SUBGROUPS OF EATING BEHAVIOR TRAITS INDEPENDENT OF OBESITY DEFINED USING FUNCTIONAL CONNECTIVITY AND FEATURE REPRESENTATION LEARNING

Hyoungshin Choi[1,2], Kyoungseob Byeon[1,2], Jong-eun Lee[1,2], Seok-Jun Hong[2,3,4], Bo-yong Park[2,5]*, and Hyunjin Park[2,6]*

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea
[2] Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Republic of Korea
[3] Center for the Developing Brain, Child Mind Institute, New York, U.S.A
[4] Department of Biomedical Engineering, Sungkyunkwan University, Suwon, Republic of Korea
[5] Department of Data Science, Inha University, Incheon, Republic of Korea
[6] School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, Republic of Korea

*** Corresponding Authors:**

Bo-yong Park, Ph.D.
Department of Data Science
Inha University
Incheon, Republic of Korea
Phone: +82-32-860-9427
Email: boyong.park@inha.ac.kr

Hyunjin Park, Ph.D.
School of Electronic and Electrical Engineering
Sungkyunkwan University
Suwon, Republic of Korea
Phone: +82-31-299-4956
Email: hyunjinp@skku.edu

Abstract

Eating behavior is highly heterogeneous across individuals, and thus, it cannot be fully explained using only the degree of obesity. We utilized unsupervised machine learning and functional connectivity measures to explore the heterogeneity of eating behaviors. This study was conducted on 424 healthy adults. We generated low-dimensional representations of functional connectivity defined using the resting-state functional magnetic resonance imaging, and calculated latent features using the feature representation capabilities of an autoencoder by nonlinearly compressing the functional connectivity information. The clustering approaches applied to latent features identified three distinct subgroups. The subgroups exhibited different disinhibition and hunger traits; however, their body mass indices were comparable. The model interpretation technique of integrated gradients revealed that these distinctions were associated with the functional reorganization in higher-order associations and limbic networks and reward-related subcortical structures. The cognitive decoding analysis revealed that these systems are associated with reward- and emotion-related systems. We replicated our findings using an independent dataset, thereby suggesting generalizability. Our findings provide insights into the macroscopic brain organization of eating behavior-related subgroups independent of obesity.

Keywords: eating behavior; subgroup; functional connectivity; autoencoder; manifold learning; representation learning; integrated gradient

Introduction

Eating behavior is a key trait associated with an individual's health [1,2]. Aberrant eating behavior can lead to a high body mass index (BMI) and cause obesity-related pathologies, such as diabetes, hypertension, and stroke [3,4]. To assess the link between eating behavior and obesity, existing studies have examined several factors that affect an individual's eating behaviors, such as hormone activity, gene enrichment, and environmental factors [4–8]. Eating behavior is highly heterogeneous across individuals, and thus, a systematic analysis is necessary to assess individual variability.

Magnetic resonance imaging (MRI) is used to investigate brain networks associated with eating behaviors *in vivo* [9–11]. In particular, resting-state functional MRI (rs-fMRI) reflects functional alterations in the brain via temporal fluctuations in brain signals. Our previous study demonstrated associations between disinhibited eating behaviors and functional connectivity in the frontoparietal network [10]. Other studies have proved associations of eating behaviors with the brain function of the prefrontal cortex, orbitofrontal cortex, and amygdala [12–15]. These findings suggest that eating behavior is associated with the brain function. However, no clear trends were observed. Some studies have shown positive associations between disinhibited eating and brain function in the reward network [12,13], whereas others have indicated opposite patterns [14,15]. This inconsistency may be owing to the heterogeneity of eating behavior traits. Thus, the brain function differences between individuals depending on the eating behavior needs to be investigated systematically.

One approach to exploring the heterogeneity of eating behaviors is clustering, which is an unsupervised machine learning technique that defines distinct clusters with relatively homogeneous data points. Clustering techniques were widely adopted in existing neuroimaging studies to identify subgroups of healthy and diseased populations [16–20]. Some studies have classified individuals with an autism spectrum disorder into several subtypes based on cortical morphologies [21] and functional connectivity [16,17]. In addition, the clustering approach was effective in schizophrenia to assess clinical heterogeneity [19,20]. Clustering techniques are purely data-driven approaches, free from an a priori hypothesis. Thus, they can be used for identifying subgroups of a particular dataset with homogeneous characteristics. We hypothesized that clustering based on neuroimaging features may identify distinct subgroups that may exhibit different clinical or behavioral traits.

Functional connectivity is a widely adopted measure to assess co-fluctuations of the brain signals, which is defined by calculating correlations of time series between brain regions. A recent study suggested a method for characterizing functional connectivity based on manifold learning techniques [22]. These techniques produce low-dimensional representations of functional connectivity by estimating principal components based on principal component analysis or scaled eigenvectors that are based on diffusion embedding in a newly defined low-dimensional space. The generated eigenvectors exhibited smooth transitions of connectome organization along the cortical mantle, and the principal eigenvector consisted of a cortical hierarchy expanding from low-level sensory to higher-order association networks [22]. These eigenvectors have been suggested as potential imaging biomarkers in studies on healthy aging [23,24] and neurodevelopment [25–30]. In our previous study, we illustrated strong associations between the BMI and low-dimensional representations of functional connectivity, indicating plausible links between functional gradients and eating behaviors [31]. Recent advances in machine learning have made strides in feature representation to learn novel features from an existing set of features for various downstream machine-learning tasks. In particular, an autoencoder creates latent features that effectively describe the original features through nonlinear data compression and reconstruction [32–34]. Autoencoders have been adopted in some studies to distinguish populations of Alzheimer's disease [35,36], schizophrenia [37,38], and autism [39] from healthy individuals. The feature representation capability of the autoencoder led to a higher performance in solving classification problems compared with conventional neuroimaging features.

In this study, we combined connectome manifolds with feature representation to identify subgroups of eating behavior traits. Briefly, we generated eigenvectors from the functional connectivity matrix and constructed an autoencoder model to identify subgroups with different behavioral traits. Subsequently, we compared the characteristics of eating behavior traits and degree of obesity among subgroups and assessed between-group differences in cortico-cortical and subcortico-cortical connectivity. Additionally, we assessed the reproducibility of our findings using an independent dataset.

Result

We studied 424 healthy adults obtained from the enhanced Nathan Kline Institute-Rockland Sample database (mean ± standard deviation [SD] age = 47.07±18.89 yr; 67% female; mean ± SD BMI = 27.82±5.77 kg/m^2, range 16.26–47.93 kg/m^2) [40]. Details of the participant selection, image processing, and analysis are described in the *Methods* section. Reproducibility of the findings was validated using an independent dataset, Leipzig Study for Mind-Body-Emotion Interactions database, which contained 212 healthy adults (mean ± SD age = 38.97±19.80 yr; 35% female; mean ± SD BMI = 24.17±3.67 kg/m^2, range 17.93–36.65 kg/m^2) [41].

Low-dimensional representation of functional connectivity and autoencoder-based feature representation

For each individual, we built a functional connectivity matrix by calculating the Pearson's correlation of the time series between different brain regions defined using the Brainnetome atlas [42]. Considering 210 cortical parcels, we computed a low-dimensional representation of the functional connectivity (henceforth, eigenvector) [22] using dimensionality reduction techniques implemented in the BrainSpace toolbox (https://github.com/MICA-MNI/BrainSpace; see *Methods*) [43]. Individual eigenvectors were linearly aligned to the template manifold and computed using the group-averaged functional connectome [43,44]. We selected three eigenvectors (E1, E2, E3) that explained approximately 54% of the information of the template affinity matrix (**Figure 1A**). Similar to the previous findings based on the Human Connectome Project dataset [22,31,43], each eigenvector exhibited different cortical axes; the first, second, and third eigenvectors expanded from the primary sensory to association cortices (E1), from visual to somatomotor (E2), and from the multiple demand network to task-negative systems (E3), respectively.

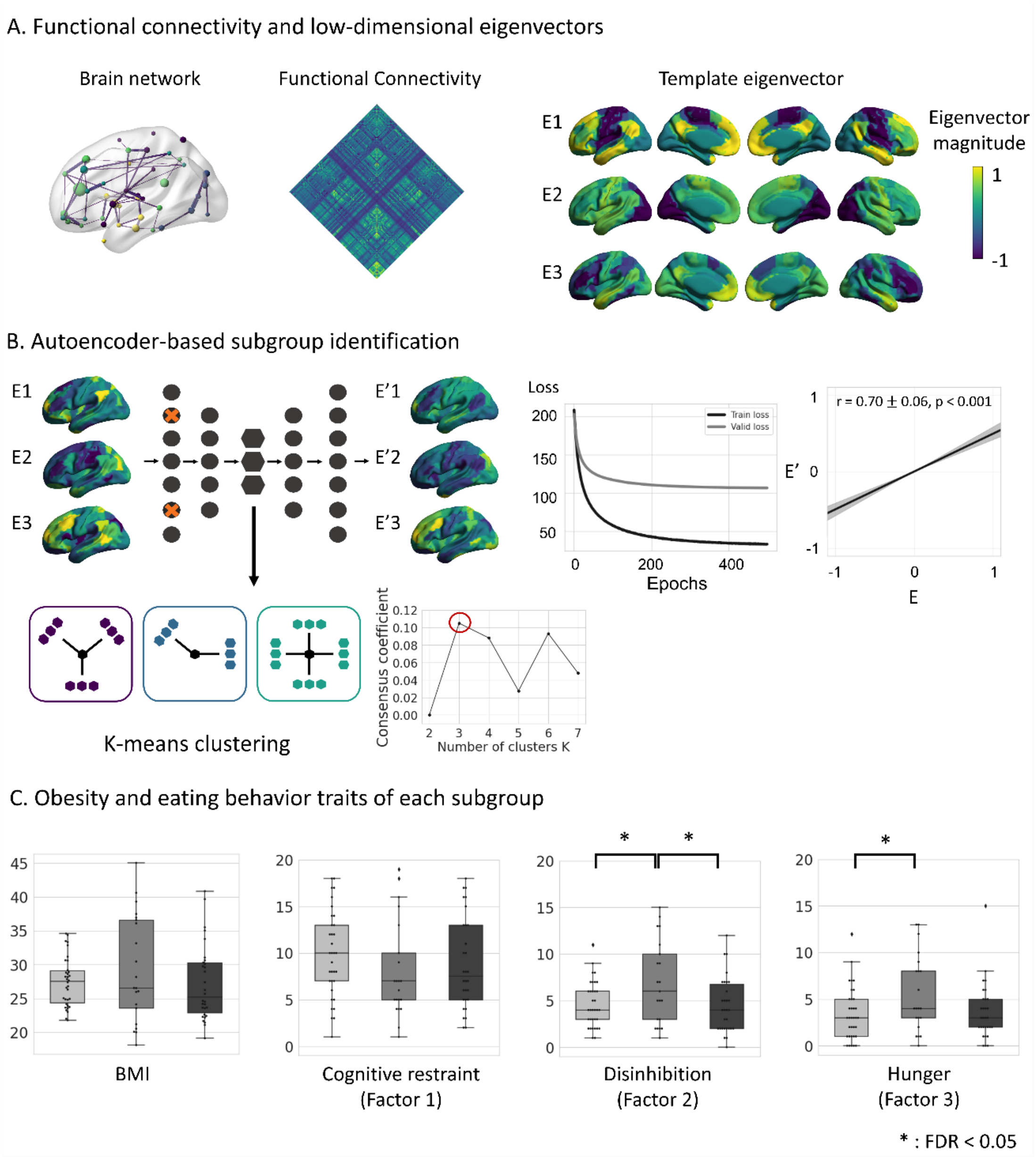


Figure 1 | Subgroup identification using the manifold learning and autoencoder-based feature representation. (A) Schematic of functional connectome organization (left) and group averaged functional connectivity matrix (middle) are reported. Template eigenvectors were generated using dimensionality reduction techniques, and three dominant eigenvectors (E1, E2, E3) were selected. **(B)** The autoencoder model learned latent features of the eigenvectors after controlling for age and sex (left top), and loss values are plotted for each epoch (middle). We calculated linear correlations between the original (E) and reconstructed (E') eigenvectors of the test dataset, and correlation coefficients across the subjects are reported with mean ± SD (right). We defined subgroups using the latent features of the autoencoder, where the number of clusters was determined using the consensus coefficient (left bottom). **(C)** Distribution of BMI and eating behavior scores of each subgroup is plotted. Significant differences in scores between subgroup pairs are indicated by asterisks. *Abbreviations:* BMI, body mass index; FDR, false discovery rate; SD, standard deviation.

The three generated eigenvectors were concatenated and used as inputs for the autoencoder model. The autoencoder model extracts latent features of the input through compression (i.e., encoding) and reconstruction (i.e., decoding) procedures (see *Methods*). The loss graph with respect to epochs demonstrated that the loss values decreased in both the training and validation datasets, indicating the appropriateness of model fitting. We applied the trained model at 499 epochs, which exhibited the highest performance in the validation dataset to the test data and found significant correlations between the original and reconstructed eigenvectors (mean±SD $r = 0.70 \pm 0.06$ across the subjects, $p<0.001$), indicating that the autoencoder learned the eigenvectors appropriately (**Figure 1B**).

Subgroup identification using features from representation learning

The latent features learned from the autoencoder (i.e., features from the hidden bottleneck layer in the middle) were subjected to an unsupervised learning framework to identify subgroups of the study population. In particular, we employed the k-means clustering, and the number of clusters was determined using the consensus clustering approach, which was set to three [45] (see *Methods*; **Figure 1B**). To assess differences between the obesity-related traits across the identified subgroups, we compared the BMI and eating behavior scores of the subgroups based on a three-factor eating questionnaire (TFEQ) [46]. We found significant differences (false discovery rate (FDR)<0.05) in the disinhibition and hunger scales in one subgroup; however, BMI did not exhibit considerable differences (**Figure 1C**). Thus, the identified subgroups may reflect different eating behavior traits, independent of the degree of obesity.

Interpretation of the latent features from representation learning

We utilized the integrated gradient interpretation model to explain the autoencoder-derived latent features [47]. The integrated gradient method computes the attribution of each element (i.e., the brain region) of the input to predict the output (i.e., latent features of the bottleneck) by progressively increasing the intensity of input values from a zero-information baseline to a particular intact input level and averaging attributions (**Figure 2A**) [47]. We used the integrated gradient technique to identify the brain regions that contributed to the latent features in the hidden layer that contained important information to reconstruct the original data (**Figure 2B**). We considered the three integrated gradient maps of eigenvectors and found that most higher-order networks, including limbic, dorsal attention, frontoparietal, and default mode networks, exhibited high contributions (**Figure 2C**). The results indicate that higher-order association and limbic regions greatly contributed to the reconstruction of the original eigenvectors.

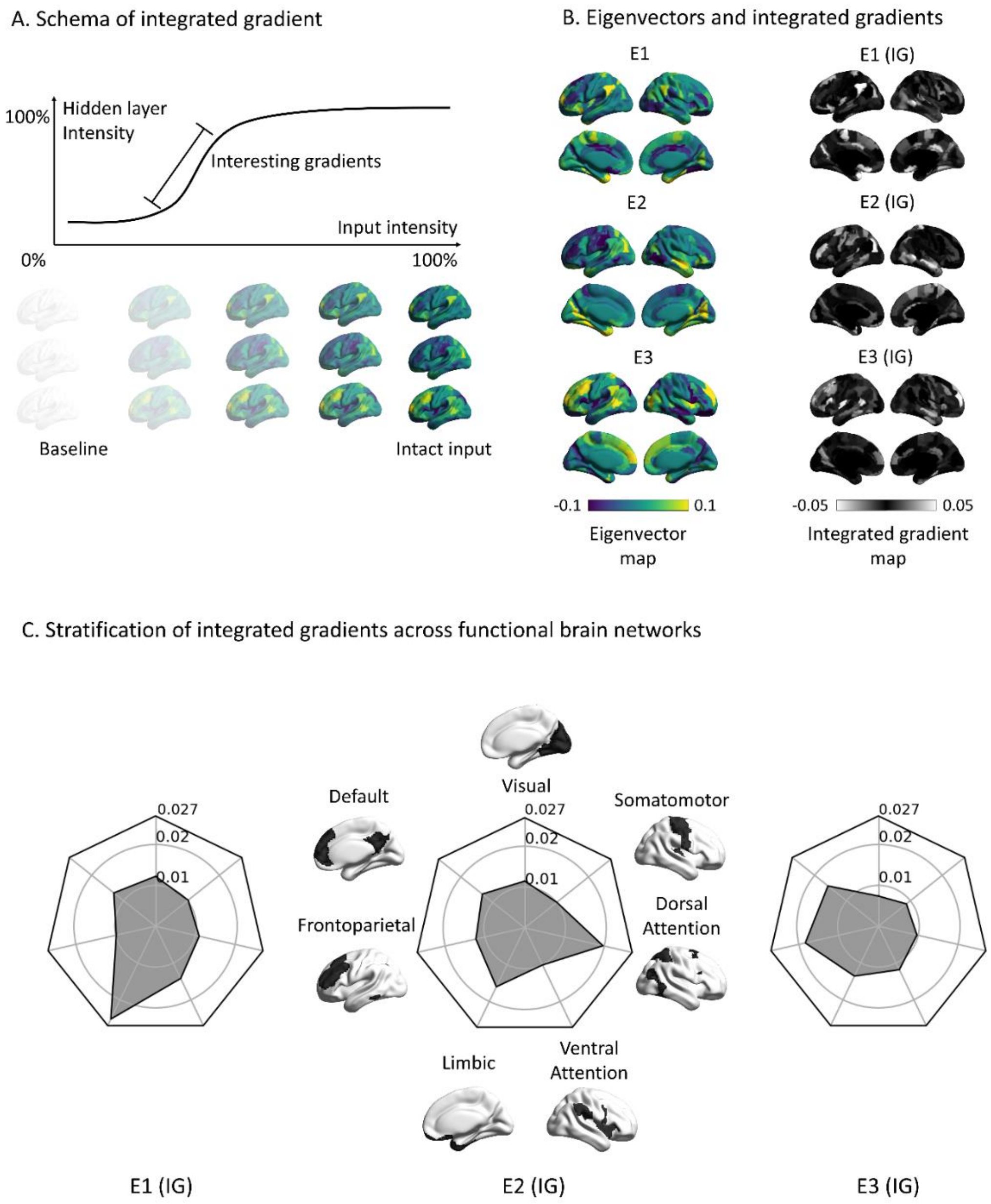
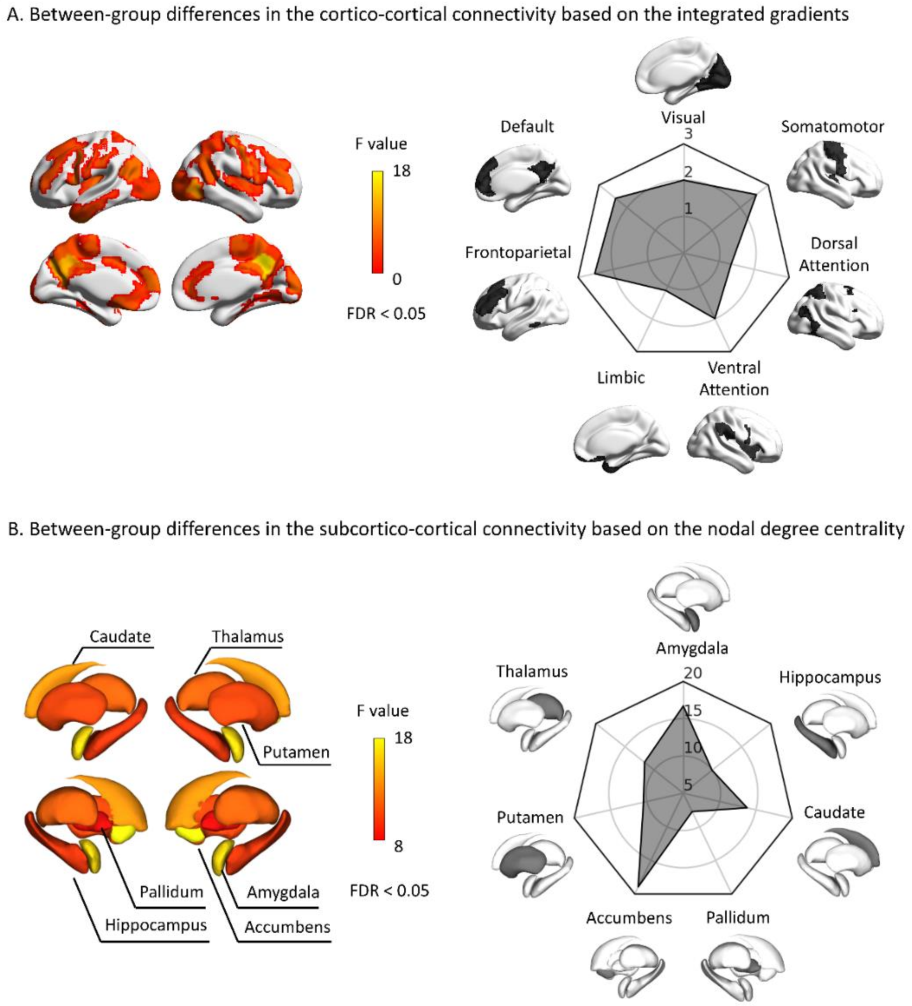


Figure 2 | Characteristics of latent features using the integrated gradient technique. (A) Integrated gradient technique estimates the attribution of input towards predicting the output by averaging contributions while changing input intensities. **(B)** Spatial maps of each eigenvector and results of the integrated gradient technique are plotted on the brain surfaces. **(C)** Effects of integrated gradient are summarized according to functional communities. *Abbreviation:* IG, integrated gradient.

Cortico-cortical and subcortico-cortical connectivity of subgroups

In addition to cortical alterations among subgroups, we hypothesized that connectivity in the reward circuit, which is known to be highly associated with the eating behavior may exhibit distinct profiles among subgroups. To prove our hypothesis, we investigated differences in the cortico-cortical connectivity based on integrated gradient maps among the three subgroups using the multivariate analysis of variance (MANOVA) [48]. We found significant differences in the precuneus, with the strongest and moderate effects in the frontoparietal and sensory regions (FDR<0.05; **Figure 3A**). Stratifying the effects according to the seven functional communities [49], somatomotor, ventral attention, frontoparietal, and default mode networks revealed strong between-group differences (**Figure 3A**). Additionally, we assessed between-group differences in the subcortico-cortical connectivity based on nodal degree centrality using ANOVA (see *Methods*). All subcortical structures exhibited considerable effects (FDR<0.05; **Figure 3B**), and stronger effects were observed in the accumbens, amygdala, and caudate (**Figure 3B**), which are involved in the reward system.


A. Between-group differences in the cortico-cortical connectivity based on the integrated gradients

B. Between-group differences in the subcortico-cortical connectivity based on the nodal degree centrality

Figure 3 | Between-group differences in the cortico-cortical and subcortico-cortical connectivity. (A) Between-group differences in the cortico-cortical connectivity based on the integrated gradients maps among the subgroups are visualized on the cortical surfaces, where the findings were multiple comparisons corrected using

FDR<0.05. Effects were stratified according to seven intrinsic functional communities. **(B)** We visualized between-group differences in subcortico-cortical nodal connectivity strengths and stratified the effects according to each subcortical structure. *Abbreviation*: FDR, false-positive discovery rate.

Cognitive associations

To provide the underlying cognitive associations of between-group differences in the cortico-cortical and subcortico-cortical connectivity, we utilized the Neurosynth meta-analysis cognitive decoding platform [50,51]. Associating the between-group differences in cortical and subcortical maps (**Figure 3**), we found high correlations with the reward-related terms, such as “anticipation,” “reward,” “incentive,” “monetary,” and “gain” (**Figure 4A**). Additionally, we associated the between-group difference maps with 24 cognitive state maps, as defined in [22], and observed strong correlations with “reward” (r = 0.40, FDR<0.05), and high associations with “emotion” (r = 0.24, FDR<0.05) and “affective” (r = 0.21; FDR<0.05; **Figure 4B**). Consequently, differences in the cortico-cortical and subcortico-cortical connectivity across the subgroups are associated with the reward-related cognitive functioning.

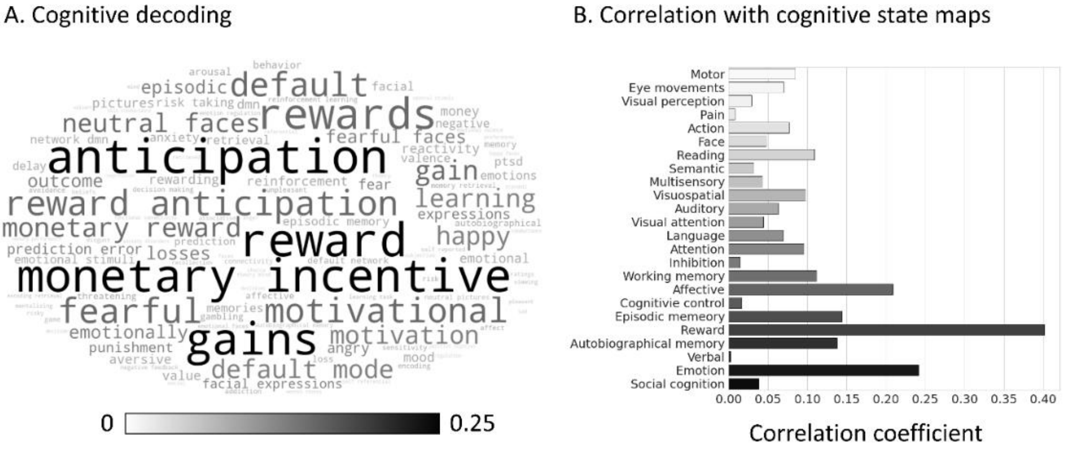


Figure 4 | Cognitive associations. (A) We conducted cognitive decoding using the F-statistic map of cortico-cortical and subcortico-cortical connectivity differences across the subgroups using Neurosynth. **(B)** Correlation coefficients between the between-group difference maps and 24 different cognitive state maps are shown with bar plots.

Replication of eating behavior traits

We performed the entire subgrouping analysis and compared the obesity and eating behavior traits across subgroups by using an independent dataset with different acquisition parameters (see *Methods*). One subgroup exhibited significant differences (FDR<0.05) in the hunger scale (**Figure 5A**), which particularly confirms the robustness of the hunger scale results. Notably, the subgroups defined from the two datasets exhibited comparable profiles (**Figure 5B**), with no statistical differences.

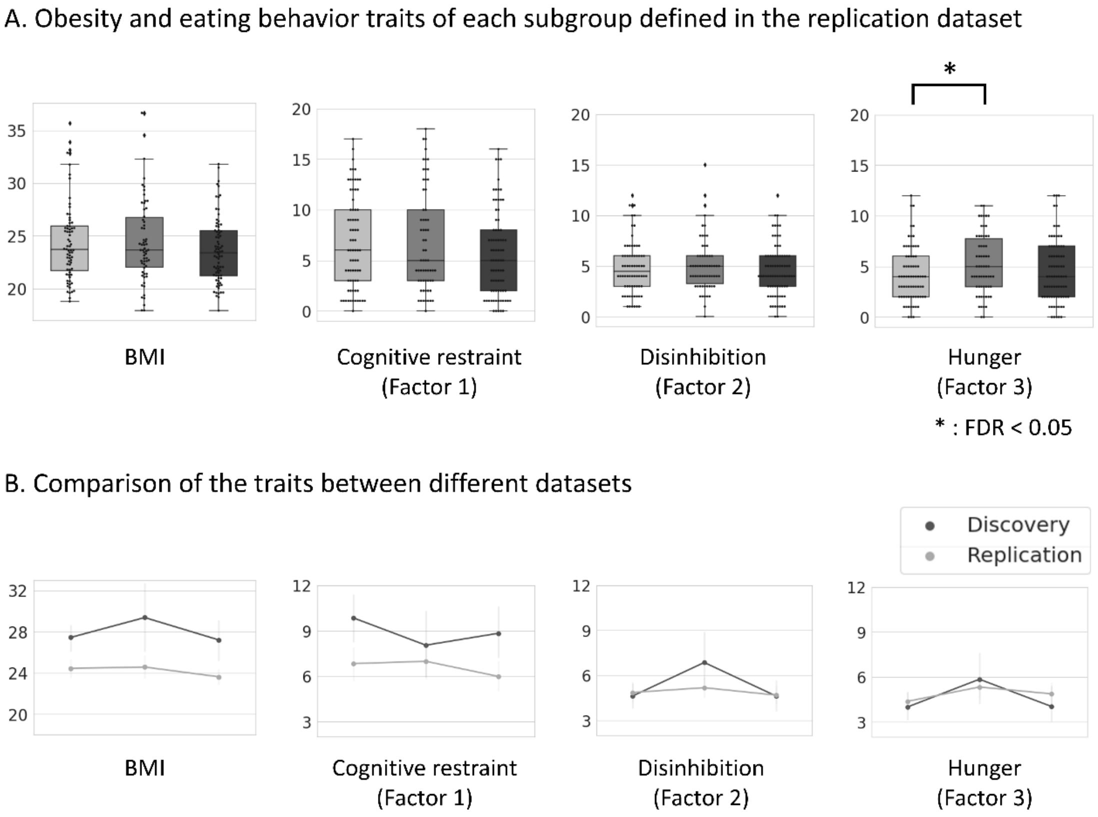


Figure 5 | Reproducibility analyses. (A) Distribution of the BMI and eating behavior scores of each subgroup using the replication dataset are shown. Significant differences between the scores of subgroups are indicated by asterisks. **(B)** We compared profiles of the BMI and eating behavior scores of two different datasets. *Abbreviation:* BMI, body mass index; FDR, false discovery rate.

Sensitivity analyses

Robustness of the findings was confirmed using the conducted analyses.

a) Subgroup identification without autoencoder. We applied k-means clustering to the concatenated eigenvectors and not to the latent features from the autoencoder. No significant differences were found between the BMI or eating behavior scores (**Figure S1**), indicating the necessity of applying the autoencoder model.

b) Bootstrapping analysis. We randomly selected 90% of participants with replacements and performed the same analyses of feature representation learning, clustering, and profiling of the obesity and eating behavior scores. We obtained consistent results that indicated the robustness of the findings (**Figure S2**).

c) Different densities of connectivity matrix. As in previous studies [22,26,43], our main findings were based on the functional connectivity with a density of 10%. Additionally, we performed the analyses based on the matrix with a density of 20 and 30% to evaluate the robustness and found consistent results (**Figure S3A**).

d) Different clustering methods. Instead of k-means clustering, we used the Gaussian mixture model clustering method, which is a probability distribution-based clustering approach (**Figure S3B**) and found that the obesity and eating behavior scores of each subgroup exhibited comparable profiles.

e) Different model architectures. We slightly changed the architecture of the autoencoder model and generated latent features. (i) We removed dropout layers; (ii) added one more layer; and (iii) removed one layer during the encoding and decoding processes (see *Methods*). The obesity and eating behavior scores were not considerably different and exhibited similar trends (**Figure S4**).

f) Manifold eccentricity. Rather than utilizing an autoencoder model to compress three eigenvectors into a single latent feature, we adopted the manifold eccentricity analysis, which depicts the distance of each brain region from the center of the template manifold [23,29] (**Figure S5A**). We defined subgroups based on the manifold eccentricity. Here, the BMI and eating behavior scores did not exhibit considerable differences, indicating that the latent features, which were defined using the autoencoder model were more useful for identifying behavioral differences across the identified subgroups (**Figure S5B**).

Discussion

Eating behavior is highly associated with the brain function [10,31]; however, owing to the heterogeneity of eating behavior traits, no conclusions have been made to relate eating behaviors with the brain. In this study, an advanced technique combining the dimensionality reduction (i.e., connectome manifolds) with representation learning of the autoencoder was used to identify three subgroups with different eating behavioral traits independent of BMI. The latent features learned from the autoencoder were used to subdivide the groups, which yielded three distinct subgroups with different eating behaviors on the disinhibition and hunger scales. Furthermore, differences in the cortico-cortical connectivity from integrated gradients and subcortico-cortical nodal strengths among the identified subgroups were associated with reward-related cognitive terms, indicating that alterations in the association and reward systems may be related to the eating behavior heterogeneity. We demonstrated the reliability and generalizability of our findings using an independent dataset with different acquisition parameters and demographic characteristics.

Machine learning is a powerful tool for analyzing neuroimaging data, and manifold learning techniques are increasingly being used to describe macroscale functional connectome organization along the cortex [27–29,52]. We generated a series of low-dimensional eigenvectors, and the spatial patterns agreed with those of existing studies based on the Human

Connectome Project data [22,31,52]. We extended previous studies by analyzing these eigenvectors using an autoencoder model consisting of encoding and decoding processes to generate latent features that contain highly compressed information of the input data. Using the latent features, we obtained three distinct subgroups, which exhibited considerable differences in eating behaviors. Indeed, our previous studies based on the conventional graph-theoretical approaches found that the functional connectivity of frontoparietal and executive control networks is associated with disinhibited eating behaviors and eating concerns [10,53]. In addition, we revealed that external stimulation of the dorsolateral prefrontal cortex affected the brain function in the frontoparietal network, which in turn yielded a reduced appetite [54]. Our current findings complemented previous studies in that eating behaviors are related to the function of higher-order brain regions. Additionally, we extended these studies by defining subgroups of study participants to investigate the heterogeneity of eating behavior traits. Interestingly, eating behaviors on the disinhibition and hunger scales exhibited significant between-group differences at a similar BMI. Therefore, although eating behavior is highly associated with obesity, the neurological underpinnings of eating behaviors may be different from those related to obesity. Further comprehensive investigations are needed to assess the convergence and divergence between obesity and eating behavior traits with respect to the brain function to explore the underlying neurological mechanisms of their relationships.

The interpretation of neural network models is important for accurate disease diagnosis and precise decision-making processes [57–59], which is uncertain because of the complex combinations of nonlinearities of the model [55,56]. Several attempts have been increasingly made [57,60–67]. Representative techniques include the layer-wise relevance propagation (LRP) and class activation map (CAM). LRP redistributes the relevance of each node from a particular layer to the previous layer using a top-down process [68], and CAM calculates the weighted sum of feature maps at the last convolutional layer, where the weights are calculated at a fully connected layer connected to the last convolutional layer using global average pooling [69]. These techniques have been used for the diagnosis of multiple sclerosis and Alzheimer's disease [65,67]. However, LRP is sensitive to the choice of network architecture [47], and CAM requires global average pooling. To overcome this limitation, a recent study introduced gradient-weighted CAM (Grad-CAM) using gradients contributing to specific outputs as weights to provide an explanation [70]. Integrated gradients expanded the prior methods, enforcing a few axiomatic properties that have a invariance toward different neural network implementations [47]. In this study, we used the integrated gradients approach and observed that limbic, dorsal attention, frontoparietal, and default mode networks exhibited strong attributions, indicating that subgroups might present distinct functional organization in large-scale networks of higher-order association and limbic regions.

Additional cortico-cortical and subcortico-cortical connectivity analyses and cognitive decoding approved that association and reward networks exhibited significant between-group differences among the identified subgroups. These systems are related to eating behaviors [4,10,14,71–74], and thus, features attributed to these networks may provide benefits in identifying subgroups related to eating behaviors. In particular, reward systems are highly associated with eating behaviors, where an imbalance between food-related reward circuits and inhibitory control systems yields an increased sensitivity to food, leading to overeating and weight gain [12,75–87]. In addition, reward circuits are regulated by dopamine-related neurotransmitters, where the atypical organization of dopaminergic circuits in mesolimbic and association cortices was observed in individuals with obesity [4,88–94]. Similarly, serotonin-related neurotransmitters control eating behaviors by inhibiting the hanger-stimulating system [95,96]. These studies collectively suggest that reward-related cognitive systems could be target regions for regulating eating behaviors independent of the degree of obesity. However, expanded validation is required to explore the biological mechanisms associated with these macroscopic brain alterations.

In this study, we identified subgroups showing different eating behavior traits, regardless of the degree of obesity by using connectome manifolds and feature representation learning. The findings were robust for independent datasets, thus suggesting generalizability. Although we interpreted latent features derived from the autoencoder model based on an integrated gradient approach, this technique measures indirect contributions of the input data. More advanced techniques for the direct inference of the contribution of latent features need to be considered in future studies. Our results provide a new evidence for eating behavior-related macroscopic imaging signatures, independent of obesity.

METHOD

Participants

Imaging and phenotypic data were obtained from the enhanced Nathan Kline Institute-Rockland Sample database (NKI-RS) [40]. We excluded participants who did not provide complete demographic information, BMI, or TFEQ scores. Of the 650 participants, 424 were selected for this study. The proportion of individuals with healthy weight (18.5≤BMI<25 kg/m^2), overweight (25≤BMI<30), and obese (BMI≥30 kg/m^2) was 144:151:121. In addition, we obtained independent data from the Leipzig Study for Mind-Body-Emotion Interactions (LEMON) database [41]. Participants without complete demographic information or obesity-related scores were excluded. In this study, we used 212 out of 229 participants. Details are presented in **Table 1**.

Table 1. Demographic information of study participants

Categories		NKI-RS	LEMON	p value[1]
Subject number		424	212	
Sex (Female:Male)		282:142	75:137	<0.001[2]
Age		47.07±18.89 (18.15–85.62)	38.97±19.80 (23–78)	<0.001
BMI		27.82±5.77 (16.26–47.93)	24.17±3.67 (17.93–36.65)	<0.001
TFEQ	Cognitive restraint (Factor 1)	8.67±4.94 (0 – 20)	6.36±4.65 (0–18)	<0.001
	Disinhibition (Factor 2)	5.03±3.45 (0–15)	4.82±2.61 (0–15)	0.39
	Hunger (Factor 3)	4.26±3.36 (0–15)	4.61±2.94 (0–12)	0.17

Mean±SD with range (minimum–maximum) are reported.
[1]The p values are calculated based on two sample *t*-tests between the discovery and replication datasets.
[2]Chi-squared test.
Abbreviations: NKI-RS, enhanced Nathan Kline Institute-Rockland Sample; LEMON, Leipzig Study for Mind-Body-Emotion Interactions; BMI, body mass index; TFEQ, three-factor eating questionnaire.

MRI acquisition

a) NKI-RS: All imaging data were obtained using a 3-T Siemens Magnetom Trio Tim scanner. The acquisition parameters of T1-weighted data were as follows: repetition time (TR), 1900 ms; echo time (TE) = 2.52 ms, flip angle, 9°; field of view (FOV), 250 mm × 250 mm; voxel resolution = 1 mm^3 isotropic, and number of slices, 176. The rs-fMRI data were as follows: TR = 645 ms, TE = 30 ms, flip angle = 60°, FOV = 222 mm × 222 mm, voxel resolution = 3 mm^3

isotropic, number of slices = 40, and number of volumes = 900.

b) LEMON: LEMON imaging data were acquired using a Siemens 3 Tesla scanner equipped with a 32-channel head coil. Scanning parameters of the T1-weighted data were as follows: TR = 5000 ms, TE = 2.92 ms, inversion time 1 (TI1) = 700 ms, TI2 = 2,500 ms, flip angle 1 (FA1) = 4°, FA2 = 5°, echo spacing = 6.9 ms, bandwidth = 240 Hz/pixel, FOV = 256 mm, voxel resolution = 1 mm^3 isotropic, acceleration factor = 3, and number of slices = 176. The parameters of the rs-fMRI data were as follows: TR = 1400 ms, TE = 30 ms, flip angle = 69°, FOV = 202 mm, voxel resolution = 2.3 mm^3 isotropic, number of slices = 64 slices, number of volumes = 657, and multiband acceleration factor = 4.

Data preprocessing

a) NKI-RS: The T1-weighted and rs-fMRI data were preprocessed using the fusion of neuroimaging preprocessing (FuNP) volume-based pipeline, which combines the AFNI, FSL, and ANTs software [97–100]. The magnetic field inhomogeneity of the T1-weighted data was corrected, and nonbrain tissues were eliminated. The rs-fMRI data were preprocessed as follows: the first 10 s of the volume were discarded, and head movements were corrected. The FIX software was used to eliminate nuisance variables, such as the cerebrospinal fluid, white matter, head motion, and cardiac- and large-vein-related abnormalities [101]. The artifact-free rs-fMRI data were registered onto the preprocessed T1-weighted data and subsequently onto the MNI152 standard space. Spatial smoothing with a full-width-at-half-maximum of 5 mm was applied.

b) LEMON: The T1-weighted data preprocessing was performed based on Nipype; the details are described in (*https://github.com/NeuroanatomyAndConnectivity/pipelines/tree/master/src/lsd_lemon)* [102]. In brief, CBS Tools [103] were used to remove the background from the T1-weighted image, and masked images were used to reconstruct cortical surfaces using FreeSurfer [104,105]. The T1-weighted data were registered onto the MNI152 standard space based on the diffeomorphic nonlinear registration using ANTs [100]. The de-identification process was performed using CBS Tools [103] by applying a brain mask to all anatomical scans. The rs-fMRI data were preprocessed using Nipype [102]. The pipeline included the following steps: the first five volumes were discarded to allow for signal equilibration and steady-state conditions [106]. Head motion and MRI-induced distortions were corrected [99]. The rigid-body transformation was applied to co-register the rs-fMRI data with the anatomical image [107]. Denoising was based on Nipype rapidart and aCompCor [108], and band-pass filtering in the frequency range of 0.01–0.1 Hz was applied. Standardization of mean centering and variance normalization was performed [109], and the preprocessed data were registered onto the MNI152 standard space [100].

Eigenvector generation

We constructed a functional connectivity matrix from the preprocessed rs-fMRI data (**Figure 1A**) by calculating Pearson's correlation of the time series between two different regions. Brain regions were defined using the Brainnetome atlas [42] and a cortico-cortical functional connectivity matrix with a size of 210 × 210. The correlation coefficient was Fisher's r-to-z transformation [110]. We generated principal eigenvectors of functional connectivity using the BrainSpace toolbox (*https://github.com/MICA-MNI/BrainSpace*) [43]. The diffusion map embedding algorithm [111], which is robust to noise and computationally efficient was used to estimate eigenvectors from the functional connectivity matrix, leaving only the top 10% elements per row. Eigenvectors of each individual were aligned to group-level template eigenvectors defined based on a group-averaged functional connectome via Procrustes alignment (**Figure 1A**) [43,44]. The age and sex were controlled using eigenvectors.

Architecture of the autoencoder model

An autoencoder was used to generate latent features from concatenated eigenvectors. The autoencoder model is defined as follows:

$$h = e_n(U) \ = tan\,h(WU + b\,),$$

$$U' = d_n(h) \ = tan\,h(Wh + b),$$

$$L = \sum \frac{|U - U'|^2}{n},$$

where n is the size of the input vector and e_n is the encoder, which transforms the input vector U into a feature representation (or hidden representation bottleneck) layer h. Furthermore, h is used to reconstruct the input data and generate U' using the decoder d_n. The model was trained by minimizing the sum of mean square errors L between the input (U) and output (U'). The autoencoder model consists of two encoder layers, two decoder layers, and one feature representation layer. The feature representation layer had 210 latent variables, and each encoder and decoder layer had 630 and 420 units, respectively (**Figure 1B**). We used a hypertangent activation function in all layers, and a dropout rate of 0.3 was applied in the input layer [112]. The model was optimized using the Adam optimizer [113] with a learning rate of 1e^{-4}, batch size of 10, and weight decay (i.e., L2-regularization) of 0.1. We concatenated three eigenvectors that provided sufficient information on the total functional connectivity data and entered it into the autoencoder model. We divided the dataset into training, validation, and test datasets with the ratios of 60, 20, and 20%, respectively. We trained the model using the training data and validated its performance using the validation data. We selected the model that exhibited the highest performance in the validation dataset for a total of 500 epochs. The selected model was applied to the test dataset, and its performance was assessed by calculating linear correlations between U and U'.

Subgroup identification

Latent features in the hidden representation layer h were used to define the participant subgroups. We applied k-means clustering, which is based on the Euclidean distance. The optimal number of subgroups was determined using the consensus clustering, which robustly assesses how a pair of data is assigned to the same cluster (i.e., consensus coefficient) [45]. We determined the optimal number of subgroups, in which the largest consensus coefficient occurred, while varying the number of clusters. To assess the clinical and behavioral traits of subgroups, we compared the measured BMI and eating behavior scores using the TFEQ. The TFEQ consists of 51 questions [46], and each element is assigned to one of the three domains: (i) cognitive restraint, (ii) disinhibition, and (iii) hunger. We applied a two-sample *t*-test to compare each score between the subgroup pairs (**Figure 1C**). Significance was assessed using 1,000 permutation tests by randomly shuffling participants. A null distribution was constructed, and the real *t*-statistic value was deemed significant if it did not belong to 95% of the distribution (two-tailed p<0.05). Multiple comparisons were corrected using FDR [114].

Integrated gradient for model explanation

To interpret latent features in the hidden representation layer, we assessed the attribution of each brain region to generate latent features using an integrated gradient approach (**Figure 2A**) [47]. The integrated gradient provides information on the extent to which a specific element (i.e., the brain region) in the input data contributes to predicting the output data (i.e., latent features). In particular, the integrated gradient (IG) from the i^{th} neuron is defined as follows:

$$IG_i(x) = (x_i - z_i) * \int_{\alpha=0}^{1} \frac{\delta f\big(z + \alpha * (x - z)\big)}{\delta x_i} d\alpha,$$

where x is the input data, z denotes the baseline, and α is the interpolation constant. The path integral can be approximated as follows:

$$IG_i(x) = (x_i - z_i) * \frac{1}{M} \sum_{m=1}^{M} \frac{\delta f\left(z + \frac{m}{M} * (x - z)\right)}{\delta x_i},$$

where m and M are the number of steps in the scaled feature perturbation constant and Riemann sum approximation of the integral, respectively. If the output has significantly changed, we assume that the attribution of the input data is high and vice versa. Thus, we can assess which brain regions of the input data considerably contributed during the compression and reconstruction processes. After calculating integrated gradients of each eigenvector, we stratified the effects according to the functional networks [49] to assess the networks that contributed the most to the reconstruction of the original data (**Figures 2B and C**).

Between-group differences in cortico-cortical and subcortico-cortical connectivity

We compared the cortico-cortical connectivity computed from the three integrated gradient maps across subgroups using MANOVA (**Figure 3A**), where multiple comparisons were corrected using FDR [114]. The effects were stratified according to seven intrinsic functional networks [49]. In addition, we compared the subcortico-cortical connectivity across the subgroups using ANOVA (**Figure 3B**). The subcortico-cortical connectivity was quantified using the nodal degree centrality, a widely used graph-theoretical measure calculated by summing the connectivity strength of a particular brain area [115–117]. The nodal degree was estimated from the functional connectivity matrix, leaving only the top 10% elements per row. Multiple comparisons were corrected using FDR.

Associations with cognitive states

Additionally, we assessed the relationships among the between-group differences in the cortico-cortical and subcortico-cortical connectivity across subgroups with cognitive terms using Neurosynth [50,51]. Neurosynth decodes the input data based on a meta-analytical method and provides correlation coefficients, whose cognitive terms are related to the data (**Figure 4A**). To systematically assess hierarchically organized cognitive maps, we performed spatial correlations between the between-group difference map in eigenvectors and 24 cognitive state maps, as defined in [22] (**Figure 4B**).

Reproducibility experiments

We performed comparing eating behavior traits among subgroups to validate the generalizability of our results using the LEMON dataset [41]. We transferred the autoencoder model trained using the NKI-RS dataset into the LEMON dataset and applied the k-means clustering to identify subgroups. The obesity and eating behavior scores were profiled across the subgroups (**Figure 5A**), and the profiles were compared between the two datasets (**Figure 5B**).

Sensitivity analyses

a) Subgroup identification without autoencoder. We applied the k-means clustering to the concatenated eigenvectors and not to latent features from the autoencoder in order to evaluate the effect of latent features on profiling clinical and behavioral traits (**Figure S1**).

b) Bootstrapping analysis. We performed 1,000 bootstraps with 90% resampled data to demonstrate the robustness of our results (**Figure S2**).

c) Different densities of connectivity matrix. We computed eigenvectors using different connectivity matrix densities of 20 to 30% and repeated the analyses (**Figure S3A**).

d) Different clustering methods. Instead of the k-means clustering, we used the Gaussian mixture model clustering approach, which creates clusters based on a probability distribution in order to assess the consistency of subgroup profiles (**Figure S3B**).

e) Different model architectures. We generated latent features by (i) removing the dropout layer (**Figure S4A**) and (ii) adding or (iii) subtracting one layer at the encoder and decoder (**Figures S4B, S4C**).

f) Manifold eccentricity. The same analyses were performed using the manifold eccentricity analysis [23,29], which computes the Euclidean distance between the center of the template manifold and all data points (i.e., brain regions) in the manifold space (**Figure S5**).

DATA AVAILABILITY

Imaging and phenotypic data were provided, in part, by the enhanced Nathan Kline Institute-Rockland Sample database, which is available after approval (http://fcon_1000.projects.nitrc.org/indi/enhanced/index.html). Data from Leipzig Study for Mind-Body-Emotion Interactions database are publicly available at (https://ftp.gwdg.de/pub/misc/MPI-Leipzig_Mind-Brain-Body-LEMON/).

CODE AVAILABILITY

Codes for eigenvector generation are available in the BrainSpace toolbox (https://brainspace.readthedocs.io/en/latest/), surface visualization in the BrainNet Viewer toolbox (http://www.nitrc.org/projects/bnv/), and enigma toolbox (https://github.com/MICA-MNI/ENIGMA). Integrative codes for the full analyses are available at https://github.com/gudtls17/EatBehav.RepresentLearning.

FUNDING

This research was supported by the National Research Foundation (NRF-2021R1F1A1052303 and NRF-2020M3E5D2A01084892), Institute for Basic Science (IBS-R015-D1), IITP grant funded by the AI Graduate School Support Program (2019-0-00421), ICT Creative Consilience program (IITP-2020-0-01821), Artificial Intelligence Convergence Research Center, Inha University (2020-0-01389), and Artificial Intelligence Innovation Hub program (2021-0-02068).

AUTHOR CONTRIBUTIONS

H. C., B. P., and H. P. designed the study, analyzed data, and wrote the manuscript. K. B. and J. L. aided in performing the experiments. S. H. reviewed the manuscript. B. P. and H. P. are the corresponding authors of this study and responsible for the integrity of data analysis.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCE

[1] J.J. Reilly, J. Armstrong, A.R. Dorosty, P.M. Emmett, A. Ness, I. Rogers, C. Steer, A. Sherriff, Early life risk factors for obesity in childhood: Cohort study, Br. Med. J. (2005). https://doi.org/10.1136/bmj.38470.670903.E0.

[2] P.K. Newby, Are dietary intakes and eating behaviors related to childhood obesity? A comprehensive review of the evidence, in: J. Law, Med. Ethics, 2007. https://doi.org/10.1111/j.1748-720X.2007.00112.x.

[3] C.F. Moore, V. Sabino, G.F. Koob, P. Cottone, Pathological Overeating: Emerging Evidence for a Compulsivity Construct, Neuropsychopharmacology. (2017). https://doi.org/10.1038/npp.2016.269.

[4] D. Val-Laillet, E. Aarts, B. Weber, M. Ferrari, V. Quaresima, L.E. Stoeckel, M. Alonso-Alonso, M. Audette, C.H. Malbert, E. Stice, Neuroimaging and neuromodulation approaches to study eating behavior and prevent and treat eating disorders and obesity, NeuroImage Clin. (2015). https://doi.org/10.1016/j.nicl.2015.03.016.

[5] P. Monteleone, M. Maj, Dysfunctions of leptin, ghrelin, BDNF and endocannabinoids in eating disorders: Beyond the homeostatic control of food intake, Psychoneuroendocrinology. (2013). https://doi.org/10.1016/j.psyneuen.2012.10.021.

[6] U. Meier, A.M. Gressner, Endocrine regulation of energy metabolism: Review of pathobiochemical and clinical chemical aspects of leptin, ghrelin, adiponectin, and resistin, Clin. Chem. (2004). https://doi.org/10.1373/clinchem.2004.032482.

[7] G. Gerlach, S. Herpertz, S. Loeber, Personality traits and obesity: A systematic review, Obes. Rev. (2015). https://doi.org/10.1111/obr.12235.

[8] H.A. Lee, W.K. Lee, K.A. Kong, N. Chang, E.H. Ha, Y.S. Hong, H. Park, The effect of eating behavior on being overweight or obese during preadolescence, J. Prev. Med. Public Heal. (2011). https://doi.org/10.3961/jpmph.2011.44.5.226.

[9] S.D. Donofry, C.M. Stillman, K.I. Erickson, A review of the relationship between eating behavior, obesity and functional brain network organization, Soc. Cogn. Affect. Neurosci. (2020). https://doi.org/10.1093/scan/nsz085.

[10] B.Y. Park, J. Seo, H. Park, Functional brain networks associated with eating behaviors in obesity, Sci. Rep. (2016). https://doi.org/10.1038/srep23891.

[11] M.A. Hege, K.T. Stingl, S. Kullmann, K. Schag, K.E. Giel, S. Zipfel, H. Preissl, Attentional impulsivity in binge eating disorder modulates response inhibition performance and frontal brain networks, Int. J. Obes. (2015). https://doi.org/10.1038/ijo.2014.99.

[12] U. Vainik, A. Dagher, L. Dubé, L.K. Fellows, Neurobehavioural correlates of body mass index and eating behaviours in adults: A systematic review, Neurosci. Biobehav. Rev. (2013). https://doi.org/10.1016/j.neubiorev.2012.11.008.

[13] L.E. Stoeckel, R.E. Weller, E.W. Cook, D.B. Twieg, R.C. Knowlton, J.E. Cox,

Widespread reward-system activation in obese women in response to pictures of high-calorie foods, Neuroimage. (2008). https://doi.org/10.1016/j.neuroimage.2008.02.031.

[14] A. Dietrich, M. Hollmann, D. Mathar, A. Villringer, A. Horstmann, Brain regulation of food craving: Relationships with weight status and eating behavior, Int. J. Obes. (2016). https://doi.org/10.1038/ijo.2016.28.

[15] L. Maayan, C. Hoogendoorn, V. Sweat, A. Convit, Disinhibited eating in obese adolescents is associated with orbitofrontal volume reductions and executive dysfunction, Obesity. (2011). https://doi.org/10.1038/oby.2011.15.

[16] H. Chen, L.Q. Uddin, X. Guo, J. Wang, R. Wang, X. Wang, X. Duan, H. Chen, Parsing brain structural heterogeneity in males with autism spectrum disorder reveals distinct clinical subtypes, Hum. Brain Mapp. (2019). https://doi.org/10.1002/hbm.24400.

[17] A.K. Easson, Z. Fatima, A.R. McIntosh, Functional connectivity-based subtypes of individuals with and without autism spectrum disorder, Netw. Neurosci. (2019). https://doi.org/10.1162/netn_a_00067.

[18] M. Hrdlicka, I. Dudova, I. Beranova, J. Lisy, T. Belsan, J. Neuwirth, V. Komarek, L. Faladova, M. Havlovicova, Z. Sedlacek, M. Blatny, T. Urbanek, Subtypes of autism by cluster analysis based on structural MRI data, Eur. Child Adolesc. Psychiatry. (2005). https://doi.org/10.1007/s00787-005-0453-z.

[19] G. Goldstein, D.N. Allen, B.E. Seaton, A comparison of clustering solutions for cognitive heterogeneity in schizophrenia, J. Int. Neuropsychol. Soc. (1998). https://doi.org/10.1017/s1355617798003531.

[20] A.E. Farmer, P. McGuffin, E.L. Spitznagel, Heterogeneity in schizophrenia: A cluster-analytic approach, Psychiatry Res. (1983). https://doi.org/10.1016/0165-1781(83)90132-4.

[21] S.J. Hong, S.L. Valk, A. Di Martino, M.P. Milham, B.C. Bernhardt, Multidimensional neuroanatomical subtyping of autism spectrum disorder, Cereb. Cortex. (2018). https://doi.org/10.1093/cercor/bhx229.

[22] D.S. Margulies, S.S. Ghosh, A. Goulas, M. Falkiewicz, J.M. Huntenburg, G. Langs, G. Bezgin, S.B. Eickhoff, F.X. Castellanos, M. Petrides, E. Jefferies, J. Smallwood, Situating the default-mode network along a principal gradient of macroscale cortical organization, Proc. Natl. Acad. Sci. U. S. A. (2016). https://doi.org/10.1073/pnas.1608282113.

[23] R.A.I. Bethlehem, C. Paquola, J. Seidlitz, L. Ronan, B. Bernhardt, C.C.A.N. Consortium, K.A. Tsvetanov, Dispersion of functional gradients across the adult lifespan, Neuroimage. (2020). https://doi.org/10.1016/j.neuroimage.2020.117299.

[24] A.J. Lowe, C. Paquola, R. Vos de Wael, M. Girn, S. Lariviere, S. Tavakol, B. Caldairou, J. Royer, D. V. Schrader, A. Bernasconi, N. Bernasconi, R.N. Spreng, B.C. Bernhardt, Targeting age-related differences in brain and cognition with multimodal imaging and connectome topography profiling, Hum. Brain Mapp. (2019).

https://doi.org/10.1002/hbm.24767.

[25] S.J. Hong, T. Xu, A. Nikolaidis, J. Smallwood, D.S. Margulies, B. Bernhardt, J. Vogelstein, M.P. Milham, Toward a connectivity gradient-based framework for reproducible biomarker discovery, Neuroimage. (2020). https://doi.org/10.1016/j.neuroimage.2020.117322.

[26] S.J. Hong, R.V. de Wael, R.A.I. Bethlehem, S. Lariviere, C. Paquola, S.L. Valk, M.P. Milham, A. Di Martino, D.S. Margulies, J. Smallwood, B.C. Bernhardt, Atypical functional connectome hierarchy in autism, Nat. Commun. (2019). https://doi.org/10.1038/s41467-019-08944-1.

[27] C. Paquola, R.A. Bethlehem, J. Seidlitz, K. Wagstyl, R. Romero-Garcia, K.J. Whitaker, R. Vos De Wael, G.B. Williams, P.E. Vértes, D.S. Margulies, B. Bernhardt, E.T. Bullmore, Shifts in myeloarchitecture characterise adolescent development of cortical gradients, Elife. (2019). https://doi.org/10.7554/eLife.50482.

[28] B. Park, S.-J. Hong, S.L. Valk, C. Paquola, O. Benkarim, R.A.I. Bethlehem, A. Di Martino, M.P. Milham, A. Gozzi, B.T.T. Yeo, J. Smallwood, B.C. Bernhardt, Differences in subcortico-cortical interactions identified from connectome and microcircuit models in autism, Nat. Commun. (2021). https://doi.org/10.1038/s41467-021-21732-0.

[29] B.Y. Park, R.A.I. Bethlehem, C. Paquola, S. Larivière, R. Rodríguez-Cruces, R. Vos de Wael, E.T. Bullmore, B.C. Bernhardt, An expanding manifold in transmodal regions characterizes adolescent reconfiguration of structural connectome organization, Elife. (2021). https://doi.org/10.7554/eLife.64694.

[30] D. Dong, C. Luo, X. Guell, Y. Wang, H. He, M. Duan, S.B. Eickhoff, D. Yao, Compression of cerebellar functional gradients in schizophrenia, Schizophr. Bull. (2020). https://doi.org/10.1093/schbul/sbaa016.

[31] B. yong Park, H. Park, F. Morys, M. Kim, K. Byeon, H. Lee, S.H. Kim, S.L. Valk, A. Dagher, B.C. Bernhardt, Inter-individual body mass variations relate to fractionated functional brain hierarchies, Commun. Biol. (2021). https://doi.org/10.1038/s42003-021-02268-x.

[32] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science (80-.). (2006). https://doi.org/10.1126/science.1127647.

[33] P. Vincent, H. Larochelle, Y. Bengio, P.A. Manzagol, Extracting and composing robust features with denoising autoencoders, in: Proc. 25th Int. Conf. Mach. Learn., 2008. https://doi.org/10.1145/1390156.1390294.

[34] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.A. Manzagol, Stacked denoising autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, J. Mach. Learn. Res. (2010).

[35] H. Il Suk, S.W. Lee, D. Shen, Latent feature representation with stacked auto-encoder for AD/MCI diagnosis, Brain Struct. Funct. (2015). https://doi.org/10.1007/s00429-013-0687-3.

[36] H. Il Suk, C.Y. Wee, S.W. Lee, D. Shen, State-space model with deep learning for functional dynamics estimation in resting-state fMRI, Neuroimage. (2016). https://doi.org/10.1016/j.neuroimage.2016.01.005.

[37] J. Kim, V.D. Calhoun, E. Shim, J.H. Lee, Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia, Neuroimage. (2016). https://doi.org/10.1016/j.neuroimage.2015.05.018.

[38] L.L. Zeng, H. Wang, P. Hu, B. Yang, W. Pu, H. Shen, X. Chen, Z. Liu, H. Yin, Q. Tan, K. Wang, D. Hu, Multi-Site Diagnostic Classification of Schizophrenia Using Discriminant Deep Learning with Functional Connectivity MRI, EBioMedicine. (2018). https://doi.org/10.1016/j.ebiom.2018.03.017.

[39] A.S. Heinsfeld, A.R. Franco, R.C. Craddock, A. Buchweitz, F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the ABIDE dataset, NeuroImage Clin. (2018). https://doi.org/10.1016/j.nicl.2017.08.017.

[40] K.B. Nooner, S.J. Colcombe, R.H. Tobe, M. Mennes, M.M. Benedict, A.L. Moreno, L.J. Panek, S. Brown, S.T. Zavitz, Q. Li, S. Sikka, D. Gutman, S. Bangaru, R.T. Schlachter, S.M. Kamiel, A.R. Anwar, C.M. Hinz, M.S. Kaplan, A.B. Rachlin, S. Adelsberg, B. Cheung, R. Khanuja, C. Yan, C.C. Craddock, V. Calhoun, W. Courtney, M. King, D. Wood, C.L. Cox, A.M.C. Kelly, A. Di Martino, E. Petkova, P.T. Reiss, N. Duan, D. Thomsen, B. Biswal, B. Coffey, M.J. Hoptman, D.C. Javitt, N. Pomara, J.J. Sidtis, H.S. Koplewicz, F.X. Castellanos, B.L. Leventhal, M.P. Milham, The NKI-Rockland sample: A model for accelerating the pace of discovery science in psychiatry, Front. Neurosci. (2012). https://doi.org/10.3389/fnins.2012.00152.

[41] A. Babayan, M. Erbey, D. Kumral, J.D. Reinelt, A.M.F. Reiter, J. Röbbig, H. Lina Schaare, M. Uhlig, A. Anwander, P.L. Bazin, A. Horstmann, L. Lampe, V. V. Nikulin, H. Okon-Singer, S. Preusser, A. Pampel, C.S. Rohr, J. Sacher, A. Thöne-Otto, S. Trapp, T. Nierhaus, D. Altmann, K. Arelin, M. Blöchl, E. Bongartz, P. Breig, E. Cesnaite, S. Chen, R. Cozatl, S. Czerwonatis, G. Dambrauskaite, M. Dreyer, J. Enders, M. Engelhardt, M.M. Fischer, N. Forschack, J. Golchert, L. Golz, C.A. Guran, S. Hedrich, N. Hentschel, D.I. Hoffmann, J.M. Huntenburg, R. Jost, A. Kosatschek, S. Kunzendorf, H. Lammers, M.E. Lauckner, K. Mahjoory, A.S. Kanaan, N. Mendes, R. Menger, E. Morino, K. Näthe, J. Neubauer, H. Noyan, S. Oligschläger, P. Panczyszyn-Trzewik, D. Poehlchen, N. Putzke, S. Roski, M.C. Schaller, A. Schieferbein, B. Schlaak, R. Schmidt, K.J. Gorgolewski, H.M. Schmidt, A. Schrimpf, S. Stasch, M. Voss, A. Wiedemann, D.S. Margulies, M. Gaebler, A. Villringer, Data descriptor: A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults, Sci. Data. (2019). https://doi.org/10.1038/sdata.2018.308.

[42] L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A.R. Laird, P.T. Fox, S.B. Eickhoff, C. Yu, T. Jiang, The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture, Cereb. Cortex. (2016). https://doi.org/10.1093/cercor/bhw157.

[43] R. Vos de Wael, O. Benkarim, C. Paquola, S. Lariviere, J. Royer, S. Tavakol, T. Xu,

S.J. Hong, G. Langs, S. Valk, B. Misic, M. Milham, D. Margulies, J. Smallwood, B.C. Bernhardt, BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets, Commun. Biol. (2020). https://doi.org/10.1038/s42003-020-0794-7.

[44] G. Langs, P. Golland, S.S. Ghosh, Predicting activation across individuals with resting-state functional connectivity based multi-atlas label fusion, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2015. https://doi.org/10.1007/978-3-319-24571-3_38.

[45] S. Monti, P. Tamayo, J. Mesirov, T. Golub, Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data, Mach. Learn. (2003). https://doi.org/10.1023/A:1023949509487.

[46] A.J. Stunkard, S. Messick, The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger, J. Psychosom. Res. (1985). https://doi.org/10.1016/0022-3999(85)90010-8.

[47] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribution for deep networks, in: 34th Int. Conf. Mach. Learn. ICML 2017, 2017.

[48] L. Sthle, S. Wold, Multivariate analysis of variance (MANOVA), Chemom. Intell. Lab. Syst. (1990). https://doi.org/10.1016/0169-7439(90)80094-M.

[49] B.T. Thomas Yeo, F.M. Krienen, J. Sepulcre, M.R. Sabuncu, D. Lashkari, M. Hollinshead, J.L. Roffman, J.W. Smoller, L. Zöllei, J.R. Polimeni, B. Fisch, H. Liu, R.L. Buckner, The organization of the human cerebral cortex estimated by intrinsic functional connectivity, J. Neurophysiol. (2011). https://doi.org/10.1152/jn.00338.2011.

[50] T. Yarkoni, R.A. Poldrack, T.E. Nichols, D.C. Van Essen, T.D. Wager, Large-scale automated synthesis of human functional neuroimaging data, Nat. Methods. (2011). https://doi.org/10.1038/nmeth.1635.

[51] T.N. Rubin, O. Koyejo, K.J. Gorgolewski, M.N. Jones, R.A. Poldrack, T. Yarkoni, Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition, PLoS Comput. Biol. (2017). https://doi.org/10.1371/journal.pcbi.1005649.

[52] R. Vos de Wael, O. Benkarim, C. Paquola, S. Lariviere, J. Royer, S. Tavakol, T. Xu, S.J. Hong, G. Langs, S. Valk, B. Misic, M. Milham, D. Margulies, J. Smallwood, B.C. Bernhardt, BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets, Commun. Biol. (2020). https://doi.org/10.1038/s42003-020-0794-7.

[53] B.Y. Park, M.J. Lee, M. Kim, S.H. Kim, H. Park, Structural and Functional Brain Connectivity Changes Between People With Abdominal and Non-abdominal Obesity and Their Association With Behaviors of Eating Disorders, Front. Neurosci. (2018). https://doi.org/10.3389/fnins.2018.00741.

[54] S.H. Kim, B.Y. Park, K. Byeon, H. Park, Y. Kim, Y.M. Eun, J.H. Chung, The effects

of high-frequency repetitive transcranial magnetic stimulation on resting-state functional connectivity in obese adults, Diabetes, Obes. Metab. (2019). https://doi.org/10.1111/dom.13763.

[55] D. Castelvecchi, Can we open the black box of AI?, Nature. (2016). https://doi.org/10.1038/538020a.

[56] G. Montavon, W. Samek, K.R. Müller, Methods for interpreting and understanding deep neural networks, Digit. Signal Process. A Rev. J. (2018). https://doi.org/10.1016/j.dsp.2017.10.011.

[57] D.T. Huff, A.J. Weisman, R. Jeraj, Interpretation and visualization techniques for deep learning models in medical imaging, Phys. Med. Biol. (2021). https://doi.org/10.1088/1361-6560/abcd17.

[58] M. Hengstler, E. Enkel, S. Duelli, Applied artificial intelligence and trust-The case of autonomous vehicles and medical assistance devices, Technol. Forecast. Soc. Change. (2016). https://doi.org/10.1016/j.techfore.2015.12.014.

[59] S. Nundy, T. Montgomery, R.M. Wachter, Promoting trust between patients and physicians in the era of artificial intelligence, JAMA - J. Am. Med. Assoc. (2019). https://doi.org/10.1001/jama.2018.20563.

[60] J.C.Y. Seah, J.S.N. Tang, A. Kitchen, F. Gaillard, A.F. Dixon, Chest radiographs in congestive heart failure: Visualizing neural network learning, Radiology. (2019). https://doi.org/10.1148/radiol.2018180887.

[61] X. Feng, J. Yang, A.F. Laine, E.D. Angelini, Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2017. https://doi.org/10.1007/978-3-319-66179-7_65.

[62] S. Hwang, H.E. Kim, Self-transfer learning for weakly supervised lesion localization, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2016. https://doi.org/10.1007/978-3-319-46723-8_28.

[63] F. Liu, B. Guan, Z. Zhou, A. Samsonov, H. Rosas, K. Lian, R. Sharma, A. Kanarek, J. Kim, A. Guermazi, R. Kijowski, Fully Automated Diagnosis of Anterior Cruciate Ligament Tears on Knee MR Images by Using Deep Learning, Radiol. Artif. Intell. (2019). https://doi.org/10.1148/ryai.2019180091.

[64] Y. Shen, M. Gao, Dynamic routing on deep neural network for thoracic disease classification and sensitive area localization, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2018. https://doi.org/10.1007/978-3-030-00919-9_45.

[65] F. Eitel, E. Soehler, J. Bellmann-Strobl, A.U. Brandt, K. Ruprecht, R.M. Giess, J. Kuchling, S. Asseyer, M. Weygandt, J.D. Haynes, M. Scheel, F. Paul, K. Ritter, Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation, NeuroImage Clin. (2019). https://doi.org/10.1016/j.nicl.2019.102003.

[66] A.W. Thomas, H.R. Heekeren, K.R. Müller, W. Samek, Analyzing Neuroimaging Data Through Recurrent Deep Learning Models, Front. Neurosci. (2019). https://doi.org/10.3389/fnins.2019.01321.

[67] M. Böhle, F. Eitel, M. Weygandt, K. Ritter, Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification, Front. Aging Neurosci. (2019). https://doi.org/10.3389/fnagi.2019.00194.

[68] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, PLoS One. (2015). https://doi.org/10.1371/journal.pone.0130140.

[69] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning Deep Features for Discriminative Localization, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2016. https://doi.org/10.1109/CVPR.2016.319.

[70] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, Int. J. Comput. Vis. (2020). https://doi.org/10.1007/s11263-019-01228-7.

[71] J. Chen, E.K. Papies, L.W. Barsalou, A core eating network and its modulations underlie diverse eating phenomena, Brain Cogn. (2016). https://doi.org/10.1016/j.bandc.2016.04.004.

[72] D.C. Castro, S.L. Cole, K.C. Berridge, Lateral hypothalamus, nucleus accumbens, and ventral pallidum roles in eating and hunger: Interactions between homeostatic and reward circuitry, Front. Syst. Neurosci. (2015). https://doi.org/10.3389/fnsys.2015.00090.

[73] N.M. White, A.E. Fisher, Relationship between amygdala and hypothalamus in the control of eating behavior, Physiol. Behav. (1969). https://doi.org/10.1016/0031-9384(69)90081-x.

[74] A.M. Douglass, H. Kucukdereli, M. Ponserre, M. Markovic, J. Gründemann, C. Strobel, P.L. Alcala Morales, K.K. Conzelmann, A. Lüthi, R. Klein, Central amygdala circuits modulate food consumption through a positive-valence mechanism, Nat. Neurosci. (2017). https://doi.org/10.1038/nn.4623.

[75] P.A. Tataranni, J.F. Gautier, K. Chen, A. Uecker, D. Bandy, A.D. Salbe, R.E. Pratley, M. Lawson, E.M. Reiman, E. Ravussin, Neuroanatomical correlates of hunger and satiation in humans using positron emission tomography, Proc. Natl. Acad. Sci. U. S. A. (1999). https://doi.org/10.1073/pnas.96.8.4569.

[76] A.M. van Opstal, M.A. Wijngaarden, J. van der Grond, H. Pijl, Changes in brain activity after weight loss, Obes. Sci. Pract. (2019). https://doi.org/10.1002/osp4.363.

[77] J. Verdejo-Román, R. Vilar-López, J.F. Navas, C. Soriano-Mas, A. Verdejo-García, Brain reward system's alterations in response to food and monetary stimuli in overweight and obese individuals, Hum. Brain Mapp. (2017). https://doi.org/10.1002/hbm.23407.

[78] N.D. Volkow, G.J. Wang, F. Telang, J.S. Fowler, P.K. Thanos, J. Logan, D. Alexoff, Y.S. Ding, C. Wong, Y. Ma, K. Pradhan, Low dopamine striatal D2 receptors are associated with prefrontal metabolism in obese subjects: Possible contributing factors, Neuroimage. (2008). https://doi.org/10.1016/j.neuroimage.2008.06.002.

[79] H. Ziauddeen, M. Alonso-Alonso, J.O. Hill, M. Kelley, N.A. Khan, Obesity and the neurocognitive basis of food reward and the control of intake, Adv. Nutr. (2015). https://doi.org/10.3945/an.115.008268.

[80] P.A. Tataranni, A. DelParigi, Functional neuroimaging: A new generation of human brain studies in obesity research, Obes. Rev. (2003). https://doi.org/10.1046/j.1467-789X.2003.00111.x.

[81] S.J. Brooks, J. Cedernaes, H.B. Schiöth, Increased Prefrontal and Parahippocampal Activation with Reduced Dorsolateral Prefrontal and Insular Cortex Activation to Food Images in Obesity: A Meta-Analysis of fMRI Studies, PLoS One. (2013). https://doi.org/10.1371/journal.pone.0060393.

[82] Y. Ding, G. Ji, G. Li, W. Zhang, Y. Hu, L. Liu, Y. Wang, C. Hu, K.M. von Deneen, Y. Han, G. Cui, H. Wang, C.E. Wiers, P. Manza, D. Tomasi, N.D. Volkow, Y. Nie, G.J. Wang, Y. Zhang, Altered Interactions Among Resting-State Networks in Individuals with Obesity, Obesity. (2020). https://doi.org/10.1002/oby.22731.

[83] G. Olivo, L. Wiemerslage, I. Swenne, C. Zhukowsky, H. Salonen-Ros, E.M. Larsson, S. Gaudio, S.J. Brooks, H.B. Schiöth, Limbic-thalamo-cortical projections and reward-related circuitry integrity affects eating behavior: A longitudinal DTI study in adolescents with restrictive eating disorders, PLoS One. (2017). https://doi.org/10.1371/journal.pone.0172129.

[84] T. Steward, A. Juaneda-Seguí, G. Mestre-Bach, I. Martínez-Zalacaín, N. Vilarrasa, S. Jiménez-Murcia, J.A. Fernández-Formoso, M.V. de las Heras, N. Custal, N. Virgili, R. Lopez-Urdiales, A. García-Ruiz-de-Gordejuela, J.M. Menchón, C. Soriano-Mas, F. Fernandez-Aranda, What difference does it make? Risk-taking behavior in obesity after a loss is associated with decreased ventromedial prefrontal cortex activity, J. Clin. Med. (2019). https://doi.org/10.3390/jcm8101551.

[85] T. Steward, R. Miranda-Olivos, C. Soriano-Mas, F. Fernández-Aranda, Neuroendocrinological mechanisms underlying impulsive and compulsive behaviors in obesity: a narrative review of fMRI studies, Rev. Endocr. Metab. Disord. (2019). https://doi.org/10.1007/s11154-019-09515-x.

[86] F. van Meer, L.N. van der Laan, G. Eiben, L. Lissner, M. Wolters, S. Rach, M. Herrmann, P. Erhard, D. Molnar, G. Orsi, M.A. Viergever, R.A.H. Adan, P.A.M. Smeets, Development and body mass inversely affect children's brain activation in dorsolateral prefrontal cortex during food choice, Neuroimage. (2019). https://doi.org/10.1016/j.neuroimage.2019.116016.

[87] A.M. Van Opstal, A.A. Van Den Berg-Huysmans, M. Hoeksma, C. Blonk, H. Pijl, S.A.R.B. Rombouts, J. Van Der Grond, The effect of consumption temperature on the homeostatic and hedonic responses to glucose ingestion in the hypothalamus and the reward system, Am. J. Clin. Nutr. (2018). https://doi.org/10.1093/ajcn/nqx023.

[88] E. Stice, S. Spoor, C. Bohon, D.M. Small, Relation between obesity and blunted striatal response to food is moderated by TaqIA A1 allele, Science (80-.). (2008). https://doi.org/10.1126/science.1161550.

[89] J.A. Felsted, X. Ren, F. Chouinard-Decorte, D.M. Small, Genetically determined differences in brain response to a primary food reward, J. Neurosci. (2010). https://doi.org/10.1523/JNEUROSCI.5483-09.2010.

[90] A.C. Choquette, S. Lemieux, A. Tremblay, V. Drapeau, C. Bouchard, M.C. Vohl, L. Pérusse, GAD2 gene sequence variations are associated with eating behaviors and weight gain in women from the Quebec family study, Physiol. Behav. (2009). https://doi.org/10.1016/j.physbeh.2009.08.004.

[91] K. Timper, J.C. Brüning, Hypothalamic circuits regulating appetite and energy homeostasis: Pathways to obesity, DMM Dis. Model. Mech. (2017). https://doi.org/10.1242/dmm.026609.

[92] L. Vong, C. Ye, Z. Yang, B. Choi, S. Chua, B.B. Lowell, Leptin Action on GABAergic Neurons Prevents Obesity and Reduces Inhibitory Tone to POMC Neurons, Neuron. (2011). https://doi.org/10.1016/j.neuron.2011.05.028.

[93] M. Durst, K. Könczöl, T. Balázsa, M.D. Eyre, Z.E. Tóth, Reward-representing D1-type neurons in the medial shell of the accumbens nucleus regulate palatable food intake, Int. J. Obes. (2019). https://doi.org/10.1038/s41366-018-0133-y.

[94] B.A. Matikainen-Ankney, A. V. Kravitz, Persistent effects of obesity: A neuroplasticity hypothesis, Ann. N. Y. Acad. Sci. (2018). https://doi.org/10.1111/nyas.13665.

[95] S.F. Leibowitz, J.T. Alexander, Hypothalamic serotonin in control of eating behavior, meal size, and body weight, Biol. Psychiatry. (1998). https://doi.org/10.1016/S0006-3223(98)00186-3.

[96] S.F. Leibowitz, G. Shor-Posner, Brain serotonin and eating behavior, Appetite. (1986). https://doi.org/10.1016/S0195-6663(86)80049-6.

[97] B.Y. Park, K. Byeon, H. Park, FuNP (fusion of neuroimaging preprocessing) pipelines: A fully automated preprocessing software for functional magnetic resonance imaging, Front. Neuroinform. (2019). https://doi.org/10.3389/fninf.2019.00005.

[98] R.W. Cox, AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages, Comput. Biomed. Res. (1996). https://doi.org/10.1006/cbmr.1996.0014.

[99] M. Jenkinson, C.F. Beckmann, T.E.J. Behrens, M.W. Woolrich, S.M. Smith, FSL - Review, Neuroimage. (2012). https://doi.org/10.1016/j.neuroimage.2011.09.015.

[100] B.B. Avants, N.J. Tustison, G. Song, P.A. Cook, A. Klein, J.C. Gee, A reproducible evaluation of ANTs similarity metric performance in brain image registration, Neuroimage. (2011). https://doi.org/10.1016/j.neuroimage.2010.09.025.

[101] G. Salimi-Khorshidi, G. Douaud, C.F. Beckmann, M.F. Glasser, L. Griffanti, S.M.

Smith, Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers, Neuroimage. (2014). https://doi.org/10.1016/j.neuroimage.2013.11.046.

[102] N. Mendes, S. Oligschläger, M.E. Lauckner, J. Golchert, J.M. Huntenburg, M. Falkiewicz, M. Ellamil, S. Krause, B.M. Baczkowski, R. Cozatl, A. Osoianu, D. Kumral, J. Pool, L. Golz, M. Dreyer, P. Haueis, R. Jost, Y. Kramarenko, H. Engen, K. Ohrnberger, K.J. Gorgolewski, N. Farrugia, A. Babayan, A. Reiter, H.L. Schaare, J. Reinelt, J. Röbbig, M. Uhlig, M. Erbey, M. Gaebler, J. Smallwood, A. Villringer, D.S. Margulies, A functional connectome phenotyping dataset including cognitive state and personality measures, Sci. Data. (2017). https://doi.org/10.1101/164764.

[103] P.L. Bazin, M. Weiss, J. Dinse, A. Schäfer, R. Trampel, R. Turner, A computational framework for ultra-high resolution cortical segmentation at 7 Tesla, Neuroimage. 93 (2014) 201–209. https://doi.org/10.1016/j.neuroimage.2013.03.077.

[104] A.M. Dale, B. Fischl, M.I. Sereno, Cortical surface-based analysis: I. Segmentation and surface reconstruction, Neuroimage. (1999). https://doi.org/10.1006/nimg.1998.0395.

[105] B. Fischl, M.I. Sereno, A.M. Dale, Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system, Neuroimage. (1999). https://doi.org/10.1006/nimg.1998.0396.

[106] M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images, Neuroimage. (2002). https://doi.org/10.1016/S1053-8119(02)91132-8.

[107] D.N. Greve, B. Fischl, Accurate and robust brain image alignment using boundary-based registration, Neuroimage. (2009). https://doi.org/10.1016/j.neuroimage.2009.06.060.

[108] Y. Behzadi, K. Restom, J. Liau, T.T. Liu, A component based noise correction method (CompCor) for BOLD and perfusion based fMRI, Neuroimage. (2007). https://doi.org/10.1016/j.neuroimage.2007.04.042.

[109] A. Rokem, M. Trumpis, F. Pérez, Nitime : time-series analysis for neuroimaging data, in: Proc. 8th Python Sci. Conf. (SciPy 2009), 2009.

[110] W.H. Thompson, P. Fransson, On Stabilizing the Variance of Dynamic Functional Brain Connectivity Time Series, Brain Connect. (2016). https://doi.org/10.1089/brain.2016.0454.

[111] R.R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. (2006). https://doi.org/10.1016/j.acha.2006.04.006.

[112] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. (2014).

[113] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 2015.

[114] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, J. R. Stat. Soc. Ser. B. (1995). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[115] R.L. Moseley, R.J.F. Ypma, R.J. Holt, D. Floris, L.R. Chura, M.D. Spencer, S. Baron-Cohen, J. Suckling, E. Bullmore, M. Rubinov, Whole-brain functional hypoconnectivity as an endophenotype of autism in adolescents, NeuroImage Clin. (2015). https://doi.org/10.1016/j.nicl.2015.07.015.

[116] A. Fallahi, M. Pooyan, N. Lotfi, F. Baniasad, L. Tapak, N. Mohammadi-Mobarakeh, S.S. Hashemi-Fesharaki, J. Mehvari-Habibabadi, M.R. Ay, M.R. Nazem-Zadeh, Dynamic functional connectivity in temporal lobe epilepsy: a graph theoretical and machine learning approach, Neurol. Sci. (2021). https://doi.org/10.1007/s10072-020-04759-x.

[117] A. Tusche, J. Smallwood, B.C. Bernhardt, T. Singer, Classifying the wandering mind: Revealing the affective content of thoughts during task-free rest periods, Neuroimage. (2014). https://doi.org/10.1016/j.neuroimage.2014.03.076.

[118] M. Xia, J. Wang, Y. He, BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics, PLoS One. (2013). https://doi.org/10.1371/journal.pone.0068910.

[119] S. Larivière, C. Paquola, B. yong Park, J. Royer, Y. Wang, O. Benkarim, R. Vos de Wael, S.L. Valk, S.I. Thomopoulos, M. Kirschner, L.B. Lewis, A.C. Evans, S.M. Sisodiya, C.R. McDonald, P.M. Thompson, B.C. Bernhardt, The ENIGMA Toolbox: multiscale neural contextualization of multisite neuroimaging datasets, Nat. Methods. (2021). https://doi.org/10.1038/s41592-021-01186-4.

Supplementary Information

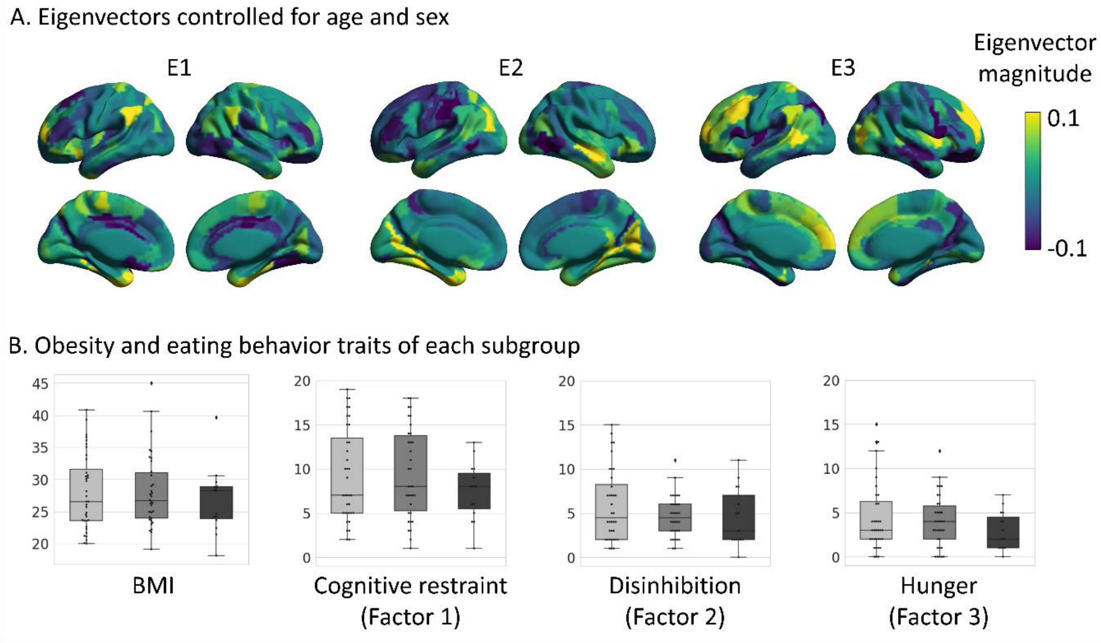


Figure S1 | Subgroup identification without the autoencoder model. (A) Three eigenvectors (E1, E2, E3) controlled for age and sex are shown on brain surfaces. **(B)** Distribution of BMI and eating behavior scores of each subgroup are plotted. *Abbreviation:* BMI, body mass index.

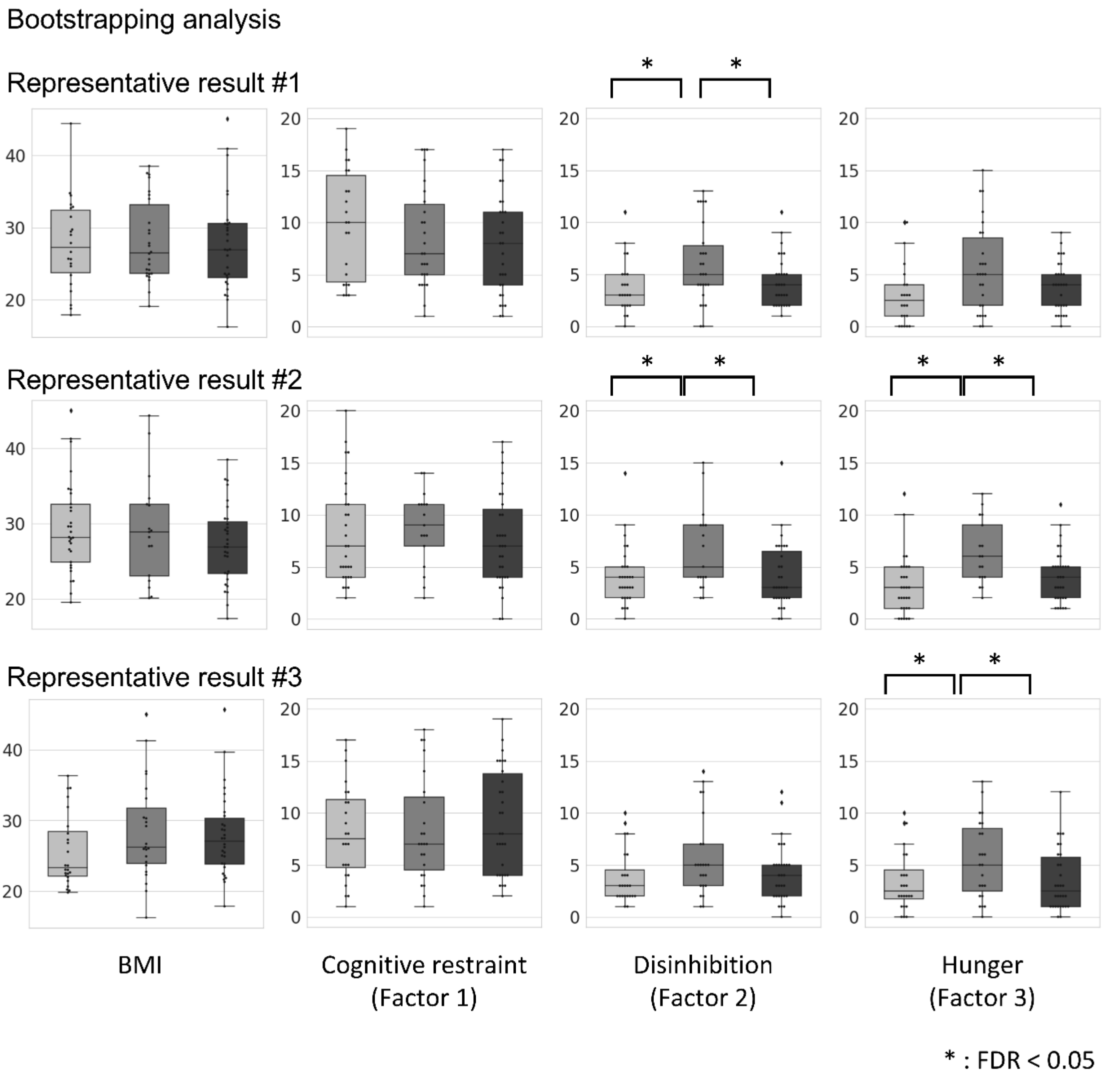


Figure S2 | Bootstrapping analysis. We performed bootstrapping analysis by selecting 90% of participants with replacement and reported the obesity and eating behavior scores. Three representative results are presented. *Abbreviations:* BMI, body mass index; FDR, false discovery rate.

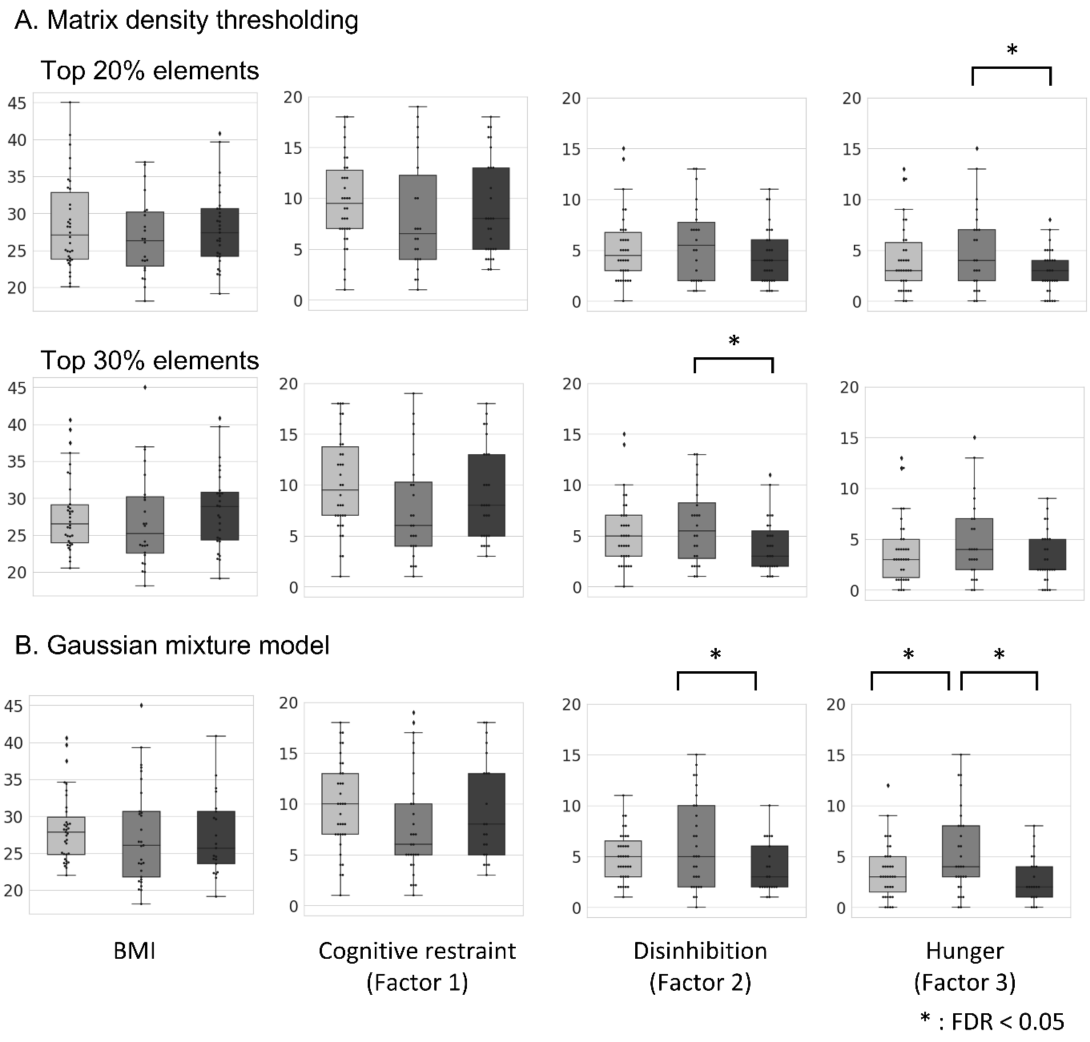


Figure S3 | Different functional connectivity matrix densities and clustering method. We plotted BMI and eating behavior scores by changing the **(A)** matrix density with 20% (top) and 30% (bottom) and **(B)** clustering method to the Gaussian mixture model. *Abbreviations:* BMI, body mass index; FDR, false discovery rate.

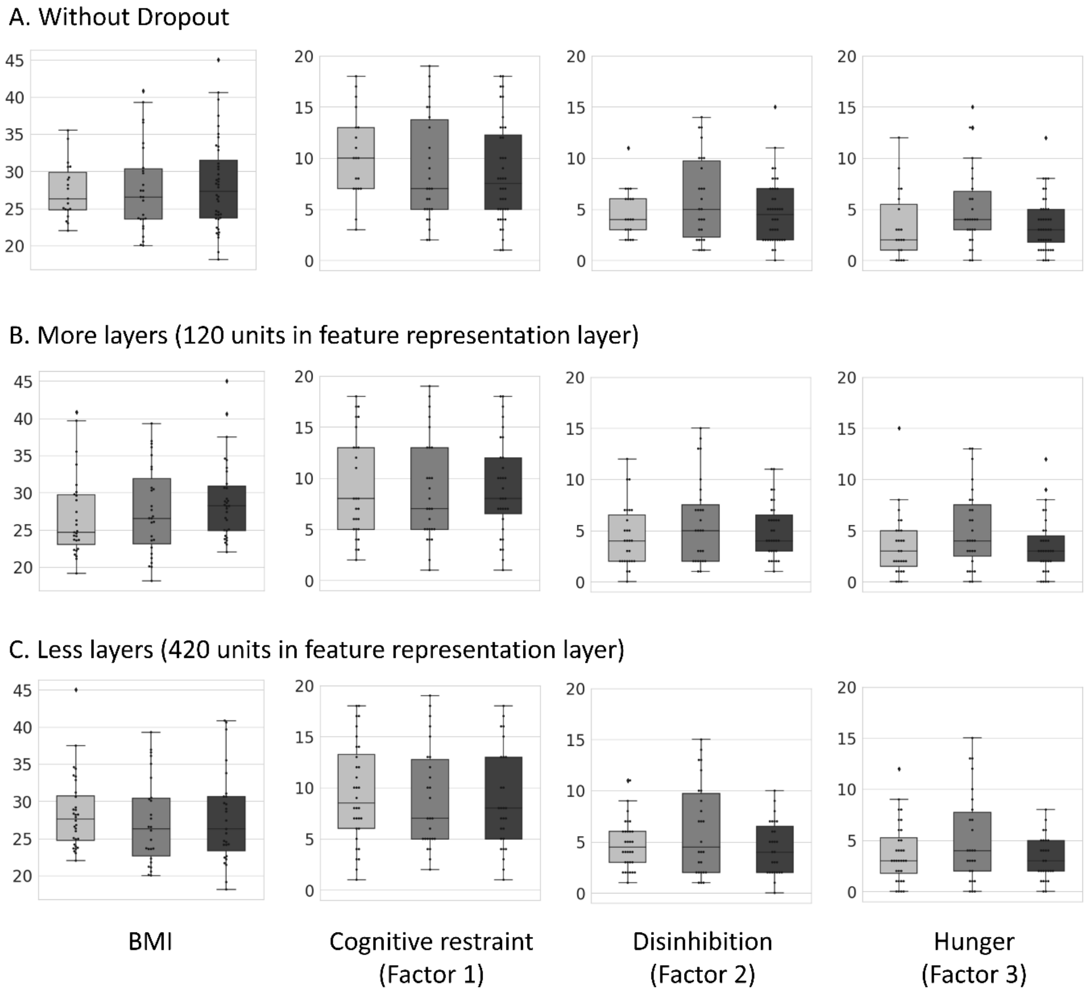


Figure S4 | Different autoencoder models. **(A)** Distribution of the BMI and eating behavior scores without dropout layers. **(B)** Score distribution with one extra encoder and one extra decoder layers. The feature representation layer had 120 latent variables. **(C)** Score distribution with one less encoder and one less decoder layers. The feature representation layer had 420 latent variables. *Abbreviation:* BMI, body mass index.

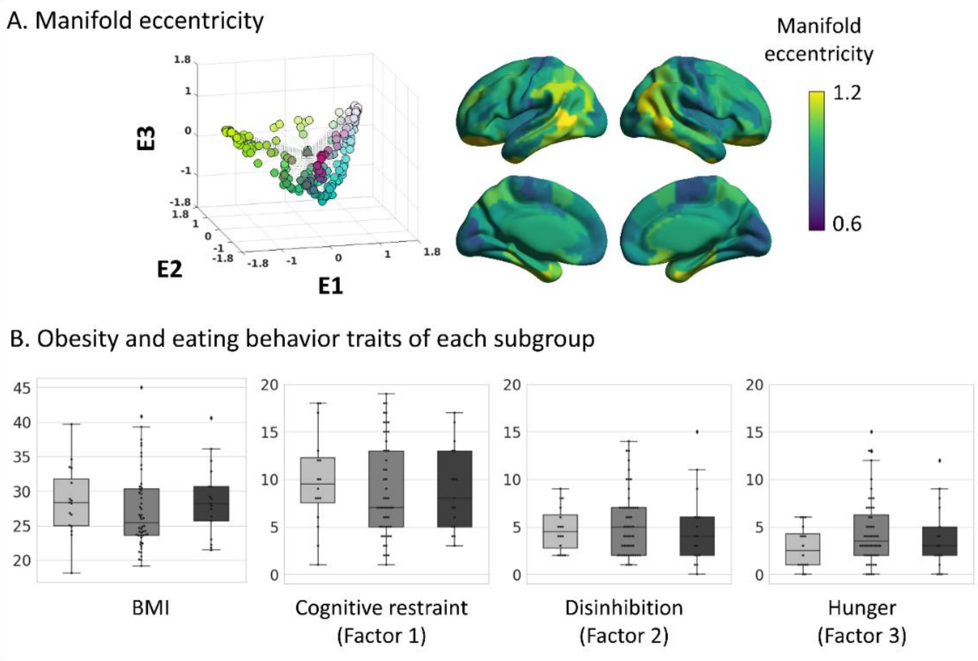


Figure S5 | Subgroup analysis using the manifold eccentricity. (A) Dots in the scatter plot represents each brain region projected onto the three-dimensional manifold space, and colors are mapped onto the brain surface for visualization. **(B)** Distribution of the BMI and eating behavior scores are plotted. *Abbreviation:* BMI, body mass index.