponding uncorrected P value for the null hypothesis that the association between functional connectivity and the behavioral dimension was not significantly different from zero. The premise of the NBS is to reject the null hypothesis for connected graph components[28] rather than for individual connections, thereby relinquishing some precision in effect localization for a gain in statistical power. To this end, connected components were identified among the set of connections with a t-statistic exceeding 4. This t-statistic threshold was chosen to ensure that the smallest effect size of interest that could be detected was approximately a Cohen's d of 0.1. The number of suprathreshold connections constituting each component defined the size of the component. Permutation testing was used to generate an empirical null distribution for component size. For each of the 10,000 permutations, an established method[98] was used to refit the general linear model to data in which the correspondence between individuals and behaviors was permuted. Permutation sequences were obtained using multilevel block permutation[51,52] to ensure that the data were shuffled in a way that respected known familial relationships among individuals. In particular, each family could only be shuffled as a whole, and family members could only be permuted within their family. Dizygotic twins and non-twin siblings were assumed to share the same kinship and they could therefore be permuted. The size of the largest connected component was stored for each permutation to ensure control of the family-wise error rate across all components identified. The family-wise-error-corrected P value for an observed component was given by the proportion of permutations for which the largest component identified was larger than or equal in size to the observed component. The NBS was repeated independently for each of the five behavioral dimensions and for the two alternative hypotheses of positive and negative correlation. This resulted in a total

of ten multiple comparisons. The FDR was controlled at 5% across these multiple comparisons.

The NBS was performed on functional connectivity mapped using 3T-rfMRI data from the second acquisition (REST2) given that the new atlas was delineated using the first acquisition (REST1). The reproducibility of any significant associations identified with the NBS in REST2 were evaluated using functional connectivity mapped using 3T-rfMRI data from REST1. To enable comparisons, functional connectivity mapped with an alternative subcortex atlas (Subcortex-II)[33] was also tested for associations with behavioral dimensions following the same procedure described above.

Statistical analysis. Null hypothesis tests were used to perform statistical analyses. Statistical significance was determined using permutation tests and resampling. Specifically, for atlas delineation, the null hypothesis was that peaks in the gradient magnitude images were consistent with the geometry of the subcortex and/or fMRI confounds. The alternative hypothesis was that the peaks were larger than attributable to subcortical geometry and confounds. Each rejection of this null hypothesis provided evidence of a discrete boundary in subcortical topography. A dedicated geometry-preserving null model was developed to perform inference on the local maxima of the gradient magnitude images (see "Null model"). This facilitated the generation of an empirical null distribution for peak magnitudes by repeatedly drawing random samples from the null model. The *P* value for each peak was estimated as the proportion of samples constituting the null distribution with peaks that equalled or exceeded the magnitude of the actual peak (see "Boundary delineation"). The FDR was controlled at 5% across the set of all peaks identified at each scale using the Benjamini–Hochberg procedure. For the behavioral analyses, the null hypothesis was that the five behavioral dimensions did not explain significant individual variation in subcortical functional connectivity. This null hypothesis was tested using linear regression analysis and the NBS[28]. All null hypothesis tests were nonparametric and did not mandate strong assumptions about data normality.

Further methodological considerations. *Connectivity gradients and subcortical geometry.* A clear peak in the diversity curve separates the hippocampus and the thalamus (Fig. 2d). However, the transition between the hippocampus and the thalamus is narrow and represents a local constriction in the geometry of the subcortex. Because our resampling preserves the geometry of the empirical data, a relatively abrupt transition is also present between the thalamus and the hippocampus in the null data. In other words, the functional separation between the hippocampus and the thalamus can be largely explained by the intrinsic geometry of the subcortex. This does not imply that the hippocampus and the thalamus are not functionally separated, but simply suggests that the functional separation is consistent with the underlying geometry of the two structures and does not require an additional disruption in the functional connectivity gradients. Conversely, there are occasional peaks in the null diversity curves that are not present in the real data (for example, the broad peak within the caudate shown in Fig. 2c). This presumably arises when the local functional connectivity gradient is smoother (more homogenous) in the real data across a geometric feature than predicted by the null data.

Multiscale organization and model selection. Multiscale (hierarchical) organization is not necessarily a fait accompli of the recursive application of model selection. Indeed, we explicitly tested for the possibility of an absence of multiscale structure for each region comprising scale I, but found that the null hypothesis could be rejected for all regions, which is suggestive of a multiscale organizational structure. If we were unable to reject the null hypothesis, this would have instead suggested an absence of any multiscale structure. Furthermore, the number of scales substantially differed between subcortical nuclei. For example, the thalamus, the hippocampus and the caudate each comprised four levels, which suggests a deeper and more complex multiscale structure relative to the amygdala and the GP, which only comprised two levels. This is consistent with the greater organizational complexity of the thalamic nuclei, from molecular[99] and cellular[39] to connectivity[20,37] and function[100,101]. In contrast, the GP is primarily involved in the regulation of movement control[102] and the amygdala plays an essential role in emotional processes[103].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Neuroimaging data analyses were undertaken using publicly available human neuroimaging datasets acquired and maintained by the HCP[18]. These datasets are available for download to anyone agreeing to the Open Access Data Use Terms (https://db.humanconnectome.org/). Access to family structure data and several behavioral measures requires acceptance of the HCP Restricted Data Use Terms (https://www.humanconnectome.org/study/hcp-young-adult/document/restricted-data-usage). The independent validation dataset used to assess the reproducibility of the parcellation homogeneity results is currently not publicly available. The new atlas is openly available in the form of NIFTI (Neuroimaging Informatics Technology Initiative) and CIFTI (Connectivity Informatics Technology Initiative) files. To facilitate mapping of whole-brain connectomes, we have also integrated the new atlas into several well-known cortex-only parcellation atlases and the combined cortex–subcortex atlases are made openly available. Supplementary Table 5 provides details about the atlas formats that can be downloaded from the GitHub repository (https://github.com/yetianmed/subcortex).

Code availability

The Matlab (R2018b) codes to compute Laplacian eigenmaps, gradient magnitudes, diversity curves and other computational analyses undertaken as part of this study are openly available at https://github.com/yetianmed/subcortex. The following additional software packages used for this study are freely and openly available: Diffusion Toolkit (v.0.6.4.1) and TrackVis (v.0.6.1): http://trackvis.org/; NBS (v.1.2): https://www.nitrc.org/projects/nbs/; Icasso (v.1.21): https://research.ics.aalto.fi/ica/icasso/; NeuroMArVL: https://immersive.erc.monash.edu/neuromarvl/; PALM (v.alpha116): https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM/ExchangeabilityBlocks; and fMRIPrep (v.1.5.9): https://fmriprep.org/en/stable/index.html.

References

51. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *NeuroImage* **92**, 381–397 (2014).
52. Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E. & Smith, S. M. Multi-level block permutation. *NeuroImage* **123**, 253–268 (2015).
53. Smith, S. M. et al. Resting-state fMRI in the Human Connectome Project. *NeuroImage* **80**, 144–168 (2013).
54. Barch, D. M. et al. Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169–189 (2013).
55. Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).
56. Salimi-Khorshidi, G. et al. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage* **90**, 449–468 (2014).
57. Griffanti, L. et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage* **95**, 232–247 (2014).
58. Parkes, L., Fulcher, B., Yucel, M. & Fornito, A. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *NeuroImage* **171**, 415–436 (2018).
59. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* **59**, 2142–2154 (2012).
60. Satterthwaite, T. D. et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage* **64**, 240–256 (2013).
61. Van Dijk, K. R., Sabuncu, M. R. & Buckner, R. L. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage* **59**, 431–438 (2012).
62. Puckett, A. M. et al. Using multi-echo simultaneous multi-slice (SMS) EPI to improve functional MRI of the subcortical nuclei of the basal ganglia at ultra-high field (7T). *NeuroImage* **172**, 886–895 (2018).
63. Glasser, M. F. et al. The Human Connectome Project's neuroimaging approach. *Nat. Neurosci.* **19**, 1175–1187 (2016).
64. Glasser, M. F. et al. Using temporal ICA to selectively remove global noise while preserving global signal in functional MRI data. *NeuroImage* **181**, 692–717 (2018).
65. Cole, M. W. et al. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* **16**, 1348–1355 (2013).
66. Esteban, O. et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
67. Percheron, G., Yelnik, J. & Francois, C. A Golgi analysis of the primate globus pallidus. III. Spatial organization of the striato–pallidal complex. *J. Comp. Neurol.* **227**, 214–227 (1984).
68. Yushkevich, P. A. et al. Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol. *NeuroImage* **111**, 526–541 (2015).
69. Guadalupe, T. et al. Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex. *Brain Imaging Behav.* **11**, 1497–1514 (2017).
70. de Hollander, G., Keuken, M. C., van der Zwaag, W., Forstmann, B. U. & Trampel, R. Comparing functional MRI protocols for small, iron-rich basal ganglia nuclei such as the subthalamic nucleus at 7 T and 3 T. *Hum. Brain Mapp.* **38**, 3226–3248 (2017).
71. Belkin, M. & Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**, 1373–1396 (2003).

72. Fornito, A., Zalesky, A. & Breakspear, M. Graph analysis of the human connectome: promise, progress, and pitfalls. *NeuroImage* **80**, 426–444 (2013).
73. Cerliani, L. et al. Probabilistic tractography recovers a rostrocaudal trajectory of connectivity variability in the human insular cortex. *Hum. Brain Mapp.* **33**, 2005–2034 (2012).
74. Jeurissen, B., Descoteaux, M., Mori, S. & Leemans, A. Diffusion MRI fiber tractography of the brain. *NMR Biomed.* **32**, e3785 (2019).
75. Alexander, D. C. in *Visualization and Processing of Tensor Fields* (eds Weickert J. & Hagen, H.) Ch. 5 (Springer, 2006).
76. Friman, O., Farneback, G. & Westin, C. F. A Bayesian approach for stochastic white matter tractography. *IEEE Trans. Med. Imaging* **25**, 965–978 (2006).
77. Peled, S., Friman, O., Jolesz, F. & Westin, C. F. Geometrically constrained two-tensor model for crossing tracts in DWI. *Magn. Reson. Imaging* **24**, 1263–1270 (2006).
78. Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F. & Woolrich, M. W. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *NeuroImage* **34**, 144–155 (2007).
79. Wang, R., Benner, T., Sorensen, A. & Wedeen, V. Diffusion Toolkit: a software package for diffusion imaging data processing and tractography. *Proc. Int. Soc. Magn. Reson. Med.* **15**, 3720 (2007).
80. O'Donnell, L. J. & Westin, C. F. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. Med. Imaging* **26**, 1562–1575 (2007).
81. Müller, M. in *Information Retrieval for Music and Motion* 69–84 (Springer, 2007).
82. Benjamini, Y. Discovering the false discovery rate. *J. R. Stat. Soc. Ser. B* **72**, 405–416 (2010).
83. Serrano, M. Á., Boguñá, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl Acad. Sci. USA* **106**, 6483–6488 (2009).
84. Meyer, F. Topographic distance and watershed lines. *Signal Process.* **38**, 113–125 (1994).
85. Eickhoff, S. B. et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* **25**, 1325–1335 (2005).
86. Lancichinetti, A., Fortunato, S. & Kertész, J. Detecting the overlapping and hierarchical community structure in complex networks. *N. J. Phys.* **11**, 033015 (2009).
87. Greene, A. S., Gao, S., Scheinost, D. & Constable, R. T. Task-induced brain state manipulation improves prediction of individual traits. *Nat. Commun.* **9**, 2807 (2018).
88. Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S. & Petersen, S. E. Intrinsic and task-evoked network architectures of the human brain. *Neuron* **83**, 238–251 (2014).
89. Medaglia, J. D., Lynall, M. E. & Bassett, D. S. Cognitive network neuroscience. *J. Cogn. Neurosci.* **27**, 1471–1491 (2015).
90. Smith, S. M. et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* **18**, 1565–1567 (2015).
91. Fornito, A., Zalesky, A., Pantelis, C. & Bullmore, E. T. Schizophrenia, neuroimaging and connectomics. *NeuroImage* **62**, 2296–2314 (2012).
92. Xia, C. H. et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat. Commun.* **9**, 3003 (2018).
93. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
94. Van der Waerden, B. Oder tests for two-sample problem and their power. *Proc. K. Nederlandse Akademie van Wetenschappen. Ser. A* **55**, 453–458 (1952).
95. Himberg, J., Hyvarinen, A. & Esposito, F. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage* **22**, 1214–1222 (2004).
96. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10**, 626–634 (1999).
97. Kuhn, H. W. The Hungarian method for the assignment problem. *Nav. Res. Logist. Q* **2**, 83–97 (1955).
98. Freedman, D. & Lane, D. A nonstochastic interpretation of reported significance levels. *J. Bus. Economic Stat.* **1**, 292–298 (1983).
99. Phillips, J. W. et al. A repeated molecular architecture across thalamic pathways. *Nat. Neurosci.* **22**, 1925–1935 (2019).
100. Hwang, K., Bertolero, M. A., Liu, W. B. & D'Esposito, M. The human thalamus is an integrative hub for functional brain networks. *J. Neurosci.* **37**, 5594–5607 (2017).
101. Shine, J. M. et al. The low-dimensional neural architecture of cognitive complexity is related to activity in medial thalamic nuclei. *Neuron* **104**, 849–855.e3 (2019).
102. Mink, J. W. The basal ganglia: focused selection and inhibition of competing motor programs. *Prog. Neurobiol.* **50**, 381–425 (1996).
103. Janak, P. H. & Tye, K. M. From circuits to behaviour in the amygdala. *Nature* **517**, 284–292 (2015).

Acknowledgements

We thank M. Glasser (Washington University) for provision of the Wishart filter code. We thank L. Cocchi (QIMR Berghofer Medical Research Institute) for provision of the additional validation dataset. All other data were provided by the HCP, the WU–Minn Consortium (1U54MH091657; Principal Investigators: D. Van Essen and K. Ugurbil) funded by the 16 National Institutes of Health (NIH) institutes and centers that support the NIH Blueprint for Neuroscience Research, and by the McDonnell Center for Systems Neuroscience at Washington University. Y.T. was supported by a NHMRC Project Grant awarded to A.Z. (APP1142801). M.B. and A.Z. were each supported by research fellowships from the NHMRC (APP1136649 and APP1118153, respectively).

Author contributions

Y.T. and A.Z. conceived the idea and designed the study. Y.T. compiled the data, performed the analyses and prepared the visualizations. Y.T. and A.Z. drafted the manuscript. D.S.M. and M.B. provided critical conceptual input. All authors provided critical feedback and editing of the final manuscript. A.Z. supervised the research.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-020-00711-6.

Correspondence and requests for materials should be addressed to Y.T. or A.Z.

Peer review information *Nature Neuroscience* thanks Janine Bijsterbosch, Rodrigo Braga and Evan Gordon for their contribution to the peer review of this work.

nature research

Corresponding author(s): Ye Tian, Andrew Zalesky

Last updated by author(s): Jul 28, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Analyses were undertaken on publicly available human neuroimaging datasets acquired and maintained by the Human Connectome Project (HCP). Specifically, 3T functional MRI data was acquired using multiband echo planar imaging (EPI) on a customized Siemens 3T MR scanner (Skyra system); 7T functional MRI data was acquired on a Siemens Magnetom 7T MR scanner. See Van Essen et al (2013) for further details. Van Essen et al (2013) The WU-Minn Human Connectome Project: An Overview. NeuroImage. 80:62-79 The independent validation dataset for assessing the reproducibility of the parcellation homogeneity results was acquired using a Siemens Prisma 3T MRI scanner.
Data analysis	Computer codes (MATLAB; R2018b) to compute Laplacian eigenmaps, gradient magnitudes, diversity curves and other computational analyses undertaken as part of this study are openly available in a community repository (https://github.com/yetianmed/subcortex). All additional software packages used for this study are freely and openly available: - Diffusion Toolkit (v.0.6.4.1) and TrackVis (v.0.6.1; gradientography visualization): http://trackvis.org/ - NBS (v.1.2; connectome inference): https://www.nitrc.org/projects/nbs/ - Icasso (v.1.21; mapping of behavioral dimensions): https://research.ics.aalto.fi/ica/icasso/ - NeuroMArVL (connectome visualization): https://immersive.erc.monash.edu/neuromarvl/ - PALM (v.alpha116; permutation sequences constrained by family structure): https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM/ExchangeabilityBlocks -fMRIPrep (v.1.5.9; fMRI data preprocessing for the independent validation dataset): https://fmriprep.org/en/stable/index.html

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All primary analyses were undertaken on publicly available human neuroimaging, demographic and behavioral datasets acquired and maintained by the Human Connectome Project (HCP). These datasets are available for download to anyone agreeing to the open access data use terms (https://db.humanconnectome.org/). All datasets were acquired by the Washington University and University of Minnesota Consortium of the HCP. In the current work, minimally preprocessed resting-state and task-evoked functional MRI datasets were used from the S1200 data release of the HCP. Access to family structure data for permutation testing and several behavioral measures required acceptance of the HCP Restricted Data Use Terms (https://www.humanconnectome.org/study/hcp-young-adult/document/restricted-data-usage).
The independent resting-state functional MRI dataset comprising ten healthy adults was acquired in Australia to test the reproducibility of the parcellation homogeneity results. This dataset is currently not publicly available.
The new atlas developed in this study is openly available (https://github.com/yetianmed/subcortex) in the form of NIFTI (Neuroimaging Informatics Technology Initiative) and CIFTI (Connectivity Informatics Technology Initiative) files. Further details are provided on the GitHub repository.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Human neuroimaging data acquired as part of the Human Connectome Project S1200 release (n=1113, Van Essen et al 2013) were analyzed in this study. The primary dataset comprised 1080 individuals (585 females) who had completed all two runs in the first 3T resting-sate functional MRI session (REST1). The second 3T resting-state functional MRI session (REST2, n=1021, 550 females) was used for replication. Participants ranged between 22 and 37 years of age. In addition to the primary datasets, 7 Tesla resting-state functional MRI (n=183, 111 females) and 3 Tesla task-evoked functional MRI (n=725, 365 females) images were analyzed to address the study aims. No statistical methods were used to pre-determine sample sizes. All relevant and appropriate data provisioned by the Human Connectome Project were included to maximize sample sizes. An independent resting-state functional MRI dataset comprising 10 healthy adults (mean age 26±2.1yrs, 6 males) was used to test the reproducibility of the parcellation homogeneity results. The independent validation dataset was used to provide an additional validation for the parcellation homogeneity results and was sufficient to replicate these results.
Data exclusions	Each resting-sate functional MRI session (REST1 and REST2) comprising two runs. Participants that had completed at least two runs in each dataset were included. As a consequence, a total of 33 of the original 1113 participants (2.9%) were excluded from REST1, and 92 (8.2%) individuals were excluded from REST2 due to absence of one of the two resting-state functional MRI runs in each session, respectively. For behavioral analyses, an additional 155 participants (13.9%) with missing behavioral responses for any one of the 109 behavioral items were excluded. This yielded a final sample size comprising 958 individuals (505 females) for the analyses relating subcortical functional networks to behavioral dimensions. All exclusions were performed before undertaking any analyses and inference.
Replication	Reproducibility of the current work was established by analyzing: i) two fMRI data sessions; ii) two different magnetic field strengths (3T and 7T); iii) two image modalities (task and rest); and, iv) an independent validation dataset (parcellation homogeneity). Specifically, the 1113 individuals comprising the primary dataset participated in two separate resting-state functional MRI sessions (REST1 and REST2). The relation between subcortical connectivity and behavior was identified in REST2 and subsequently reproduced in REST1. This established reproducibility in the context of MRI inter-session variability. The relation between subcortical connectivity and behavior was reproduced for each of the four atlas scales (Scale I-IV). The new atlas was delineated using 3 Tesla and 7 Tesla resting-state functional MRI. While 7 Tesla unveiled finer subdivisions of hippocampus and amygdala, compared to 3 Tesla, the atlas showed significant spatial concordance between the two field strengths (normalized mutual information > 0.8). The parcellation homogeneity was estimated using the second 3T MRI session (REST2), whereas the first session (REST1) was used for atlas delineation. Moreover, the parcellation homogeneity results were replicated in an independent validation dataset comprising ten healthy adults.
Randomization	Participants were not allocated to experimental groups. Randomization was therefore not applicable. Further details pertaining to experimental design are described by Van Essen et al (2013). For permutation testing, permutation sequences were constrained to control for family structure. For behavioral analyses, covariates that were controlled as part of the general linear model included head motion parameters and subcortical volumes. Age and sex were regressed from all behavioral items before performing ICA.
Blinding	Participants were not allocated to experimental groups during recruitment, data processsing and/or statistical analyses. Blinding was therefore not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The primary dataset comprised 1080 healthy young adults ranging between 22 and 37 years of age (585 females). The Human Connectome Project aimed to recruit a cohort that was generally representative of the population at large, thereby capturing a wide range of variability in healthy individuals with respect to behavioral, ethnic, and socioeconomic diversity (Van Essen et al 2013). See Van Essen et al (2013) for further details. Van Essen et al (2013) The WU-Minn Human Connectome Project: An Overview. NeuroImage. 80:62-79 The independent validation dataset was acquired in Australia and comprised 10 healthy adults (mean age 26±2.1yrs, 6 males).
Recruitment	The Human Connectome Project made efforts to recruit participants broadly reflecting the ethnic and racial composition of the U.S. population as represented in the 2000 decennial census (Van Essen et al 2013). Exclusion criteria included severe neurodevelopmental disorders, documented neuropsychiatric disorders or neurologic disorders. Individuals with illnesses such as diabetes or high blood pressure were also excluded during recruitment. See Van Essen et al (2013) for further details. Detailed exclusion criteria: 1. Significant history of psychiatric disorder, substance abuse, neurological, or cardiovascular disease a) Participant report of diagnosis by a treating physician; or b) Hospitalization for the condition for two days or longer; or c) Pharmacologic or behavioral treatment by a cardiologist, psychiatrist, neurologist, or endocrinologist for a period of 12 months or longer, other than treatment for childhood-only ADHD. 2. Participant report of diagnosis by a treating physician; or 3. Hospitalization for the condition for two days or longer; or 4. Participant report of diagnosis by a treating physician; or 5. Hospitalization for the condition for two days or longer; or 6. Pharmacologic or behavioral treatment by a cardiologist, psychiatrist, neurologist, or endocrinologist for a period of 12 months or longer, other than treatment for childhood-only ADHD. 7. Two or more seizures after age 5 or a diagnosis of epilepsy 8. Any genetic disorder, such as cystic fibrosis or sickle cell disease 9. Multiple sclerosis, cerebral palsy, brain tumor or stroke 10. Any of the following head injuries a) Loss of consciousness for > 30 minutes; or b) Amnesia for > 24 hours; or c) Change in mental status for > 24 hours; or d) CT findings consistent with traumatic brain injury; or e) Three or more concussive (mild) incidences of head injury 11. Premature birth (for twins, before 34 weeks; for non-twin siblings, before 37 weeks. If weeks unknown, less than 5 lbs. at birth for non-twins) 12. Currently on chemotherapy or immunomodulatory agents, or history of radiation or chemotherapy that could affect the brain. 13. Thyroid hormone treatment in the past month 14. Treatment for diabetes in the past month (other than gestational or diet-controlled diabetes) 15. Use of daily prescription medications for migraines in the past month 16. A score of 25 or below on the Folstein Mini Mental State Exam on visit Day 1 17. Moderate or severe claustrophobia 18. Pregnancy 19. Unsafe metal in the body The independent validation dataset was recruited from the general community. Individuals aged between 18 and 65 years were eligible. Exclusion criteria included neurodevelopmental disorders, neurologic disorders and/or neuropsychiatric disorders.
Ethics oversight	The HCP datasets were acquired by the WU-Minn HCP consortium with local human research ethics approval and shared with

Ethics oversight	the authors in accordance with the WU-Minn HCP Consortium Open Access and Restrict Data Use Terms. Our use of HCP datasets and our recruitment of the independent dataset complied with the Australian NHMRC national framework for the conduct of human research. Informed consent was obtained from all the participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Resting-state and task block/event design function MRI. Resting state: 3 and 7 Tesla. Task: 3 Tesla.
Design specifications	The resting-state functional MRI design is described in Glasser et al (2013), while the task-evoked functional MRI design is described in Barch et al (2014). Seven distinct tasks were chosen to tap a broad range of cognitive and affective processes and activated a wide range of neural systems. The tasks considered were emotion processing, gambling, language processing, motor execution, relational matching, social inference and working memory. The number of blocks/events varied according to the specific task. Glasser et al (2013) The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage. 80:105-24 Barch et al (2014) Function in the human connectome: task-fMRI and individual differences in behavior. NeuroImage. 80:169-89
Behavioral performance measures	An extensive set of task performance measures were acquired by the Human Connectome Project. Behavioral items relating to task performance in emotional processing, language processing, social recognition, motor execution, working memory, relational matching and reward/decision making processing were included in the mapping of behavioral dimensions. Further details pertaining to task-evoked functional MRI are provided in Barch et al (2014).

Acquisition

Imaging type(s)	Functional magnetic resonance imaging
Field strength	3 Tesla and 7 Tesla
Sequence & imaging parameters	HCP datasets: 3T functional MRI sequence: Gradient-echo EPI; Imaging parameters: TR=720ms, TE=33.1ms, flip angle=52 degrees, FOV=208x180mm, Matrix=104x90, Slice thickness=2.0mm, 72 slices and 2.0 mm isotropic voxels, alternated LR and RL phase encoding. 7T functional MRI sequence: Gradient-echo EPI; Imaging parameters: TR=1000ms, TE=22.2ms, flip angle=45 degrees, FOV=208x208mm, Matrix=130x130, Slice thickness=1.6mm, 85 slices and 1.6 mm isotropic voxels, alternated AP and PA phase encoding. Independent validation dataset: 3T functional MRI sequence: Gradient-echo EPI; Imaging parameters: TR=810ms, TE=30ms. flip angel=53 degrees, FOV=240x256mm, Matrix=106x106, Slice thickness=2.0mm, 72 slices and 2.0 mm isotropic voxels, PA phase encoding.
Area of acquisition	Whole brain scan

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	HCP datasets: Minimally preprocessed resting-state and task fMRI data were sourced from the Human Connectome Project. Details of the minimal preprocessing pipeline can be found elsewhere (Glasser et al., 2013). The minimally preprocessed data were subjected to two additional operations. First, spatial smoothing was performed with a Gaussian smoothing kernel of 6mm FWHM (full width at half maximum) for 3T and 4mm FWHM for 7T images. Second, to further improve subcortical SNR and suppress the effects of unstructured and autocorrelated noise that share similar temporal and spatial properties, the fMRI time series were reconstructed from a principal component analysis (PCA), where the components were filtered according to an empirically-fitted Wishart distribution to dampen the effects of noise (Glasser et al., 2016). Custom code to perform these two steps are available in a community repository (https://github.com/yetianmed/subcortex). Independent validation dataset: Resting-state fMRI data were preprocessed using fMRIPrep-1.5.9 (Estenban et al., 2019). Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, Gorgolewski KJ. 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature methods. 16:111-116.
Normalization	The functional images were spatially aligned to the MNI (Montreal Neurological Institute) standard space using the FNIRT nonlinear registration algorithm as part of the minimal preprocessing pipeline (Glasser et al., 2013).

Normalization template: MNI ICBM 152 nonlinear 6th generation

Noise and artifact removal:
HCP datasets:
The correction of head motion and spatially structured physiological noise were performed with ICA+FIX (Salimi-Khorshidi et al.,2014; Griffanti et al., 2014) as part of the minimal preprocessing pipeline (Glasser et al., 2013). A summary measure of head motion was included as a covariate when testing for associations between behavioral associations and connectivity.

Independent validation dataset:
Confounds including the 24 head motion parameters (six basic motion parameters + six temporal derivatives + 12 quadratic terms and their six temporal derivatives), mean white matter and CSF signals were computed and regressed from the preprocessed fMRI data for each individual. The residuals of this regression were then subjected to spatial smoothing with a Gaussian smoothing kernel of 6mm FWHM and Wishart filtering.

Volume censoring: Volume censoring was not performed on this data.

Statistical modeling & inference

Model type and settings: Univariate (linear regression), multivariate (principle component analysis and independent component analysis) and predictive (support vector machine learning)

Effect(s) tested:
For atlas delineation, the null hypothesis was that peaks in the gradient magnitude images could be explained by geometry and/or other confounds. A dedicated geometry-preserving null model was developed to perform inference on the local maxima of the gradient magnitude images.
For the behavioral analyses, the null hypothesis was that the five behavioral dimensions did not explain significant individual variation in subcortical functional connectivity. This null hypothesis was tested using linear regression analysis and the network-based statistic.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both

Anatomical location(s): The probabilistic Harvard-Oxford Subcortical Structural Atlas (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases) was used to delineate a binary mask for each of the following subcortical regions: thalamus, caudate, putamen, nucleus accumbens, globus pallidus, hippocampus and amygdala.

Statistic type for inference (See Eklund et al. 2016): Non-parametric permutation testing, network-based statistic (Zalesky et al., 2010)

Correction: Non-parametric permutation testing, FDR correction and Bonferroni correction.

Models & analysis

n/a	Involved in the study
☐	☒ Functional and/or effective connectivity
☐	☒ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Functional and/or effective connectivity: Pearson correlation and Fisher transformation.

Graph analysis: The Louvain community detection algorithm (Blondel et al., 2008) was used to determine the modular structure in the subcortical functional connectivity matrix.

Multivariate modeling and predictive analysis:
In the analysis of mapping functional connectivity and eigenmaps, whole-brain functional connectivity was mapped for each subcortex voxel. Spatial gradients in the resulting maps were then computed to yield a continuous representation of functional connectivity variation across the subcortex. Connectopic mapping (Haak et al., 2017) was used to map spatial gradients for each individual, which involved computing a sequence of eigendecompositions to yield a Laplacian eigenmap (Belkin et al.,2003).

In the analysis of personalized parcellation, a machine learning approach was used to personalize the group-consensus parcellations and distinguish individual differences in functional organization, yielding a unique parcellation for each individual. A binary support vector machine (SVM) classifier was trained to classify whether a voxel resided in the region or its uncertainty zone. One hundred individuals were randomly selected to train the SVM and the trained SVM was then applied to delineate a personalized parcellation for the remaining 921 individuals. The SVM feature space was defined over the symmetric similarity matrix, where each cell in this matrix stored the similarity in functional connectivity fingerprints between a pair of subcortex voxels. The feature vector for a particular voxel was selected as the relevant row in this matrix. The feature space was mapped separately for each individual and required computation of individual similarity matrices. A radial basis function kernel was used and the feature space was standardized. For each of the 921 individuals in the test set, the SVM was used to predict the likelihood (posterior probability) of each voxel belonging to the region under consideration, as opposed to belonging to the region's uncertainty zone.

In the analysis of relating subcortical functional network to behaviors, independent component analysis (ICA) was used to reduce the set of behavioral measures to a set of latent dimensions, with the aim of performing inference on the dimensions rather than the full set of behaviors. Before performing ICA, behavioral items with continuously distributed scores (87 of 109) were normalized with a rank-based inverse Gaussian transformation. Age and sex were then regressed from all items and ICA was performed on the residuals that resulted from this regression.