




## Meta-connectomic analysis maps consistent, reproducible, and transcriptionally relevant functional connectome hubs in the human brain

Zhilei Xu <sup>1,2,3</sup>, Mingrui Xia <sup>1,2,3</sup>, Xindi Wang<sup>1,2,3</sup>, Xuhong Liao<sup>4</sup>, Tengda Zhao<sup>1,2,3</sup> & Yong He <sup>1,2,3,5</sup>✉

Human brain connectomes include sets of densely connected hub regions. However, the consistency and reproducibility of functional connectome hubs have not been established to date and the genetic signatures underlying robust hubs remain unknown. Here, we conduct a worldwide harmonized meta-connectomic analysis by pooling resting-state functional MRI data of 5212 healthy young adults across 61 independent cohorts. We identify highly consistent and reproducible connectome hubs in heteromodal and unimodal regions both across cohorts and across individuals, with the greatest effects observed in lateral parietal cortex. These hubs show heterogeneous connectivity profiles and are critical for both intra- and inter-network communications. Using post-mortem transcriptome datasets, we show that as compared to non-hubs, connectome hubs have a spatiotemporally distinctive transcriptomic pattern dominated by genes involved in the neuropeptide signaling pathway, neurodevelopmental processes, and metabolic processes. These results highlight the robustness of macroscopic connectome hubs and their potential cellular and molecular underpinnings, which markedly furthers our understanding of how connectome hubs emerge in development, support complex cognition in health, and are involved in disease.

<sup>1</sup>State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China. <sup>2</sup>Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China. <sup>3</sup>IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China. <sup>4</sup>School of Systems Science, Beijing Normal University, Beijing, China. <sup>5</sup>Chinese Institute for Brain Research, Beijing, China. ✉email: [yong.he@bnu.edu.cn](mailto:yong.he@bnu.edu.cn)

Functional connectome mapping studies have identified sets of densely connected regions in large-scale human brain networks, which are known as hubs<sup>1</sup>. Connectome hubs play a crucial role in global brain communication<sup>1,2</sup> and support a broad range of cognitive processing, such as working memory<sup>3</sup> and semantic processing<sup>4</sup>. Growing evidence suggests that these highly connected brain hubs are preferentially targeted by many neuropsychiatric disorders<sup>5–8</sup>, which provides critical clues for understanding the biological mechanisms of disorders and establishing biomarkers for disease diagnosis<sup>8,9</sup> and treatment evaluation<sup>10</sup> (refs. 1,2,11,12 for reviews).

Despite such importance, there is considerable inconsistency in anatomical locations of functional connectome hubs among existing studies. For example, components of the default-mode network (DMN) have been frequently reported as connectome hubs, yet the spatial pattern is highly variable across studies. In particular, several studies have shown highly connected hubs in the lateral parietal regions of the DMN<sup>7,8,13,14</sup>, whereas others have reported midline structures of the DMN<sup>15–19</sup>. Several works have identified primary sensorimotor and visual regions as connectome hubs<sup>13,14,16–19</sup>, yet others did not replicate these findings<sup>7,8,15</sup>. Subcortical regions, such as the thalamus and amygdala, have also been inconsistently reported as hubs<sup>8,15,16,18</sup> and non-hubs<sup>7,13,14,17,19</sup>. Thus, the consistency and reproducibility of functional connectome hubs have been difficult to establish to date, which can be attributed to inadequate sample size and differences in imaging scanner, imaging protocol, data processing, and connectome analysis strategies. Here, we aimed to establish a harmonized meta-analysis model to identify robust functional connectome hubs in healthy young adults by combining multiple cohorts with uniform protocols for data quality assurance, image processing, and connectome analyses.

Once the robust connectome hubs are identified, we will further examine their genetic signatures. It has been well demonstrated that the connectome architecture of the human brain is inheritable, such as functional connectivity of the DMN<sup>20</sup> and the cost-efficiency optimization<sup>21</sup>. Moreover, the functional connectomes can be regulated by genotypic variation both during rest<sup>22</sup> and in cognitive tasks<sup>23</sup>, especially involving the DMN<sup>22,23</sup> and frontoparietal network (FPN)<sup>23</sup>. Growing evidence also suggests spatial correspondence between transcriptomic profiles and connectome architectures<sup>24–26</sup> (ref. 27 for review). Thus, we reasoned that the robust macroscopic connectome hubs could be associated with microscopic genetic signatures. Elucidating these genetic signatures will substantially benefit our understanding of how connectome hubs emerge in development, function in complex cognition, and are involved in disease.

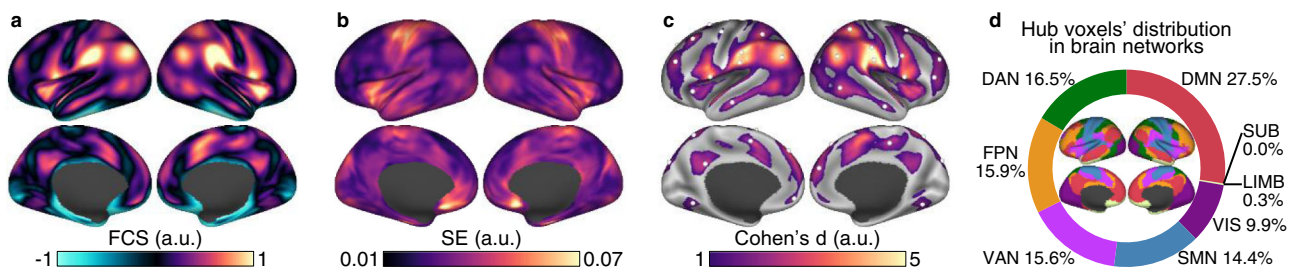
To address these issues, we provided, to the best of our knowledge, the first worldwide harmonized meta-connectomic analysis of functional brain hubs by pooling a large-sample

resting-state functional MRI (rsfMRI) dataset of 5212 healthy young adults (aged 18–36 years, 2377 males) across 61 independent cohorts. We identified highly consistent and reproducible functional connectome hubs in multiple heteromodal and unimodal regions, with the most robust findings occurring in several lateral parietal regions. These connectome hubs showed unique and heterogeneous connectivity profiles to provide support for both intra- and inter-network communications. To uncover the genetic signatures underlying these connectome hubs, we conducted machine learning approaches to distinguish connectome hubs from non-hubs using transcriptomic data from the Allen Human Brain Atlas (AHBA), explored their developmental evolutions using the BrainSpan Atlas, and assessed their neural relevance by contextualizing them relative to established neuroimaging patterns. We demonstrated that these robust connectome hubs were associated with a spatiotemporal transcriptomic pattern dominated by genes enriched for the neuropeptide signaling pathway, neurodevelopmental processes, and metabolic processes.

## Results

**Identifying consistent connectome hubs using a harmonized meta-analysis model.** Prior to the meta-analysis, we constructed a voxelwise functional connectome matrix for each individual by computing the Pearson's correlation coefficient between pre-processed rsfMRI time series of all pairs of gray matter voxels (47,619 voxels). Then, the functional connectivity strength (FCS) of each voxel was computed as the sum of connection weights between the given voxel and all the other voxels. This resultant FCS map was further normalized with respect to its mean and standard deviation across voxels<sup>7</sup>. For each cohort, we performed a general linear model on these normalized FCS maps to reduce age and sex effects. As a result, we obtained a mean FCS map and its corresponding variance map for each cohort that were used for subsequent meta-analyses.

To identify the most consistent connectome hubs, we conducted a voxelwise random-effects meta-analysis on the mean and variance FCS maps of the 61 cohorts. Such an analysis addressed the across-cohort heterogeneity of functional connectomes, resulting in a robust FCS pattern (Fig. 1a) and its corresponding standard error (SE) map (Fig. 1b). Then, we identified consistent connectome hubs whose FCS values were significantly ( $p < 0.001$ , cluster size  $> 200 \text{ mm}^3$ ) higher than the global mean (i.e., zero) using a voxelwise  $Z$  value map computed by dividing the FCS map by the SE map. To determine the statistical significances of these observed  $Z$  values, a nonparametric permutation test<sup>28</sup> with 10,000 iterations was performed. Finally, we estimated voxelwise effect sizes using Cohen's  $d$  metric computed by dividing the  $Z$  value map by the square root of the



**Fig. 1** Identifying consistent connectome hubs using a harmonized meta-analysis model. **a, b** Robust FCS pattern (**a**) and its corresponding variance (standard error, SE) map (**b**) estimated using a harmonized voxelwise random-effects meta-analysis across 61 cohorts. **c** The most consistent functional connectome hubs ( $p < 0.001$ , cluster size  $> 200 \text{ mm}^3$ ). White spheres represent hub peaks. **a–c** a.u. arbitrary unit. **d** Hub voxels' distribution in eight large-scale brain networks. Insets depicts the seven large-scale cortical networks<sup>29</sup>. SUB subcortical network, LIMB limbic network.

**Table 1 Highly consistent functional connectome hubs.**

No.	Hub	Location	MNI coordinates			Cohen's <i>d</i>	FCS	SE
			x	y	z			
1	Right Pft	Pft (superoanterior BA 40)	60	-21	45	6.267	1.072	0.022
2	Left Pft	Pft (superoanterior BA 40)	-60	-24	36	6.151	0.949	0.020
3	Right PF	PF (posterior BA 40)	60	-27	24	5.785	1.239	0.027
4	Left SCEF	Supplementary and cingulate eye field	0	0	51	5.635	1.000	0.023
5	Left PGi	PGi (inferior BA 39)	-51	-66	30	5.168	1.075	0.027
6	Left PFop	PF opercular (inferoanterior BA 40)	-63	-27	18	5.160	1.095	0.027
7	Left 43	Area 43	-57	3	3	4.927	1.114	0.029
8	Right 6r	Rostral area 6	57	6	0	4.916	1.184	0.031
9	Right PGi	PGi (inferior BA 39)	54	-60	30	4.739	1.007	0.027
10	Right 8BL	Area 8B lateral	21	36	51	4.655	0.713	0.020
11	Right 7PC	Area 7PC	36	-45	54	4.414	0.712	0.021
12	Left 9p	Area 9 posterior	-15	45	45	4.199	0.639	0.019
13	Right 6v	Ventral area 6	54	9	33	4.037	0.766	0.024
14	Left 8Av	Ventral area 8A	-39	18	48	3.990	0.561	0.018
15	Left AIP	Anterior intra-parietal area	-33	-45	45	3.474	0.567	0.021
16	Right FST	Fundus of the superior temporal area	54	-60	0	3.156	0.729	0.030
17	Right 9m	Area 9 middle	3	54	24	3.128	0.609	0.025
18	Left 31pv	Area 31p ventral	-3	-51	33	3.049	0.784	0.033
19	Right VIP	Ventral intra-parietal complex	18	-63	57	2.984	0.572	0.025
20	Right 6a	Area 6 anterior	33	3	63	2.975	0.454	0.020
21	Left FOP4	Frontal opercular area 4	-33	21	6	2.858	0.828	0.037
22	Right 5mv	Area 5m ventral	12	-30	45	2.822	0.701	0.032
23	Right 46	Area 46	36	42	30	2.779	0.656	0.030
24	Left 10v	Area 10v	0	57	-9	2.769	0.731	0.034
25	Left p9-46v	Area posterior 9-46v	-42	36	27	2.591	0.561	0.028
26	Left V3A	Area V3A	-15	-90	33	2.575	0.684	0.034
27	Left TE1a	Area TE1 anterior	-63	-15	-15	2.527	0.595	0.030
28	Right TE1a	Area TE1 anterior	60	-9	-21	2.494	0.580	0.030
29	Right IFSa	Anterior inferior frontal sulcus	48	39	12	2.468	0.480	0.025
30	Left 7Am	Medial area 7A	-12	-60	60	2.461	0.475	0.025
31	Right V3A	Area V3A	18	-87	36	2.442	0.645	0.034
32	Right V4	Fourth visual area	24	-63	-9	2.339	0.446	0.024
33	Left 6a	Area 6 anterior	-24	3	63	2.317	0.331	0.018
34	Left VMV1	Ventromedial visual area 1	-18	-60	-6	1.937	0.397	0.026
35	Left FEF	Frontal eye fields	-45	-9	57	1.412	0.640	0.058

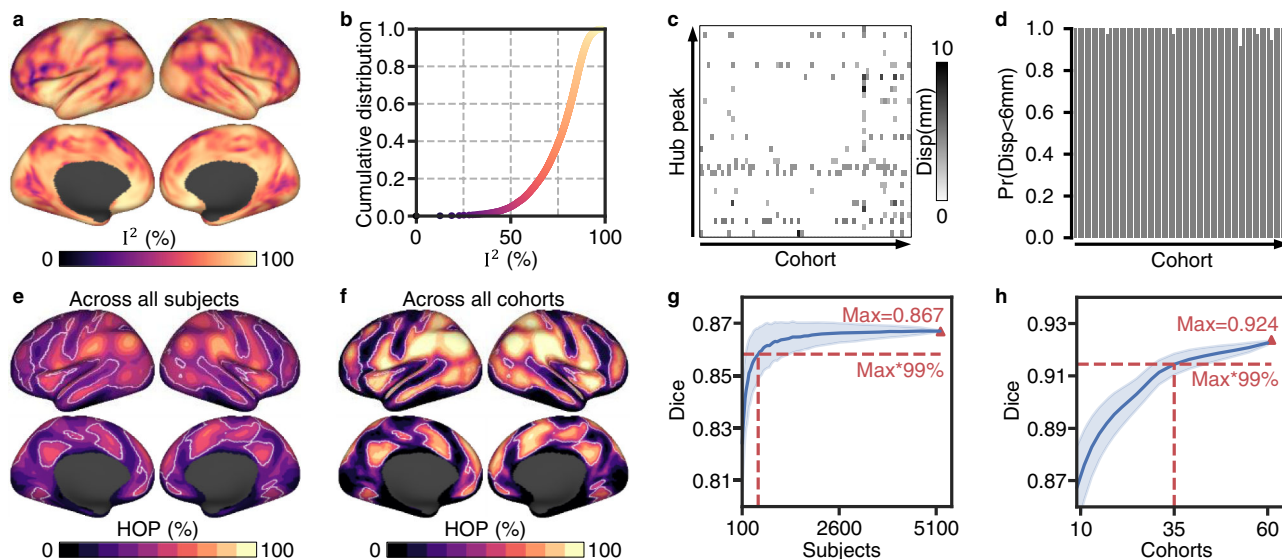
BA Brodmann area.

cohort number (Fig. 1c). According to prior brain network parcellations<sup>29,30</sup>, these identified hub voxels (15,461 voxels) were spatially distributed in multiple brain networks, including the DMN (27.5%), dorsal attention network (DAN) (16.5%), FPN (15.9%), ventral attention network (VAN) (15.6%), somatomotor network (SMN) (14.4%), and visual network (VIS) (9.9%) (Fig. 1d). Using a local maxima localization procedure, we identified 35 robust brain hubs across 61 cohorts (Fig. 1c and Table 1), involving various heteromodal and unimodal areas. Specifically, the most robust findings resided in several lateral parietal regions, including the bilateral ventral postcentral gyrus, supramarginal gyrus, and angular gyrus.

**The identified connectome hubs are reproducible across cohorts and individuals.** During identifying the above highly consistent connectome hubs, the random-effects meta-analysis revealed high heterogeneity of FCS across cohorts (Fig. 2a). The cumulative distribution function plot shows more than 95% voxels with  $I^2$  (heterogeneity score) exceeding 50% (Fig. 2b), indicating high heterogeneity across cohorts in almost all brain areas (see also Supplementary Fig. 1). To determine whether the connectome hubs identified here are dominated by certain cohorts or are reproducible across cohorts and individuals, we performed a leave-one-cohort-out validation analysis and an across-subject/cohort conjunction analysis.

**Leave-one-cohort-out validation analysis.** We repeated the above harmonized meta-analysis hub identification procedure after leaving one cohort out at a time. Comparing the identified hubs using all cohorts (Fig. 1c) with those after leaving one cohort out obtained extremely high Dice's coefficients ( $mean \pm sd$ :  $0.990 \pm 0.006$ ; range: 0.966-0.997). For hub peaks, leaving one cohort out resulted in very few displacements (mostly fewer than 6 mm, Fig. 2c, d). Thus, connectome hubs identified using the 61 cohorts were not dominated by specific cohorts.

**Across-subject/cohort conjunction analysis.** We defined the top  $N$  ( $N = 15,461$ , which is the voxel number of hubs in Fig. 1c) voxels with the highest FCS values of a subject or a cohort as connectome hubs for that subject or that cohort. Then, for each voxel, we assessed hub occurrence probability values across subjects and cohorts. The identified hubs using all cohorts were highly overlapped with the top  $N$  voxels with the highest hub occurrence probability values both across all subjects and across all cohorts, indicated by a high Dice's coefficient ( $Dice = 0.867$ , Fig. 2e;  $Dice = 0.924$ , Fig. 2f). When the identified hubs using all cohorts were compared with the top  $N$  voxels with the highest hub occurrence probability values across randomly selected subjects or across randomly selected cohorts, the Dice's coefficient approached 99% of its maximum value after exceeding 510 subjects (Fig. 2g) and 35 cohorts (Fig. 2h), respectively. This



**Fig. 2** The identified connectome hubs are reproducible across cohorts and individuals. **a** Heterogeneity measurement  $I^2$  estimated through the random-effects meta-analysis. **b** Cumulative distribution function plot of  $I^2$ . **c** Heatmap of displacements of the 35 hub peaks after leaving one cohort out. **d** Bar plot of the probability across the 35 hub peaks whose displacement was less than 6 mm after leaving one cohort out. **e, f** Hub occurrence probability (HOP) map across all subjects (**e**) and all cohorts (**f**). White lines delineate boundaries of the identified hubs in Fig. 1c. **g, h** Dice's coefficient of the identified hubs in Fig. 1c compared with the top  $N$  (voxel number of the identified hubs in Fig. 1c) voxels with the highest hub occurrence probability values across randomly selected subjects (**g**) and randomly selected cohorts (**h**). Blue shading represents the standard deviation across 2000 random selections.

indicated that the identified connectome hubs were highly reproducible both across cohorts and across individuals.

Validation analysis demonstrated that the above results did not depend on analysis parameters, such as the connection threshold (Supplementary Figs. 2 and 3), and were not driven by the size of the brain network to which they belong<sup>31</sup> (Supplementary Fig. 4), suggesting the robustness of our main findings.

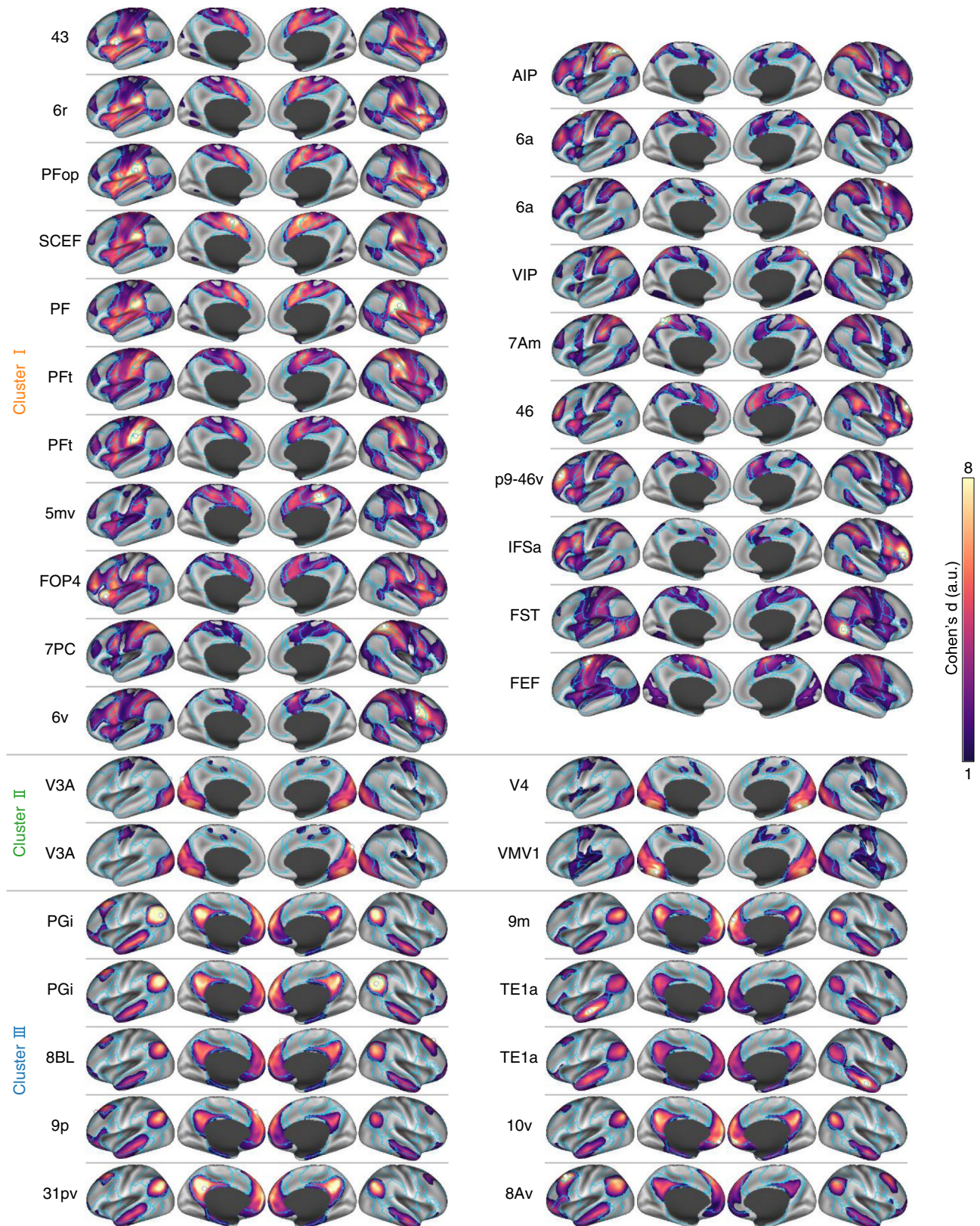
**Connectome hubs have heterogeneous functional connectivity profiles.** Next, we further examined whether these robust brain hubs (Fig. 1c and Table 1) have distinctive functional connectivity profiles that represent their unique roles in network communication. To gain detailed and robust functional connectivity profiles of each hub region, we conducted a seed-to-whole-brain connectivity meta-analysis in a harmonized protocol again. For each of the 35 hub regions, we obtained an estimated Cohen's  $d$  effect size map that characterizes the robust whole-brain connectivity pattern relevant to the seed region across the 61 cohorts (Fig. 3). We then divided the connectivity map of each hub into eight brain networks according to prior parcellations<sup>29,30</sup>, resulting in an  $8 \times 35$  connectivity matrix with each column representing the voxel percentage of each of the eight networks connected with a hub.

Hierarchical clustering analysis on the connectivity matrix clearly divided the 35 hubs into three clusters (Fig. 4a, b). Cluster I consists of 21 hubs that are primarily connected with extensive areas in the DAN, VAN, FPN, and SMN (orange, Fig. 4c). Cluster II consists of four hubs that are densely connected with VIS (green, Fig. 4c). Cluster III consists of 10 hubs that have robust connections with the DMN and LIMB (blue, Fig. 4c). Of particular interest is that within Cluster III, a left posterior middle frontal hub called ventral area 8A (8Av) shows a distinctive connectivity profile in contrast to the other nine hubs, manifested as having robust connections with bilateral frontal FPN regions (Fig. 3 and Supplementary Fig. 5). This implies that the left 8Av hub is a key connector between the DMN and FPN, which can be supported by the recent finding of a control-default connector located in the posterior middle frontal gyrus<sup>32</sup>.

Although both Cluster I and III hubs are connected with subcortical structure (Fig. 4c), they are connected with different subcortical nuclei (Supplementary Fig. 6). Finally, whereas all hubs possess dense intranetwork connections, most also retain substantial internetwork connections (Supplementary Fig. 7), which preserves efficient communication across the whole brain network feasible.

**Transcriptomic data distinguishes connectome hubs from non-hubs.** A supervised machine learning classifier based on XGBoost<sup>33</sup> and 10,027 genes' transcriptomic data from the AHBA<sup>34</sup> was trained to distinguish connectome hubs from non-hubs (Fig. 5a). The sensitivity, specificity, and accuracy rate of the XGBoost classifier were stably estimated by repeating the training and testing procedure 1000 times. This classifier performed better than chance in all 1000 repetitions and achieved an overall accuracy rate of 65.3% (Fig. 5b). In cross-validation, connectome hubs and non-hubs were classified with a sensitivity of 71.1% and specificity of 63.4%, respectively. The testing procedure yielded a comparable sensitivity of 69.7% and specificity of 62.0%. After training the classifier, each gene's contribution to the optimal prediction model was determined. We noted that some key genes contributed two or three orders of magnitude more than other genes (Fig. 5c and Supplementary Data 1). The contributions of the top 300 genes with the greatest contributions to the XGBoost classifier were consistent between the first 500 repetitions and the second 500 repetitions (Pearson's  $r = 0.958$ ,  $p < 10^{-6}$ , Fig. 5d), suggesting a high reproducibility.

To exclude the XGBoost model's potential bias relating to the mostly contributed key genes, we replicated the above classification results using another machine learning model based on the support vector machine (SVM) that was trained using only the top  $N$  key genes with the greatest contributions to the XGBoost classifier (Fig. 5e). Because no data were available to determine how many key genes were sufficient to train an SVM classifier, we examined the count  $N$  from 100 to 300. The SVM classifier achieved a very high peak accuracy rate of 91.8% with approximately the top 150 key genes in the easiest classification

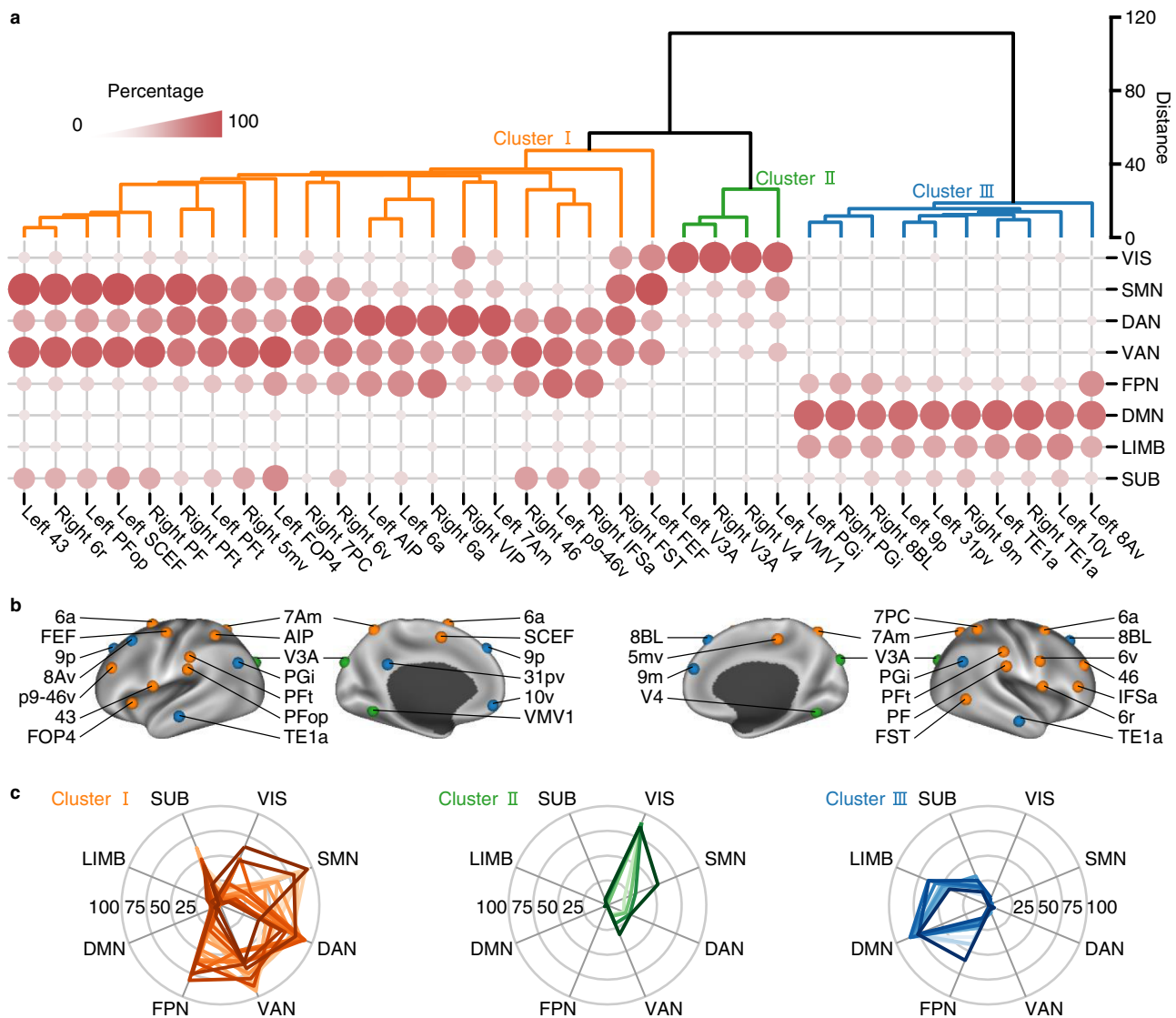


**Fig. 3 Functional connectivity maps of connectome hubs.** Cluster labels were derived by the hierarchical clustering solution in Fig. 4a. White spheres represent hub seeds. Blue lines delineate boundaries of the seven cortical networks shown in Fig. 1d. a.u. arbitrary unit.

task (Fig. 5f) and also achieved a reasonable peak accuracy rate of 67.8% with approximately the top 150 key genes even in the most difficult classification task (Fig. 5g). By contrast, SVM classifiers trained using 150 randomly selected genes performed

worse than that using the top 150 key genes in all 1000 repetitions (Fig. 5h).

Validation analyses showed that the XGBoost and SVM classifiers trained using surrogate hub identification maps with



**Fig. 4 Hierarchical clustering analysis on connectome hubs' functional connectivity maps.** **a** Dendrogram derived by hierarchical clustering on the connectivity percentage matrix. **b** The 35 hubs were rendered using three different colors according to the hierarchical clustering solution. **c** Radar charts showing heterogeneous connectivity profiles of the three hub clusters.

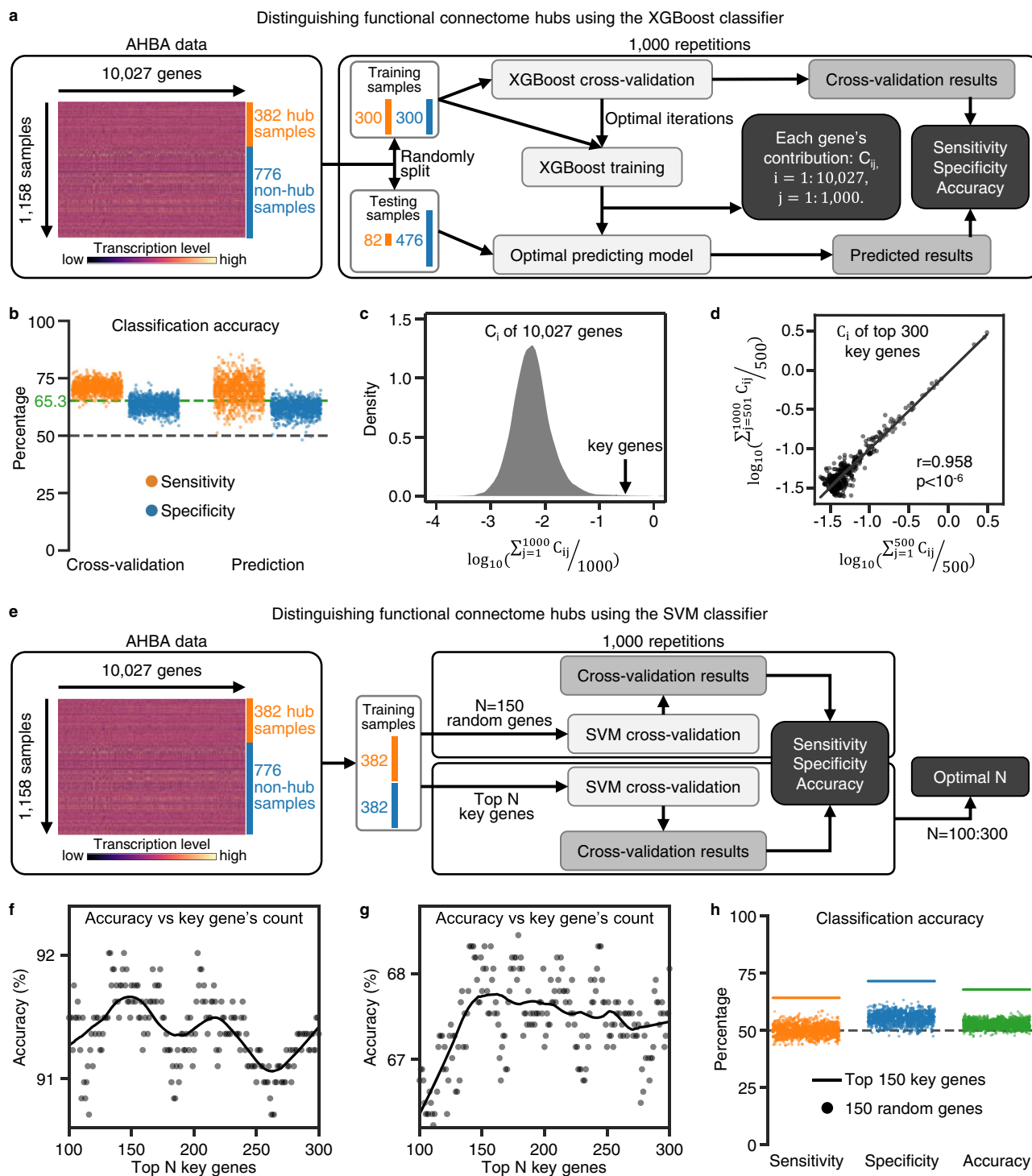
the spatial autocorrelations being corrected performed no better than the chance level (Supplementary Fig. 8), confirming that the performance of the XGBoost and SVM classifiers was not driven by the effects of spatial autocorrelation inherent to the hub localization and the transcriptomic data. Thus, these robust connectome hubs were apparently associated with a transcriptomic pattern dominated by approximately 150 key genes.

**Connectome hubs have a spatiotemporally distinctive transcriptomic pattern.** Gene Ontology (GO) enrichment analysis using GOrilla<sup>35</sup> demonstrated that the above 150 key genes were mostly enriched in the neuropeptide signaling pathway (*fold enrichment* ( $FE$ ) = 8.9, uncorrected  $p = 1.2 \times 10^{-5}$ , Supplementary Data 2). GO enrichment analysis using the ranked 10,027 genes according to their contributions to the XGBoost classifier also confirmed the most enriched GO term of the neuropeptide signaling pathway ( $FE = 5.7$ , uncorrected  $p < 10^{-6}$ , Supplementary Data 3). The ranked 10,027 genes were also associated with the developmental process ( $FE = 1.2$ ), cellular developmental process ( $FE = 1.3$ ), anatomical structure development ( $FE = 1.3$ ), and neuron projection arborization ( $FE = 13.7$ ) (uncorrected

$p < 5.5 \times 10^{-4}$ , Supplementary Data 3). We speculated that connectome hubs have a distinctive transcriptomic pattern of neurodevelopmental processes in contrast to non-hubs.

We repeated the GO enrichment analysis of the above 150 key genes using DAVID<sup>36,37</sup> and confirmed the mostly enriched GO term of the neuropeptide signaling pathway ( $FE = 8.7$ , uncorrected  $p = 5.8 \times 10^{-4}$ , Supplementary Data 4). In addition, there were 10 GO terms associated with metabolic process, such as the positive regulation of cellular metabolic process ( $FE = 1.4$ , uncorrected  $p = 0.031$ , Supplementary Data 4). Disease association analysis demonstrated metabolic disease associated with the greatest number of key genes (60 genes,  $FE = 1.2$ , uncorrected  $p = 0.094$ , Supplementary Data 5). Accordingly, it is rational to speculate that connectome hubs have a distinctive transcriptomic pattern of metabolic processes in contrast to non-hubs.

To confirm the above two speculations of GO enrichment analysis results, we examined transcription level differences between hub and non-hub regions for genes previously implicated in key neurodevelopmental processes<sup>38</sup> (Supplementary Data 6) and main neuronal metabolic pathways<sup>39</sup> (oxidative phosphorylation<sup>40</sup> and aerobic glycolysis<sup>41</sup>, Supplementary Data 7). Permutation tests



**Fig. 5** Transcriptomic data distinguishes connectome hubs from non-hubs. **a** Schematic diagram of using the XGBoost model to classify brain samples as a hub or non-hub. **b** Performance of the XGBoost classifier. Each dot represents one repetition in **a**. The horizontal gray dashed line represents the chance level accuracy rate (50%). The horizontal green dashed line represents the average accuracy rate (65.3%) of the XGBoost classifier across 1000 repetitions. **c** Density plot of 10,027 genes' logarithmic average contributions across 1000 repetitions to the XGBoost classifier. Genes with the greatest contributions were regarded as key genes. **d** Regression plot of the logarithmic average contributions of the top 300 key genes across the first 500 repetitions versus those across the second 500 repetitions. Each dot represents one gene. **e** Schematic diagram of using the SVM model to classify brain samples as a hub or non-hub. **f, g** Accuracy rate of the SVM classifier versus the count of key genes used to distinguish 382 hub samples from 382 non-hub samples with the highest rate (**f**) or lowest rate (**g**) to be correctly classified by the XGBoost classifier. Each dot represents one SVM classifier. Black curves were estimated by locally weighted regression. **h** Performance of the SVM classifier. Horizontal lines correspond to the SVM classifier trained using top 150 key genes in **g**. Each dot represents one repetition using 150 randomly selected genes in **e**. The horizontal gray dashed line represents the chance level accuracy rate (50%).

revealed hub regions with significantly higher transcription levels for genes associated with dendrite development, synapse development, and aerobic glycolysis than non-hub regions (one-sided Wilcoxon rank-sum tests, Bonferroni-corrected  $ps \leq 0.032$ , Fig. 6a). In addition, hub regions had a weak trend of lower transcription levels for genes associated with axon development, myelination, and neuron migration, and a higher transcription level for genes associated with oxidative phosphorylation (Fig. 6a). These differences in transcription level were consistent with our speculations of GO enrichment analysis results.

These above transcriptomic results were derived from the AHBA, an adult transcriptomic dataset. To explore their developmental evolutions, we inspected the developmental trajectory of transcription level in hub and non-hub regions respectively using the BrainSpan Atlas<sup>42</sup>. We observed diverging developmental trajectories of transcription level between hub and non-hub regions in these key neurodevelopmental processes and main neuronal metabolic pathways (Fig. 6b and Supplementary Fig. 9a). The magnitude of differences in developmental trajectory between hub and non-hub regions continuously exceeds the median absolute deviation of transcription level across brain regions during some periods (Fig. 6c and Supplementary Fig. 9b), suggesting a trend of greater difference than expected. Specifically, hub regions have higher transcription levels for neuron migration during the late-fetal period, higher transcription levels for dendrite and synapse development from the late-childhood to mid-adolescence period, and lower transcription levels for axon development and myelination from the mid-childhood to late-adolescence period than non-hub regions. These results are in agreement with the observation of primary somatosensory, auditory, and visual (V1/V2) cortices with lower synapse density but higher myelination than the prefrontal area<sup>43,44</sup>. Moreover, hub regions have higher transcription levels than non-hub regions for aerobic glycolysis since the early childhood period. These transcriptome analyses achieved convergent results between the AHBA and BrainSpan Atlas.

Together, functional connectome hubs have a spatiotemporally distinctive transcriptomic pattern in contrast to non-hubs, which is dominated by genes involved in the neuropeptide signaling pathway, neurodevelopmental processes, and metabolic processes.

**Neural contextualization of connectome hubs' transcriptomic pattern.** To assess the neural relevance of the above identified transcriptomic pattern underlying functional connectome hubs, we contextualized it relative to prior established neuroimaging maps. The identified transcriptomic pattern is dominated by genes with the highest enrichment for the neuropeptide signaling pathway. Considering that neuropeptides are a main type of indirect neurotransmitter widely distributed in the human central nervous system and their vital role in modulating direct excitatory and inhibitory transmission<sup>45</sup>, it is rational to speculate that there are apparent differences in neurotransmitter systems between hub and non-hub regions. Using neurotransmitter maps derived from positron emission tomography and single photon emission computed tomography<sup>46</sup>, we found that hub regions have higher density of GABAa, glutamate, mu opioid, cannabinoid, dopamine D2, and serotonin receptor and norepinephrine transporter but lower density of dopamine transporter and fluorodopa than non-hub regions (one-sided Wilcoxon rank-sum tests, Bonferroni-corrected  $ps \leq 0.015$ , Fig. 7a).

Growing evidence has suggested a striking spatial correspondence between transcriptomic profile and structural connectivity in the human brain<sup>27</sup>. We speculated that the above differences in microscale transcriptome between hub and non-hub regions in

key neurodevelopmental processes may result in differences in macroscale structural connectivity profile. Using a fiber length profiling dataset<sup>47</sup>, we observed that hub regions possess more fibers with a length exceeding 40 mm but less fibers with a length shorter than 40 mm (one-sided Wilcoxon rank-sum tests, Bonferroni-corrected  $ps \leq 0.007$ , Fig. 7b), suggesting a more intricate fiber configuration in hub regions.

The above transcriptome analyses have shown a higher transcription level of oxidative phosphorylation and aerobic glycolysis in hub regions than in non-hubs. We validated this observation using a metabolism dataset derived from positron emission tomography<sup>48</sup> and found that hub regions not only have a higher metabolic rate than non-hubs in oxidative phosphorylation (indicated by the cerebral metabolic rate for oxygen) and aerobic glycolysis (indicated by the glycolytic index), but also have more blood supply (indicated by the cerebral blood flow) (one-sided Wilcoxon rank-sum tests, Bonferroni-corrected  $ps < 0.001$ , Fig. 7c). This is in agreement with prior observations of a tight coupling between FCS and blood supply<sup>1,49</sup>.

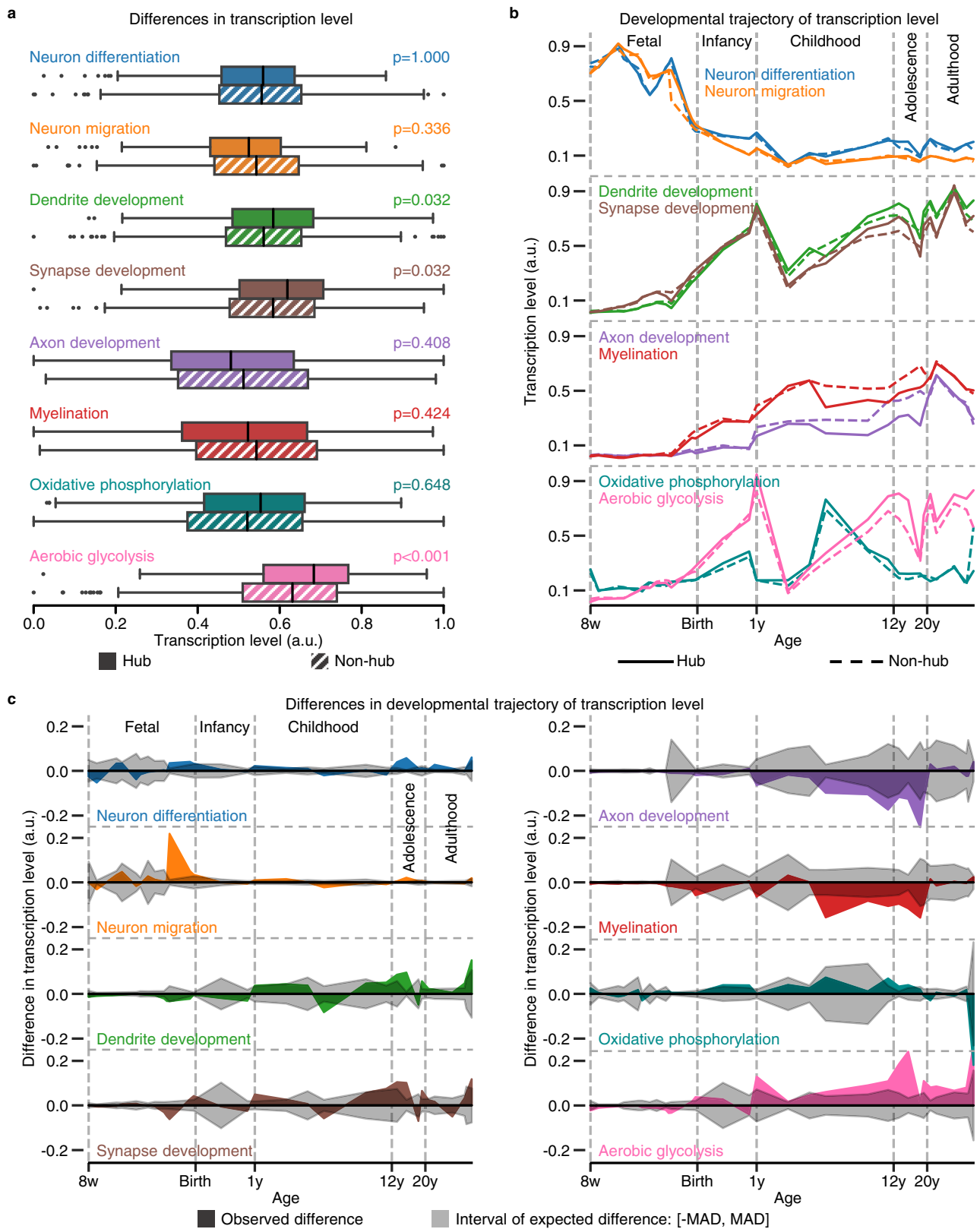
In addition, we also noted that the above 150 key genes are enriched for several psychiatric disorders ( $FE = 3.5$ , uncorrected  $p = 5.5 \times 10^{-4}$ , Supplementary Data 5). This finding is in accordance with prior observations of hub regions being preferentially targeted by neuropsychiatric disorders<sup>5-8</sup>. This implies that connectome hubs may have different susceptibility to neuropsychiatric disorders in contrast to non-hubs. We validated it by performing an association analysis between the effect size of connectome hub and the effect size of cortical thickness atrophy in neuropsychiatric disorders<sup>50</sup>. We observed that the Cohen's  $d$  of connectome hub is negatively correlated with the Cohen's  $d$  of cortical thickness atrophy in 22q deletion syndrome (Pearson's  $r = -0.292$ , uncorrected  $p = 0.009$ ) and autism spectrum disorder (Pearson's  $r = -0.333$ , uncorrected  $p = 0.019$ ) but positively correlated with the Cohen's  $d$  of cortical thickness atrophy in bipolar disorder (Pearson's  $r = 0.418$ , uncorrected  $p = 0.003$ ) and schizophrenia (Pearson's  $r = 0.247$ , uncorrected  $p = 0.040$ ) (Fig. 7d). This suggests that connectome hubs have a trend of higher susceptibility to cortical thickness atrophy in bipolar disorder and schizophrenia but lower susceptibility to cortical thickness atrophy in 22q deletion syndrome and autism spectrum disorder than non-hubs.

## Discussion

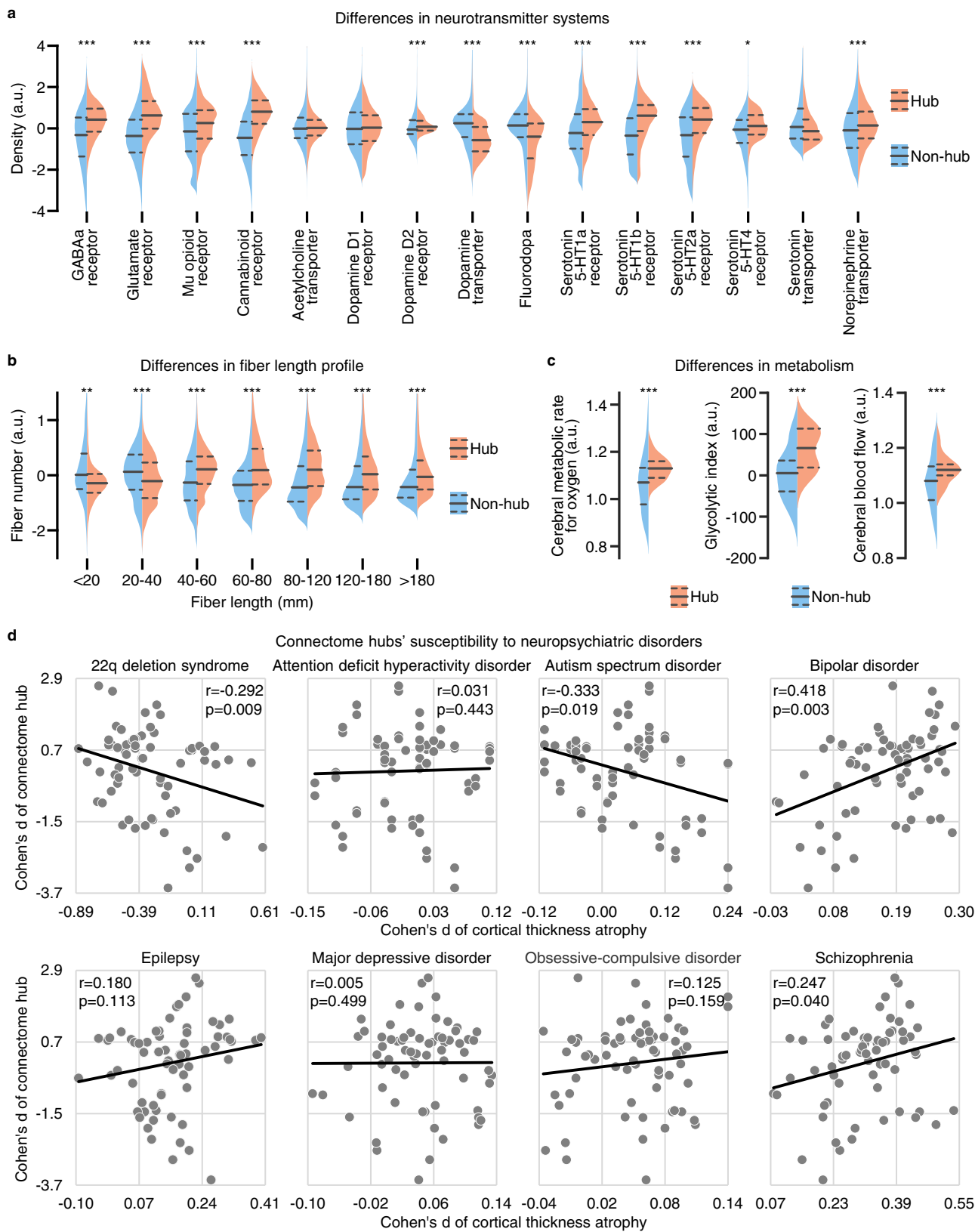
Using a worldwide harmonized meta-connectomic analysis of 5212 healthy young adults across 61 cohorts, we provided, to the best of our knowledge, the first description of highly consistent and reproducible functional connectome hubs in the resting human brain. Using transcriptomic data from the AHBA and BrainSpan Atlas, we reported that these robust connectome hubs have a spatiotemporally distinctive transcriptomic pattern in contrast to non-hub regions. These results advanced our knowledge of the robustness of macroscopic functional connectome hubs and their potential cellular and molecular substrates.

Extant reports have shown largely inconsistent and less reproducible hub localizations<sup>7,8,13-19</sup>, which may arise from high heterogeneity in the included subjects, data acquisition, and analysis strategies across studies. To diminish these potential confounding factors, we employed stringent participant inclusion criteria that included only healthy young adults aged 18–36 years and adopted harmonized data preprocessing and connectome analysis protocols across cohorts. Nevertheless, the random-effects meta-analysis revealed high heterogeneity among cohorts in almost all brain areas, which implied that heterogeneity of imaging scanners and/or imaging protocols could be an important cause for inconsistent and less reproducible results across





**Fig. 6 Connectome hubs have a spatiotemporally distinctive transcriptomic pattern.** **a** Differences in transcription level between hub samples ( $n = 382$ ) and non-hub samples ( $n = 776$ ) for genes associated with key neurodevelopmental processes<sup>38</sup> and main neuronal metabolic pathways<sup>39</sup>. Boxplot left and right edges, vertical black lines, and whiskers and dots depict the 25th and 75th percentiles, median, and extreme nonoutlier and outlier values, respectively. The statistical significances of one-sided Wilcoxon rank-sum tests were determined by 1000 permutation tests and were labeled with Bonferroni-corrected  $p$  values. **b** Developmental trajectory of transcription level in hub and non-hub regions for genes involved in key neurodevelopmental processes<sup>38</sup> and main neuronal metabolic pathways<sup>39</sup>. **c** Differences in the developmental trajectory of transcription level between hub and non-hub regions shown in **b**. MAD, the median absolute deviation of transcription level across brain regions. w post-conceptional week, y postnatal year, a.u. arbitrary unit.



prior studies. Thus, our study was indispensable by conducting a harmonized random-effects meta-analysis model in which both intracohort variation (i.e., sampling errors) and intercohort heterogeneity were considered<sup>51</sup>. In addition, our validation results showed that the spatial distribution of functional connectome hubs was relatively stable when using more than 510 subjects and

35 cohorts, demonstrating that 5212 subjects from 61 cohorts were adequate to minimize the effects of both sampling errors and heterogeneity among cohorts. Considering only dozens of subjects in most prior studies<sup>7,8,13-15,17,19</sup>, the low statistical power attributed to inadequate subjects could be another cause for prior inconsistent and less reproducible hub localizations.

**Fig. 7 Neural contextualization of connectome hubs' transcriptomic pattern.** **a–c** Differences between hub (red) and non-hub (blue) regions in density of neurotransmitter receptor and transporter (**a**, hub voxels  $n = 15,461$ , non-hub voxels  $n = 32,158$ ), fiber number for different fiber length bins (**b**, hub vertices  $n = 25,944$ , non-hub vertices  $n = 33,195$ ), and metabolic rate for oxygen, aerobic glycolysis, and blood supply (**c**, hub regions  $n = 29$ , non-hub regions  $n = 60$ ). For each violin plot, dashed gray lines depict the 25th and 75th percentiles, solid gray line depicts median value. The statistical significances of one-sided Wilcoxon rank-sum tests were determined by 1000 permutation tests and were labeled with Bonferroni-corrected  $p$  values. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . a.u. arbitrary unit. **d** Regression plot of the Cohen's  $d$  value of connectome hub versus the Cohen's  $d$  value of cortical thickness atrophy across 68 cortical areas for eight disorders. Positive Cohen's  $d$  value indicates thinning of cortical thickness in patients. Each dot represents one cortical area. The statistical significances of Pearson's correlation coefficients were determined by 1000 permutation tests and were labeled with uncorrected  $p$  values.

Finally, we used harmonized image processing and connectome analysis protocols across cohorts, which avoided methodological variation and reduced potential methodological defects that have not been resolved in prior studies. See an extension discussion in Supplementary Note 1.

The present results demonstrated that the 35 highly consistent and reproducible connectome hubs show heterogeneous functional connectivity profiles, forming three clusters. Twenty-one hubs (Cluster I) are connected with extensive areas in the DAN, VAN, FPN, and SMN. Previous investigations indicated that they are core regions of the DAN (left AIP, right 7PC, left 7Am, bilateral PFt, left FEF, bilateral 6a, right 6v, and right FST)<sup>29,52</sup>, VAN (left 43, left FOP4, right 46, right 6r, right PF, left PFop, left SCEF, right 5mv)<sup>29,52</sup>, and FPN (left p9-46v and right IFSa)<sup>29,53</sup>. In addition, hub regions involved in the sensorimotor pathway (right VIP, right FST, left 7Am, and left FEF)<sup>54</sup> are also connected with the visual association cortex, acting as connectors between the VIS and the SMN, DAN, and VAN. Information flow along the primary visual, visual association, and higher-level sensorimotor cortices is undertaken by the four occipital hubs (Cluster II) left VMV1, right V4, and bilateral V3A that are all densely connected with the VIS and portions of the SMN, DAN, and VAN. This is supported by the report of their dense connections with both the visual system and SMN region the frontal eye field, DAN region the superior parietal cortex, and VAN region the parietal operculum and anterior insula<sup>55</sup> and also aligns with the role of their homologous regions in the non-human primate cerebral cortex<sup>54</sup>. The remaining 10 hubs (Cluster III) are all located in canonical DMN regions<sup>56</sup>. One of them, the left 8Av hub, is robustly connected with both DMN and lateral prefrontal FPN regions, acting as a connector between the DMN and FPN. This can be supported by the recent finding of a control-default connector located in the posterior middle frontal gyrus<sup>32</sup> and may also be a case of the hypothesis of parallel interdigitated subnetworks<sup>57</sup> where the posterior middle frontal gyrus is connected with a subnetwork of the DMN and some regions of the FPN. This observation offers a crucial complementary interpretation to the conventional assumption that the DMN is anticorrelated with other networks<sup>56</sup>. Considering that communication between the DMN and other networks is of particular relevance to neuropsychiatric disorders<sup>58</sup>, such as autism spectrum disorders<sup>59</sup>, we speculated that the left 8Av hub may be a promising target region for therapeutic interventions.

We demonstrated that these robust brain hubs have a spatiotemporally distinctive transcriptomic pattern dominated by genes with the highest enrichment for the neuropeptide signaling pathway. Because neuropeptides are a main type of indirect neurotransmitter that is widely distributed in the human central nervous system<sup>45</sup>, robust neuropeptide signaling pathways are indispensable for efficient synaptic signal transduction that sustains dense and flexible functional connections of hub regions. This is also supported by our observation of differences in neurotransmitter receptor and transporter density between hub and non-hub regions. In addition, hub regions have higher transcription levels for main neuronal metabolic pathways in contrast

to non-hubs. This is reasonable because massive synaptic activities in hub regions demand high material and metabolic costs, which is in accordance with our observation of more blood supply and higher oxidative phosphorylation and aerobic glycolysis levels in hub regions. This is also consistent with prior observations of a tight coupling between FCS and blood supply<sup>1,49</sup>.

We found that connectome hubs possess a spatiotemporally distinctive transcriptomic pattern of key neurodevelopmental processes in contrast to non-hubs. Specifically, connectome hubs have higher transcription levels for dendrite and synapse development and lower transcription levels for axon development and myelination during childhood, adolescence, and adulthood. These findings are compatible with previous observations of the prefrontal area having higher synapse density but lower myelination than primary somatosensory, auditory, and visual (V1/V2) cortices<sup>43,44</sup>. Higher transcription levels for dendrite and synapse development in hub regions are necessary for the overproduction of synapses that will be selectively eliminated based on the demand of the environment and gradually stabilized before full maturation<sup>60</sup>, which has been proposed as the major mechanism of creating diverse neuronal connections beyond their genetic determination<sup>60</sup>. Lower transcription levels for axon development and myelination will prolong the myelination period in hub regions, which characterizes a delayed maturation phase<sup>61</sup>. Marked delay of anatomical maturation in human prefrontal and lateral parietal cortices has been frequently observed both in human development<sup>62,63</sup> and in primate evolution<sup>61</sup>, which provides more opportunities for social learning to establish diverse neuronal circuits that contribute to our complex<sup>63</sup> and species-specific<sup>61</sup> cognitive capabilities. We also observed higher transcription levels for neuron migration in hub regions from mid-fetal period to early infancy. This is in agreement with the report of extensive migration of young neurons persisting for several months after birth in the human frontal cortex<sup>64</sup>. Meanwhile, the migration and final laminar positioning of postmitotic neurons are regulated by common transcription factors<sup>65</sup>, which suggests that a higher transcription level for neuron migration in hub regions facilitates the construction of more intricate inter-laminar connectivity. These microscale divergences of key neurodevelopmental processes may result in a more intricate macroscale structural connectivity profile in hub regions.

Human neurodevelopment is an intricate and protracted process, during which the transcriptome of the human brain requires precise spatiotemporal regulation<sup>38</sup>. Thus, in addition to contributing to our complex cognitive capabilities, the spatiotemporal differences in transcriptomic pattern of neurodevelopment between hub and non-hub regions may also increase brain connectome's susceptibility to neuropsychiatric disorders<sup>61,63</sup>, which means small disturbance in the magnitude or the timing of this transcriptomic pattern may have long-term consequences on brain anatomical topography or functional activation. This is in line with the result of several psychiatric disorders being the most significant disease associated with the top 150 key genes and is also supported by our observation of differences in susceptibility

to cortical thickness atrophy in neuropsychiatric disorders between hub and non-hub regions. This implies that uncovering the intricate transcriptomic pattern, diverse neuronal circuits, anatomical topography, and functional activation of connectome hubs provide crucial and promising routes for understanding the pathophysiological mechanisms underlying neurodevelopmental disorders, such as autism spectrum disorders<sup>38,59</sup> and schizophrenia<sup>5,38,61,63</sup>.

Of note, we conducted transcriptome–connectome association analysis using machine learning approaches in which non-linear mathematical operations were implemented rather than linear operations, such as linear correlation<sup>24</sup>, linear regression<sup>25</sup>, or partial least squares<sup>26</sup>. It has been argued that observations of transcriptome–connectome spatial association have a high false-positive rate through linear regression<sup>66</sup> and linear correlation<sup>67</sup> and may be largely shifted toward the first principal component axis of the dataset through partial least squares<sup>68</sup>. These investigations imply that prior transcriptome–connectome association results by linear mathematical operations may include high false-positive observations that are independent of connectome measurements, such as genes enriched for ion channels<sup>24–26</sup>. By contrast, high reproducibility across different machine learning models and across different GO enrichment analysis tools and convergent results from the AHBA and BrainSpan Atlas made it very unlikely that our findings were false-positive observations.

Some results of the present study should be interpreted cautiously because of methodological issues. First, we identified the robust connectome hubs using preprocessed rsfMRI data with global signal regression because of its great promise in minimizing physiological artifacts on functional connectomes<sup>69</sup>. Validation analysis demonstrated that hub distribution identified without global signal regression was more likely derived from physiological artifacts rather than by ongoing neuronal activity (Supplementary Note 2 and Supplementary Fig. 10). Second, we conducted a voxel-based connectome analysis in order to directly compare our results with the extant voxel-based reports<sup>7,8,14–19</sup> and increase the sensitivity of identifying spatially focal (e.g., voxel-sized) hubs<sup>70</sup>. The effects of parcellation-based<sup>70</sup> and surface-based<sup>71</sup> analysis on hub localizations should be resolved in future studies. Third, the AHBA dataset only includes partial human genes, of which approximately half were excluded in data preprocessing<sup>34</sup>, which may have induced incomplete observations in our data-driven analysis. Finally, our transcriptomic signature results addressed only the association between connectome hubs and transcriptomic patterns and did not explore causation between them. Exploring more detailed mechanisms underlying this association is attractive and may be practicable for non-human primate brains in future studies.

## Methods

**Dataset.** We collected a large-sample rsfMRI dataset ( $N = 7202$ ) from public data-sharing platforms and in-house cohorts, which consists of 73 cohorts from Asia, Europe, North America, and Australia. Data of each cohort were collected with participants' written informed consent and with approval by the respective local institutional review boards.

**Image preprocessing and quality control.** We first reviewed T1-weighted structural MRI data for all participants with the assistance of a neuroradiologist and a clinical neurologist to confirm no identifiable lesion or structural abnormality (e.g., regional atrophy and posterior cranial fossa arachnoid cyst). All rsfMRI data for the remaining participants were preprocessed routinely using SPM12 v6470 and GREYNA<sup>72</sup> v2.0.0 with a uniform pipeline. For each individual, we discarded the first 10 s' volumes for magnetic field stabilization and the participant's adaptation to the scanner. Next, slice-timing was corrected within each volume. To correct for head motion, all volumes were realigned to the mean image. Participants with significant head motion (translation above 3 mm or rotation above 3° in any direction) were excluded from the subsequent analyses. Then, all volumes were normalized to the 3-mm isotropic space of the Montreal Neurological Institute (MNI) using the EPI template provided by SPM12. The normalized

volumes were spatially smoothed using a 6-mm full-width at half-maximum Gaussian kernel. After that, the time series of each voxel underwent the procedure of linear trend removal, nuisance signal regression (24 head motion parameters, white matter, cerebrospinal fluid, and global brain signals), and temporal band-pass filtering (0.01–0.1 Hz). Finally, scrubbing was performed to minimize head motion effects<sup>73</sup>. Specifically, volumes with framewise displacement exceeding 0.5 mm and their adjacent volumes (1 back and 2 forward) were replaced with linearly interpolated data. We excluded participants with more than 25% interpolated volumes. Notably, slice-timing was not corrected in the Human Connectome Project cohort due to multiband acquisition<sup>74</sup>. For participants with more than one rsfMRI scan, we used only one of them. To reduce the potential effects of development and aging on our results, we restricted our analysis to healthy young adults aged 18 to 36 years. To ensure sufficient statistical power, twelve cohorts were discarded due to having fewer than 10 participants that passed quality controls. After these stringent quality controls, we included preprocessed rsfMRI data of 5212 healthy young adults (2377 males) from 61 independent cohorts in the final analysis. The sample size and age ranges of each cohort were summarized in Fig. 8. Supplementary Data 8 provides detailed information on the individual cohorts.

**Identifying robust functional connectome hubs using a harmonized meta-analysis.** For each individual, we constructed a voxelwise functional connectome matrix by computing the Pearson's correlation coefficient between preprocessed rsfMRI time series of all pairs of voxels within a predefined gray matter mask (47,619 voxels). The gray matter mask was divided into seven large-scale cortical networks<sup>29</sup> and a subcortical network<sup>30</sup>. The cerebellum was not included due to largely incomplete coverage during rsfMRI scanning in most cohorts. Negative functional connections were excluded from our analysis due to neurobiologically ambiguous interpretations<sup>75</sup>. To further reduce the bias of signal noise and simultaneously avoid the effect of potential sharing signals between nearby voxels, both weak connections (Pearson's  $r < 0.1$ ) and connections terminating within 20 mm were set to zero<sup>76</sup>. We validated the threshold of weak connections using 0.05 and 0.2 (Supplementary Figs. 2 and 3). For each voxel, we computed the FCS as the sum of connection weights between the given voxel and all the other voxels. We further normalized this resultant FCS map with respect to its mean and standard deviation across voxels<sup>7</sup>.

For each cohort, we performed a general linear model on these normalized FCS maps to reduce age and sex effects. For each voxel, we constructed the general linear model as:

$$FCS_i = \beta_0 + \beta_{Age} * (Age_i - MeanAge) + \beta_{Sex} * Sex_i + \epsilon_i \quad (1)$$

$FCS_i$ ,  $Age_i$ ,  $Sex_i$ , and  $\epsilon_i$  indicate the FCS, age, sex, and residual of the  $i$ th individual, respectively.  $MeanAge$  indicates the mean age of that cohort.  $\beta_0$  indicates the mean FCS of that cohort. The general linear model exported a mean FCS map and its corresponding variance map for each cohort.

The mean and variance FCS maps of the 61 cohorts were submitted to a random-effects meta-analysis model<sup>51</sup> to address across-cohort heterogeneity of functional connectomes. A short summary of the random-effects meta-analysis was provided in the following section. The detailed computational procedures are described in the book<sup>51</sup>. This resulted in a consistent FCS pattern (Fig. 1a) and its corresponding SE map (Fig. 1b). We compared the FCS of each voxel with the average of the whole brain (i.e., zero) using a  $Z$  value and estimated effect size using Cohen's  $d$  metric<sup>51</sup>:

$$Z = \frac{FCS - 0}{SE} \quad (2)$$

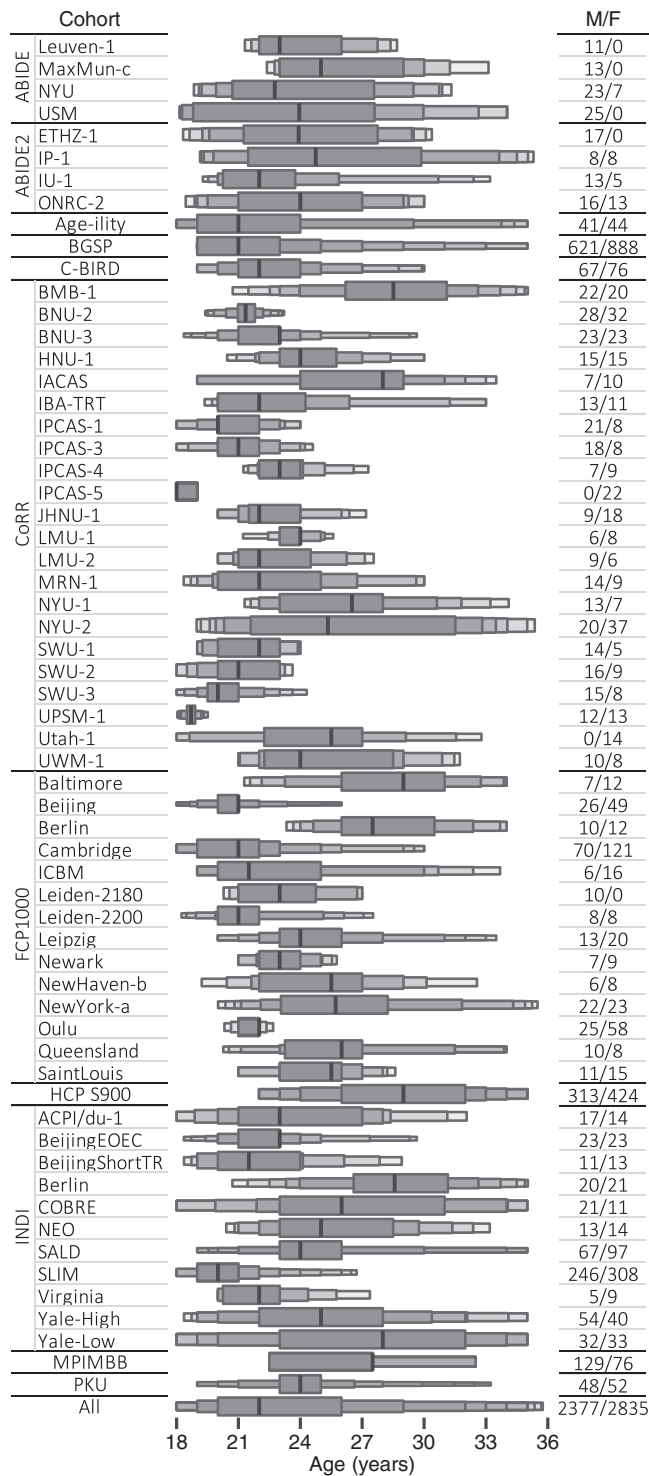
$$d = \frac{Z}{\sqrt{k}} \quad (3)$$

$k$  is the number of cohorts in the meta-analysis.

In line with a previous neuroimaging meta-analysis study<sup>77</sup>, we performed 10,000 one-sided nonparametric permutation tests<sup>28</sup> to assign a  $p$  value to the observed  $Z$  value. For each iteration, after randomizing the spatial correspondence among cohorts' mean FCS maps (the spatial correspondence between a cohort's mean FCS map and its variance map was not changed), we repeated the computation procedure of the random-effects meta-analysis for each voxel and extracted the maximum  $Z$  value of all voxels to construct a null distribution. A  $p$  value was assigned to each voxel by comparing the observed  $Z$  value to the null distribution. For a statistical significance level below 0.05, this  $p$  value closely tracks the Bonferroni threshold<sup>28</sup>.

Finally, we defined functional connectome hubs as brain regions with a  $p$  value less than 0.001 and cluster size greater than 200 mm<sup>3</sup> (Fig. 1c). The thresholds of  $p$  value and cluster size were similar with the activation likelihood estimation algorithm<sup>77</sup>. We extracted MNI coordinates for each local peak  $Z$  value terminating beyond 15 mm within each brain cluster using the `wb_command -volume-extrema` command in Connectome Workbench v1.4.2.

**Random-effects meta-analysis.** For each voxel,  $M_i$ ,  $SD_i$ , and  $N_i$  indicate the mean and standard deviation value and the participant number of the  $i$ th cohort, respectively. The original weight assigned to the  $i$ th cohort is the inverse of its



**Fig. 8 Enhanced box plot of the age ranges of each cohort.** Vertical black lines depict the median value. Left and right edges of the incrementally narrower boxes depict the lower and upper fourths, eighths, sixteenths, etc. M/F males/females.

variance:

$$W_i = \frac{N_i}{SD_i^2} \tag{4}$$

The heterogeneity between cohort means was calculated as:

$$Q = \sum W_i M_i^2 - \frac{(\sum W_i M_i)^2}{\sum W_i} \tag{5}$$

The expected value of  $Q$  is the degrees of freedom:

$$df = k - 1 \tag{6}$$

where  $k$  is the number of cohorts in the meta-analysis. Therefore, the estimated variance of the cohort mean distribution was calculated as:

$$T^2 = \frac{Q - df}{\sum W_i - \frac{\sum W_i^2}{\sum W_i}} \tag{7}$$

The percentage of total variability that reflects heterogeneity among cohorts was calculated as:

$$I^2 = \frac{Q - df}{Q} \times 100\% \tag{8}$$

The weight assigned to the  $i$ th cohort was updated as:

$$W_i^* = \frac{1}{\frac{SD_i^2}{N_i} + T^2} \tag{9}$$

The result of the random-effects meta-analysis was calculated as:

$$M^* = \frac{\sum W_i^* M_i}{\sum W_i^*} \tag{10}$$

The variance of  $M^*$  was estimated as:

$$V_{M^*} = \frac{1}{\sum W_i^*} \tag{11}$$

The standard error of  $M^*$  was calculated as:

$$SE_{M^*} = \sqrt{V_{M^*}} \tag{12}$$

This random-effects meta-analysis model exported a mean value  $M^*$ , its corresponding standard error value  $SE_{M^*}$ , and the heterogeneity score  $I^2$ .

**Mapping seed-to-whole-brain connectivity maps of functional connectome hubs.**

We modeled each hub seed region as a sphere with a 6-mm radius centered on the hub peak and computed Pearson's correlation coefficients between the seed region's preprocessed rsfMRI time series and the time series of all gray matter voxels. The time series of the seed region was computed by averaging the time series of all gray matter voxels in the seed sphere. These correlation coefficients were further transformed to Fisher's  $z$  for normality.

For each cohort, we performed a general linear model on these Fisher's  $z$  value maps to reduce age and sex effects. For each voxel, we constructed the general linear model as:

$$\text{Fisher's } z_i = \beta_0 + \beta_{\text{Age}} * (\text{Age}_i - \text{MeanAge}) + \beta_{\text{Sex}} * \text{Sex}_i + \varepsilon_i \tag{13}$$

Fisher's  $z_i$ ,  $\text{Age}_i$ ,  $\text{Sex}_i$ , and  $\varepsilon_i$  indicate the Fisher's  $z$ , age, sex, and residual of the  $i$ th individual, respectively.  $\text{MeanAge}$  indicates the mean age of that cohort.  $\beta_0$  indicates the mean Fisher's  $z$  value of that cohort. The general linear model exported a mean Fisher's  $z$  value map and its corresponding variance map for each cohort.

The mean and variance Fisher's  $z$  value maps of the 61 cohorts were submitted to the random-effects meta-analysis model<sup>51</sup> to address across-cohort heterogeneity of functional connections, resulting in a robust Fisher's  $z$  pattern and its corresponding SE map. We compared the Fisher's  $z$  value of each voxel with zero using a  $Z$  value and estimated effect size using Cohen's  $d$  metric<sup>51</sup>:

$$Z = \frac{\text{Fisher's } z - 0}{SE} \tag{14}$$

$$d = \frac{Z}{\sqrt{k}} \tag{15}$$

$k$  is the number of cohorts in the meta-analysis.

We performed 10,000 one-sided nonparametric permutation tests<sup>28</sup> to assign a  $p$  value to the observed  $Z$  value. For each iteration, after randomizing the spatial correspondence among cohorts' mean Fisher's  $z$  value maps (the spatial correspondence between a cohort's mean Fisher's  $z$  value map and its variance map was not changed), we repeated the computation procedure of the random-effects meta-analysis for each voxel and extracted the maximum  $Z$  value of all voxels to construct a null distribution. Then, we assigned a  $p$  value to each voxel by comparing the observed  $Z$  value to the null distribution.

Finally, we defined the most consistent functional connectivity map as brain regions with a  $p$  value less than 0.001 and cluster size greater than 200 mm<sup>3</sup> (Fig. 3). To illustrate the left 8Av hub's connectivity map, we also mapped its contralateral region the right 8Av region's connectivity map (Supplementary Fig. 5).

**Hierarchical clustering analysis on connectivity maps of functional connectome hubs.**

To address the effect of network size, we first divided the most consistent functional connectivity map of each hub into eight brain networks

mentioned above and represented the functional connectivity profile of a hub as the voxel percentage of each of the eight networks connected with it. Thus, we obtained an 8×35 connectivity matrix with each column representing the voxel percentage of each of the eight networks connected with a hub (Fig. 4a). Then, the 8×35 connectivity matrix was submitted to a hierarchical clustering model to obtain an agglomerative hierarchical cluster tree (Fig. 4a) that indicates the similarity of these functional connectivity profiles. We implemented hierarchical clustering model using the *linkage* function from MATLAB R2013a with default parameters.

### Identifying transcriptomic pattern underlying functional connectome hubs.

We trained machine learning classifiers based on XGBoost<sup>33</sup> and SVM to distinguish connectome hubs from non-hubs using transcriptomic features from the preprocessed AHBA dataset<sup>34</sup> (Fig. 5). The original AHBA dataset consists of microarray expression data of more than 20,000 genes in 3702 spatially distinct brain samples taken from six neurotypical adult donors<sup>78</sup>. Only two out of six donors were sampled from both hemispheres and the other four were sampled from only the left hemisphere. Because no statistically significant hemispheric difference was identified in the original AHBA dataset<sup>78</sup>, a prior study<sup>34</sup> provided a publicly available preprocessed AHBA dataset that includes 10,027 genes' transcriptomic data from 1285 left cortical samples. Preprocessing steps taken by the study<sup>34</sup> mainly include probe-to-gene re-annotation, intensity based data filtering, probe selection, accounting for individual variability, and gene filtering. Of the 1,285 samples, 382 were identified as hub samples and 776 as non-hub samples according to their MNI coordinates and the hub identification map in Fig. 1c. The remaining 127 samples were not included in our analysis because they were out of our gray matter mask. The brain samples used in our analysis are listed in Supplementary Data 9.

We built a supervised machine learning classifier based on XGBoost, a scalable tree boosting system with state-of-the-art resource efficiency and superior performance in many machine learning challenges<sup>33</sup>, to distinguish hub samples from non-hub samples using 10,027 genes' transcriptomic data from the preprocessed AHBA dataset. We used equal amounts of positive samples (hub samples) and negative samples (non-hub samples) in the classifier training procedure to ensure that the optimal classifier was unbiased toward any type of sample. To balance the time complexity and the prediction accuracy, we trained the classifier with 300 randomly selected hub samples and 300 randomly selected non-hub samples and tested it with the remaining 82 hub samples and 476 non-hub samples (Fig. 5a). Each gene's contribution to the optimal prediction model was determined after training the classifier. Based on previous experience<sup>79</sup>, we performed a 30-fold cross-validation procedure to identify the optimal number of model training iterations. The sensitivity, specificity, and accuracy rate of the XGBoost classifier and each gene's contribution to the classification results were stably estimated by repeating the randomly selecting training samples, cross-validation, and classifier training and testing procedures 1000 times. We implemented XGBoost using the XGBoost package<sup>33</sup> v1.2.0.1 in R 4.0.2 with following parameters: *nrounds* = 1500, *early\_stopping\_rounds* = 50, *eta* = 0.05, *objective* = "binary:logistic".

To exclude the XGBoost model's potential bias relating to mostly contributed key genes, we reproduced classification results using another machine learning model based on SVM (Fig. 5e). Instead of using all 10,027 genes' transcriptomic features, we only used genes with the greatest contributions to the XGBoost classifier to train the SVM classifier. If the key genes with the greatest contributions to the classification results are independent of the XGBoost model, the SVM classifier will achieve a comparable or higher accuracy rate than the XGBoost classifier because of the exclusion of redundant genes whose contribution to the classification results is negligible. In line with the XGBoost model, we used equal amounts of hub samples and non-hub samples in the classifier training procedure. For the easiest classification task, the SVM classifier was trained to distinguish all 382 hub samples from 382 non-hub samples with the highest rate to be correctly classified by the XGBoost classifier. For the most difficult classification task, the SVM classifier was trained to distinguish all 382 hub samples from 382 non-hub samples with the lowest rate to be correctly classified by the XGBoost classifier. To balance the time complexity and the prediction accuracy, we performed a 382-fold cross-validation procedure to obtain the optimal SVM classifier. We implemented SVM using the *svm* function from the *scikit-learn* package<sup>80</sup> v0.23.2 in Python 3.8.3 with default parameters.

To exclude the potential effect of spatial autocorrelation inherent to the transcriptomic data and the hub localization, we repeated the above described XGBoost and SVM classifiers training and testing procedures using surrogate hub identifications with the spatial autocorrelations being corrected using a generative model<sup>81</sup>. As shown in Supplementary Fig. 8a, we first constructed a surrogate Z value map with the spatial autocorrelation being corrected using a generative model<sup>81</sup> based on the unthresholded Z value map corresponding to the hub identification map in Fig. 1c. Then, for the 1158 AHBA brain samples within our gray matter mask, we assigned the 382 samples with the highest surrogate Z values as hub samples and the 776 samples with the lowest surrogate Z values as non-hub samples. For the XGBoost classifier, we trained the classifier with 300 randomly selected surrogate hub samples and 300 randomly selected surrogate non-hub samples using 10,027 genes' transcriptomic data from the preprocessed AHBA

dataset and tested it with the remaining 82 surrogate hub samples and 476 surrogate non-hub samples. We performed a 30-fold cross-validation procedure to identify the optimal number of model training iterations. We implemented XGBoost using the XGBoost package<sup>33</sup> v1.2.0.1 in R 4.0.2 with following parameters: *nrounds* = 1,500, *early\_stopping\_rounds* = 50, *eta* = 0.05, *objective* = "binary:logistic". For the SVM classifier, we built a supervised SVM classifier through a 382-fold cross-validation procedure to distinguish all 382 surrogate hub samples from 382 randomly selected surrogate non-hub samples using the transcriptomic data of the top 150 genes listed in Supplementary Data 1. We implemented SVM using the *svm* function from the *scikit-learn* package<sup>80</sup> v0.23.2 in Python 3.8.3 with default parameters. Finally, we repeated the surrogate hub identification generating, XGBoost classifier training and testing, and SVM classifier training procedures 1000 times.

The top 150 key genes (Supplementary Data 1) mostly contributed to the classification results were submitted to GO enrichment analyses using GOrilla<sup>35</sup> and DAVID<sup>36,37</sup> v6.8. We conducted two GO enrichment analyses using GOrilla. The first analysis used the 150 key genes as the target list and all 10,027 genes as the background list. The second analysis used the ranked 10,027 genes according to their contributions to the XGBoost classifier. Of note, we performed GO enrichment analyses for all three ontology categories: biological process, molecular function, and cellular component. However, only analysis for biological process yielded significant GO terms. We repeated GO enrichment analysis for biological process using DAVID with the 150 key genes as the target list and all 10,027 genes as the background list. In addition, we also performed GO enrichment analysis for disease association using DAVID with the 150 key genes as the target list and all 10,027 genes as the background list.

Based on GO enrichment analysis results, we tested transcription level differences for gene sets involved in key neurodevelopmental processes<sup>38</sup> (Supplementary Data 6) and main neuronal metabolic pathways<sup>39</sup> (oxidative phosphorylation<sup>40</sup> and aerobic glycolysis<sup>41</sup>, Supplementary Data 7) between connectome hubs and non-hubs through one-sided Wilcoxon rank-sum test. In line with prior studies<sup>38,41</sup>, we used the first principal component of each gene set's transcription level to plot and to perform the statistical analysis (Fig. 6a). For illustration purposes, we normalized the first principal component of each gene set's transcription level respect to its minimum and maximum values across all brain samples to the range 0–1.

To explore developmental details, we inspected the developmental trajectory of transcription level of the above gene sets (Supplementary Data 6 and 7) in hub and non-hub regions respectively using the BrainSpan Atlas<sup>42</sup>. The normalized BrainSpan Atlas was generated using 524 brain samples from 42 donors aged from eight post-conceptual weeks to 40 postnatal years, including 52,376 genes' transcriptomic data from 11 neocortical areas and five additional regions of the human brain. The brain regions used in our analysis are listed in Supplementary Data 10. We plotted the developmental trajectory using locally weighted regression by smoothing the first principal component of each gene set's transcription level against  $\log_2[\text{post-conceptual days}]$  as in a prior study<sup>38</sup> (Fig. 6b). For most developmental periods, there are only no more than five hub brain samples and 10 non-hub brain samples at a specific age (Supplementary Fig. 11). Such small sample size makes it practically impossible to determine the statistical significance level of difference in transcription level between hub and non-hub regions at a specific age. We compared the magnitude of differences in developmental trajectory between hub and non-hub regions to the median absolute deviation of transcription level across brain regions at a specific age (Fig. 6c). The magnitude of differences in developmental trajectory exceeding the median absolute deviation indicates a trend of greater difference in transcription level between hub and non-hub regions than expected at a specific age. Of note, considering apparent transcriptomic differences compared to the neocortex<sup>38</sup>, we excluded the striatum, mediodorsal nucleus of the thalamus, and cerebellar cortex in the developmental trajectory analysis but not the amygdala and hippocampus whose developmental trajectories of transcription level are more similar to those of the neocortex than to those of other subcortical structures<sup>38</sup>. Analysis using only neocortical regions revealed similar results (Supplementary Fig. 9).

**Contextualizing functional connectome hubs' transcriptomic pattern using established neuroimaging patterns.** We assessed the neural relevance of the above identified transcriptomic pattern underlying functional connectome hubs by contextualizing it relative to established neuroimaging patterns of neurotransmitter<sup>46</sup>, cortical fiber length<sup>47</sup>, brain metabolism<sup>48</sup>, and cortical thickness atrophy in neuropsychiatric disorders<sup>50</sup>.

The JuSpace toolbox<sup>46</sup> provided 15 neurotransmitter receptor and transporter density maps in the MNI volume space. For each of the 15 density maps, we tested difference in density between hub and non-hub voxels through one-sided Wilcoxon rank-sum test (Fig. 7a). For illustration purposes, we normalized the density value respect to its median and median absolute deviation across voxels.

The cortical fiber length profiling dataset<sup>47</sup> provided fiber number data across different length bins in a standard brain surface space. We resampled the identified hub distribution mask in Fig. 1c from the MNI volume space to the standard brain surface space provided by the dataset<sup>47</sup> and tested difference in fiber number between hub and non-hub vertices for each length bin through one-sided Wilcoxon

rank-sum test (Fig. 7b). For illustration purposes, we normalized the fiber number value respect to its mean and standard deviation across voxels.

The brain metabolism dataset provided by the positron emission tomography study<sup>48</sup> was assigned to 82 Brodmann areas in a standard brain surface space and seven subcortical structures in the MNI volume space. We first resampled the identified hub distribution mask in Fig. 1c from the MNI volume space to the standard brain surface space provided by the dataset<sup>48</sup> and identified Brodmann areas with more than 50% vertices within the hub distribution mask as hub regions. Then, we identified subcortical structures with more than 50% voxels within the hub distribution mask as hub regions. After that, we examined differences between hub and non-hub regions in metabolic measurements of blood supply (the cerebral blood flow), oxidative phosphorylation (the cerebral metabolic rate for oxygen), and aerobic glycolysis (the glycolytic index) through one-sided Wilcoxon rank-sum test (Fig. 7c).

The Cohen's  $d$  value of cortical thickness atrophy in neuropsychiatric disorders was assigned to 68 cortical areas in a standard brain surface space<sup>50</sup>. We first resampled the unthresholded Cohen's  $d$  map of connectome hub in Fig. 1c from the MNI volume space to the standard brain surface space provided by the dataset<sup>50</sup> and computed the Cohen's  $d$  value for each of the 68 cortical areas by averaging Cohen's  $d$  value across vertices within each cortical area. Then, we computed Pearson's correlation coefficient between the Cohen's  $d$  value of connectome hub and the Cohen's  $d$  value of cortical thickness atrophy across 68 cortical areas for each of the eight disorders (Fig. 7d). To reduce the potential effects of development on our results, we used cortical thickness atrophy data from adults for the attention deficit hyperactivity disorder, bipolar disorder, major depressive disorder, and obsessive-compulsive disorder.

**Statistics and reproducibility.** We performed statistical analyses using MATLAB R2013a. The statistical significances of brain clusters in Figs. 1c and 3 and Supplementary Figs. 2c, 3c, 5, and 10a were determined by comparing the observed  $Z$  values with their corresponding null distributions constructed by above mentioned 10,000 one-sided nonparametric permutation tests<sup>28</sup>. To determine the statistical significances of one-sided Wilcoxon rank-sum tests in Figs. 6a, 7a–c and Supplementary Fig. 10e, we constructed 1000 surrogate hub identification maps with the spatial autocorrelations being corrected using a generative model<sup>81</sup> and repeated calculating rank-sum statistics using these surrogate hub identification maps to construct a null distribution. Then,  $p$  values of these rank-sum statistics were determined by comparing the observed values with their corresponding null distributions and were Bonferroni-corrected. Surrogate hub identification maps for Figs. 6a and 7a–c were constructed based on the hub identification map in Fig. 1c. Surrogate hub identification maps for Supplementary Fig. 10e were constructed based on the hub identification map in Supplementary Fig. 10a. To determine the statistical significances of Pearson's correlation coefficients in Fig. 7d, we constructed 1000 surrogate maps of the unthresholded Cohen's  $d$  map in Fig. 1c with the spatial autocorrelations being corrected using a generative model<sup>81</sup> and repeated calculating Pearson's correlation coefficients using these surrogate Cohen's  $d$  maps to construct a null distribution. Then,  $p$  values of these Pearson's correlation coefficients were determined by comparing the observed values with their corresponding null distributions.

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

## Data availability

The MRI data of the first 60 cohorts listed in Supplementary Data 8 are available at the International Neuroimaging Data-sharing Initiative ([http://fcon\\_1000.projects.nitrc.org](http://fcon_1000.projects.nitrc.org)), Brain Genomics Superstruct Project<sup>82</sup> (<https://doi.org/10.7910/DVN/25833>), Human Connectome Project (<https://www.humanconnectome.org>), MPI-Leipzig Mind-Brain-Body Project (<https://openneuro.org/datasets/ds000221>), and Age-ility Project (<https://www.nitrc.org/projects/age-ility>). The MRI data of the PKU cohort are under active use by the reporting laboratory and will be available from the corresponding author upon reasonable request. The preprocessed AHBA dataset is available at <https://doi.org/10.6084/m9.figshare.6852911>. The normalized BrainSpan Atlas dataset is available at <http://brainspan.org/static/download.html>. The neurotransmitter receptor and transporter density maps provided by the JuSpace toolbox<sup>46</sup> are available at <https://github.com/juryxy/JuSpace>. The fiber length profiling dataset<sup>47</sup> is available at <https://balsa.wustl.edu/study/1K3l>. The cortical thickness atrophy dataset provided by the ENIGMA Toolbox<sup>50</sup> is available at <https://github.com/MICA-MNI/ENIGMA>. Numerical source data to reproduce all figure panels is available at <https://doi.org/10.6084/m9.figshare.21194128>.

## Code availability

The code to reproduce the results and visualizations of this manuscript is available on Zenodo<sup>83</sup>. Software packages used in this manuscript include MATLAB R2013a (<https://www.mathworks.com/products/matlab.html>), SPM12 v6470 (<https://www.fli.ion.ucl.ac.uk/spm/software/spm12>), GREYNA<sup>72</sup> v2.0.0 (<https://www.nitrc.org/projects/gretna>),

Connectome Workbench v1.4.2 (<https://www.humanconnectome.org/software/connectome-workbench>), cifti-matlab v2 (<https://github.com/Washington-University/cifti-matlab>), R 4.0.2 (<https://www.r-project.org>), XGBoost package<sup>33</sup> v1.2.0.1 (<https://cran.r-project.org/web/packages/xgboost>), Python 3.8.3 (<https://www.python.org>), and scikit-learn package<sup>80</sup> v0.23.2 (<https://scikit-learn.org>). Online analysis tools used in this manuscript include GOrilla<sup>35</sup> (<http://cbl-gorilla.cs.technion.ac.il>) and DAVID<sup>36,37</sup> v6.8 (<https://david.ncifcrf.gov>).

Received: 28 April 2022; Accepted: 23 September 2022;  
Published online: 04 October 2022

## References

- van den Heuvel, M. P. & Sporns, O. Network hubs in the human brain. *Trends Cogn. Sci.* **17**, 683–696 (2013).
- Bullmore, E. & Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349 (2012).
- Liu, J. et al. Intrinsic brain hub connectivity underlies individual differences in spatial working memory. *Cereb. Cortex* **27**, 5496–5508 (2016).
- Xu, Y., Lin, Q., Han, Z., He, Y. & Bi, Y. Intrinsic functional network architecture of human semantic processing: Modules and hubs. *NeuroImage* **132**, 542–555 (2016).
- van den Heuvel, M. P. et al. Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry* **70**, 783–792 (2013).
- Crossley, N. A. et al. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382–2395 (2014).
- Buckner, R. L. et al. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *J. Neurosci.* **29**, 1860–1873 (2009).
- Dai, Z. et al. Identifying and mapping connectivity patterns of brain network hubs in Alzheimer's disease. *Cereb. Cortex* **25**, 3723–3742 (2014).
- Wang, J. et al. Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. *Biol. Psychiatry* **73**, 472–481 (2013).
- Wang, L. et al. The effects of antidepressant treatment on resting-state functional brain networks in patients with major depressive disorder. *Hum. Brain Mapp.* **36**, 768–778 (2015).
- Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nat. Rev. Neurosci.* **16**, 159–172 (2015).
- Gong, Q. & He, Y. Depression, neuroimaging and connectomics: a selective overview. *Biol. Psychiatry* **77**, 223–235 (2015).
- Achard, S., Salvador, R., Whitcher, B., Suckling, J. & Bullmore, E. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* **26**, 63–72 (2006).
- Liao, X.-H. et al. Functional brain hubs and their test-retest reliability: a multiband resting-state functional MRI study. *NeuroImage* **83**, 969–982 (2013).
- Cole, M. W., Pathak, S. & Schneider, W. Identifying the brain's most globally connected regions. *NeuroImage* **49**, 3132–3148 (2010).
- Tomasi, D. & Volkow, N. D. Functional connectivity density mapping. *Proc. Natl Acad. Sci. USA* **107**, 9885–9890 (2010).
- Fransson, P., Aden, U., Blennow, M. & Lagercrantz, H. The functional architecture of the infant brain as revealed by resting-state fMRI. *Cereb. Cortex* **21**, 145–154 (2011).
- Tomasi, D. & Volkow, N. D. Association between functional connectivity hubs and brain networks. *Cereb. Cortex* **21**, 2003–2013 (2011).
- de Pasquale, F. et al. The connectivity of functional cores reveals different degrees of segregation and integration in the brain at rest. *NeuroImage* **69**, 51–61 (2013).
- Glahn, D. C. et al. Genetic control over the resting brain. *Proc. Natl Acad. Sci. USA* **107**, 1223–1228 (2010).
- Fornito, A. et al. Genetic influences on cost-efficient organization of human cortical functional networks. *J. Neurosci.* **31**, 3261–3270 (2011).
- Wiggins, J. L. et al. The impact of serotonin transporter (5-HTTLPR) genotype on the development of resting-state functional connectivity in children and adolescents: a preliminary report. *NeuroImage* **59**, 2760–2770 (2012).
- Gordon, E. M., Stollstorff, M., Devaney, J. M., Bean, S. & Vaidya, C. J. Effect of dopamine transporter genotype on intrinsic functional connectivity depends on cognitive state. *Cereb. Cortex* **22**, 2182–2196 (2012).
- Richiardi, J. et al. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).
- Diez, I. & Sepulcre, J. Neurogenetic profiles delineate large-scale connectivity dynamics of the human brain. *Nat. Commun.* **9**, 3876 (2018).
- Vertes, P. E. et al. Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks. *Philos. Trans. R. Soc. B* **371**, 20150362 (2016).

27. Fornito, A., Arnatkevičiūtė, A. & Fulcher, B. D. Bridging the gap between connectome and transcriptome. *Trends Cogn. Sci.* **23**, 34–50 (2019).
28. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* **15**, 1–25 (2002).
29. Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
30. Tzourio-Mazoyer, N. et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–289 (2002).
31. Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N. & Petersen, S. E. Evidence for hubs in human functional brain networks. *Neuron* **79**, 798–813 (2013).
32. Gordon, E. M. et al. Three distinct sets of connector hubs integrate human brain function. *Cell Rep.* **24**, 1687–1695 (2018).
33. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794 (2016).
34. Arnatkevičiūtė, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *NeuroImage* **189**, 353–367 (2019).
35. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinforma.* **10**, 48 (2009).
36. Huang, D., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
37. Huang, D., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–457 (2009).
38. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
39. Harris, J. J., Jolivet, R. & Attwell, D. Synaptic energy use and supply. *Neuron* **75**, 762–777 (2012).
40. Arroyo, J. D. et al. A genome-wide CRISPR death screen identifies genes essential for oxidative phosphorylation. *Cell Metab.* **24**, 875–885 (2016).
41. Goyal, M. S., Hawrylycz, M., Miller, J. A., Snyder, A. Z. & Raichle, M. E. Aerobic glycolysis in the human brain is associated with development and neonatal gene expression. *Cell Metab.* **19**, 49–57 (2014).
42. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
43. Sherwood, C. C. & Gómez-Robles, A. Brain plasticity and human evolution. *Annu. Rev. Anthropol.* **46**, 399–419 (2017).
44. Huttenlocher, P. R. & Dabholkar, A. S. Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178 (1997).
45. Nicholls J. G., Martin A. R., Fuchs P. A., Brown D. A., Diamond M. E., Weisblat D. A. *From Neuron to Brain*, Fifth Edition. 273–298 Sinauer Associates, (2012).
46. Dukat, J. et al. JuSpace: A tool for spatial correlation analyses of magnetic resonance imaging data with nuclear imaging derived neurotransmitter maps. *Hum. Brain Mapp.* **42**, 555–566 (2021).
47. Bajada, C. J., Schreiber, J. & Caspers, S. Fiber length profiling: a novel approach to structural brain organization. *NeuroImage* **186**, 164–173 (2019).
48. Vaishnavi, S. N. et al. Regional aerobic glycolysis in the human brain. *Proc. Natl Acad. Sci. USA* **107**, 17757–17762 (2010).
49. Liang, X., Zou, Q., He, Y. & Yang, Y. Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain. *Proc. Natl Acad. Sci. USA* **110**, 1929–1934 (2013).
50. Lariviere, S. et al. The ENIGMA Toolbox: multiscale neural contextualization of multisite neuroimaging datasets. *Nat. Methods* **18**, 698–700 (2021).
51. Cumming, G. *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. 207–230, 281–319 Routledge, (2013).
52. Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L. & Raichle, M. E. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl Acad. Sci. USA* **103**, 10046–10051 (2006).
53. Marek, S. & Dosenbach, N. U. F. The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues Clin. Neurosci.* **20**, 133–140 (2018).
54. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
55. Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H. & Johnson, K. A. Stepwise connectivity of the modal cortex reveals the multimodal organization of the human brain. *J. Neurosci.* **32**, 10649–10661 (2012).
56. Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. The brain’s default network anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* **1124**, 1–38 (2008).
57. Braga, R. M. & Buckner, R. L. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron* **95**, 457–471.e455 (2017).
58. Menon, V. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn. Sci.* **15**, 483–506 (2011).
59. Xie, Y. et al. Alterations in connectome dynamics in autism spectrum disorder: a harmonized mega- and meta-analysis study using the autism brain imaging data exchange dataset. *Biol. Psychiatry* **91**, 945–955 (2022).
60. Changeux, J.-P. & Danchin, A. Selective stabilisation of developing synapses as a mechanism for the specification of neuronal networks. *Nature* **264**, 705–712 (1976).
61. Miller, D. J. et al. Prolonged myelination in human neocortical evolution. *Proc. Natl Acad. Sci. USA* **109**, 16480–16485 (2012).
62. Lebel, C., Gee, M., Camicioli, R., Wieler, M., Martin, W. & Beaulieu, C. Diffusion tensor imaging of white matter tract evolution over the lifespan. *NeuroImage* **60**, 340–352 (2012).
63. Petanjek, Z. et al. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl Acad. Sci. USA* **108**, 13281–13286 (2011).
64. Paredes, M. F. et al. Extensive migration of young neurons into the infant human frontal lobe. *Science* **354**, aaf0703 (2016).
65. Kwan, K. Y., Sestan, N. & Anton, E. S. Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. *Development* **139**, 1535–1546 (2012).
66. Wei, Y. et al. Statistical testing in transcriptomic-neuroimaging studies: a how-to and evaluation of methods assessing spatial and gene specificity. *Hum. Brain Mapp.* **43**, 885–901 (2022).
67. Fulcher, B. D., Arnatkevičiūtė, A. & Fornito, A. Overcoming false-positive gene-category enrichment in the analysis of spatially resolved transcriptomic brain atlas data. *Nat. Commun.* **12**, 2669 (2021).
68. Helmer, M. et al. On stability of Canonical Correlation Analysis and Partial Least Squares with application to brain-behavior associations. *bioRxiv*, <https://doi.org/10.1101/2020.08.25.265546> (2020).
69. Ciric, R. et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage* **154**, 174–187 (2017).
70. Fornito, A., Zalesky, A. & Bullmore, E. T. Network scaling effects in graph analytic studies of human resting-state fMRI data. *Front. Syst. Neurosci.* **4**, 22 (2010).
71. 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).
72. Wang, J., Wang, X., Xia, M., Liao, X., Evans, A. & He, Y. GRETNA: a graph theoretical network analysis toolbox for imaging connectomics. *Front. Hum. Neurosci.* **9**, 386 (2015).
73. 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).
74. Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).
75. Schwarz, A. J. & McGonigle, J. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *NeuroImage* **55**, 1132–1146 (2011).
76. Power, J. D. et al. Functional network organization of the human brain. *Neuron* **72**, 665–678 (2011).
77. Eickhoff, S. B. et al. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* **30**, 2907–2926 (2009).
78. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
79. Kaufmann, T. et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat. Neurosci.* **22**, 1617–1623 (2019).
80. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
81. Burt, J. B., Helmer, M., Shinn, M., Anticevic, A. & Murray, J. D. Generative modeling of brain maps with spatial autocorrelation. *NeuroImage* **220**, 117038 (2020).
82. Holmes, A. J. et al. Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. *Sci. Data* **2**, 150031 (2015).
83. Xu, Z. et al. Functional connectome hubs. *Zenodo*, <https://doi.org/10.5281/zenodo.7106831> (2022).



## Acknowledgements

We thank Drs. Huali Wang and Xiaodan Chen for data acquisition of the PKU cohort and Drs. Qiushi Wang and Nan Zhang for valuable advice on MRI data quality controls. This work was supported by the National Natural Science Foundation of China [82021004, 31830034, and 81620108016 to Y.H., 82071998 and 81671767 to M.X., 81971690 to X.L., 81801783 to T.Z.], Changjiang Scholar Professorship Award [T2015027 to Y.H.], National Key Research and Development Project [2018YFA0701402 to Y.H.], Beijing Nova Program [Z191100001119023 to M.X.], and Fundamental Research Funds for the Central Universities [2020NTST29 to M.X.].

## Author contributions

Conceptualization: Z.X., Y.H.; Methodology: Z.X., Y.H., M.X., X.W., X.L., T.Z.; Investigation: Z.X.; Visualization: Z.X.; Supervision: Y.H.; Writing—original draft: Z.X., Y.H.; Writing—review & editing: Y.H., Z.X., M.X., X.L., T.Z., X.W.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-04028-x>.

**Correspondence** and requests for materials should be addressed to Yong He.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: George Inglis. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022