nature communications

Article

https://doi.org/10.1038/s41467-026-71923-w

Integrating optogenetic fMRI and spatial transcriptomics to reveal circuit-specific gene signatures in fronto- and hippo-thalamic networks

Received: 26 March 2025

Accepted: 31 March 2026

Published online: 18 April 2026

Check for updates

Changjiang Zhang[1,2,7], Yijuan Zou[2,7], Liangchen Zhuo[1,3,7], Xiya Liu[1,3,7], Suijuan Zhong[4,7], Chonghai Yin[2], Mayuqing Li[4], Mengdi Wang[1], Wei Wang[1], Xin Zhou[4], Bo Zeng[2], Changsheng Dong[1,3], Le Sun[5] ✉, Zheng Wang[6] ✉, Ang Li[1,3] ✉, Qian Wu[4] ✉ & Xiaoqun Wang[1,2,4] ✉

Elucidating how microscale molecular architecture shapes macroscale brain network organization requires integrative frameworks that link transcriptional profiles, structural connectivity, and functional dynamics across scales. However, existing evidence largely relies on cross-modal data acquired from different individuals, limiting causal and mechanistic inference. To overcome this, we performed concurrent awake optogenetic functional MRI (opto-fMRI) and spatial transcriptomics (opto-ST) within the same mice, enabling coupled analysis of neural activity and gene expression in fronto-thalamic and hippo-thalamic networks. Optogenetic stimulation of the medial prefrontal cortex (mPFC) or subiculum (SUB) evoked distinct thalamic BOLD response patterns. Multivariate spatial modeling revealed that energy metabolism and core biochemical processes underlie these GLM-derived activation patterns—a finding further supported by data-driven PCA, which identified enrichments in similar biological processes. These results suggest that conventional linear models predominantly capture linear gene–process relationships, highlighting the need for nonlinear computational approaches in fMRI analysis. Overall, our multi-omics framework reveals that genes regulating metabolic pathways underlie region-specific neurovascular coupling, thereby bridging local transcriptomics and large-scale functional imaging signatures.

The advent of spatially resolved transcriptomic resources, particularly the Allen Human Brain Atlas (AHBA), has catalyzed a paradigm shift in systems neuroscience by enabling multiscale spatial covariance analyses that integrate macroscale neuroimaging phenotypes with microscale transcriptomic architectures[1–3]. This methodological convergence has established *imaging transcriptomics* as a transformative discipline that deciphers molecular determinants of neuroimaging biomarkers through advanced spatial registration algorithms and multivariate spatial statistics[4–7]. Such approaches have provided unprecedented mechanistic insights into the neuromolecular foundations of functional network organization. For instance, early research revealed that a small subset of genes, particularly enriched in human supra-granular layers can explain corticocortical connectivity[8], while conversely, brain functional connectivity architecture gradients can also largely predict ex vivo cell type distributions[9]. Furthermore, researchers have linked brain functional abnormalities in numerous

A full list of affiliations appears at the end of the paper. ✉e-mail: sunle@ibp.ac.cn; zheng.wang@pku.edu.cn; al@ibp.ac.cn; qianwu@bnu.edu.cn; xiaoqunwang@ibp.ac.cn

brain disorders—including schizophrenia, depression, autism, and Alzheimer's disease—to spatially-informed transcriptomic templates, attempting to elucidate the molecular mechanisms behind macroscale brain circuit dysfunction[10–18].

Despite these advances, a key limitation persists across previous studies: data from different scales typically originate from different samples, with spatial analyses often matching functional imaging from living subjects against transcriptomic templates derived from post-mortem tissue of healthy donors[3]. This limitation largely stems from the expensive and technically challenging nature of acquiring spatially resolved transcriptomic data. Such fragmented comparisons significantly constrain the field's ability to establish causal molecular mechanisms and develop effective cross-scale computational models, with validation largely infeasible or only limited to patients with well-defined genetic disorders[19].

To overcome this fundamental limitation, we developed an integrative multimodal approach in awake mice that captures both brain-wide functional responses and region-specific molecular changes within the same subjects. As a critical proof of concept, we focused on the thalamus—a central hub that integrates diverse cortical and subcortical inputs—to investigate how distinct large-scale inputs shape both neurovascular and transcriptomic responses in a shared downstream target. The thalamus is involved in a variety of brain functions, including sensory processing, attention, decision-making, and memory[20,21]. Traditionally, it has been believed that the functional diversity of the thalamus is determined by the organization of its different nuclei[20,22]. The input types that innervate thalamic neurons are another feature that differentiates thalamic functions[20,23]. We targeted two major thalamic inputs with dissociable yet overlapping functions: the fronto-thalamic pathway (implicated in cognitive control and motivation) and the hippo-thalamic pathway (central to memory processing)[24–32]. Despite their distinct primary functions, both pathways converge on the thalamus and are implicated in shared processes such as arousal and emotional regulation[33,34]. We therefore hypothesized that optogenetic activation of these inputs would elicit distinct but partially overlapping signatures in thalamic gene expression, reflecting their unique and shared functional roles.

Specifically, we implemented a multimodal approach integrating optogenetic fMRI (opto-fMRI) with spatial transcriptomics (opto-ST), Fos targeted recombination in active populations(opto-FosTRAP), and anatomical tracing to systematically dissect medial prefrontal cortex (mPFC)- and subiculum (SUB)- driven thalamic responses in awake mice. We hypothesized that these two inputs would elicit distinct BOLD response patterns and gene expression changes, illuminating shared or unique molecular processes involved in cognition, arousal, and emotion.

Through a series of image processing and registration approaches, we integrated optogenetically-evoked BOLD responses, neuronal activation, anatomical projections, and transcriptomic data within a common standardized space, enabling more causally multi-scale correlation analyses. Using partial least squares (PLS) modeling, we show that SUB projections strongly predict thalamic BOLD activation, whereas mPFC projections correlate more weakly, consistent with higher-level circuits having more diffuse or multi-tiered connectivity. Our study highlights that energy metabolism and basic biochemical processes emerge as key correlates of BOLD signal generation, suggesting a tight coupling between neuronal activation and metabolic gene regulation. This cross-scale approach provides a path for exploring how molecular signals shape large-scale network activity and paves the way for integrating multimodal datasets across levels of brain organization.

Results

Circuit-specific BOLD activation in awake mice

We combined optogenetics with fMRI in awake mice (Fig. 1a–d), FosTRAP (Fig. 1e), and anterograde axon tracing (Fig. 1f) in the mPFC of the SUB to determine whether functional activation is directly governed by neural/glial cell activation caused by corresponding anatomical projections from mPFC/SUB neurons. After rigorous noise regression of the whole-brain BOLD signals collected in the awake state (Fig. S2), optogenetic activation of either region induced extensive BOLD responses with distinct spatial patterns across the brain (Fig. 1g; Fig. S1–S6). No statistically significant voxel-level differences in activation patterns between male and female mice after correction for multiple comparisons (Fig. S6). Both topological data analysis (TDA) and whole-brain correlation analysis demonstrated that even individual animals showed robust and consistent activation patterns within each group - TDA visually segregated activation patterns into two distinct clusters corresponding to mPFC and SUB stimulation (Fig. 1l), while correlation analysis confirmed the reproducibility of these region-specific signatures (Fig. S4). Within the thalamus, both regions elicited significant activation with visually distinct spatial organization: mPFC stimulation tended to show a medial-to-lateral gradient, while SUB stimulation showed a dorsolateral-to-ventromedial pattern.

To understand how anatomical projections and cellular activation contribute to the BOLD activations, we performed anterograde tracing from mPFC or SUB, along with opto-FosTRAP labeling of Fos-expressing cells (Fig. 1e, f). Spatial correlation analyses across the whole brain revealed moderate alignment between projection maps and BOLD activation, whereas correlation with FosTRAP activation was lower (Fig. 1j, k). The observed spatial discrepancy (mPFC: $r = 0.39$, $p = 1.1*10^{-5}$; SUB: $r = 0.22$, $p = 0.016$; Fig. 1j, k) between fMRI and FosTRAP patterns suggests that the fMRI signal may be influenced by additional factors beyond the strong primary inputs that are effectively captured by FosTRAP. This interpretation is grounded in the fundamental differences of the two methods: FosTRAP relies on activity-dependent c-Fos induction, which is known to have a high threshold and can be regulated by stimulation parameters[35–37] and intracellular molecular gating mechanisms[38–40]. Therefore, the comprehensive fMRI signal likely aggregates both supra-threshold activity (captured by FosTRAP) and sub-threshold synaptic activity from polysynaptic, secondary projections, which may contribute to the hemodynamic response without reliably triggering c-Fos expression.

Coregistration of opto-fMRI-ST data in the thalamus

To link gene expression changes with BOLD activations, we focused on rostral and caudal thalamic slices for high-resolution spatial transcriptomics (opto-ST) (Fig. 2a). A total of 32 samples from 10 individuals in four groups with technical replicates were included to ensure data stability. Each animal's ST data were coregistered to a standardized coordinate space (CCFv3) (Fig. 2d, e). The 40-minute (~90-minute total from stimulation onset) time point (Fig. 2d) was selected to align with reported peaks in early stimulus-induced transcription[38,41–43]. The rostral mPFC group, rostral SUB group, and analogous caudal groups showed consistent gene expression patterns, underscoring data reliability (Fig. S8).

Among the thalamic subregions, the mPFC group exhibited prominent activation primarily in the bilateral ventral group of the dorsal thalamus (VENT), subparafascicular nucleus (SPF), subparafascicular area (SPA), peripeduncular nucleus (PP), lateral group of the dorsal thalamus (LAT), anterior group of the dorsal thalamus (ATN), medial group of the dorsal thalamus (MED), midline group of the dorsal thalamus (MTN), intralaminar nuclei of the dorsal thalamus (ILM), reticular nucleus of the thalamus (RT), and epithalamus (EPI) (Fig. 2a–c). Conversely, the SUB group showed significant activation in the ipsilateral VENT, LAT, ATN, MED, MTN, ILM, RT and EPI (Fig. 2a–c). Furthermore, we found that different subregions of the thalamus exhibited distinct response dynamics (Fig. 2c). Overall, the mPFC group exhibited comparable peak values and consistent temporal dynamics between the ipsilateral and contralateral thalamic subregions. In contrast, the SUB group showed marked hemispheric

https://doi.org/10.1038/s41467-026-71923-w

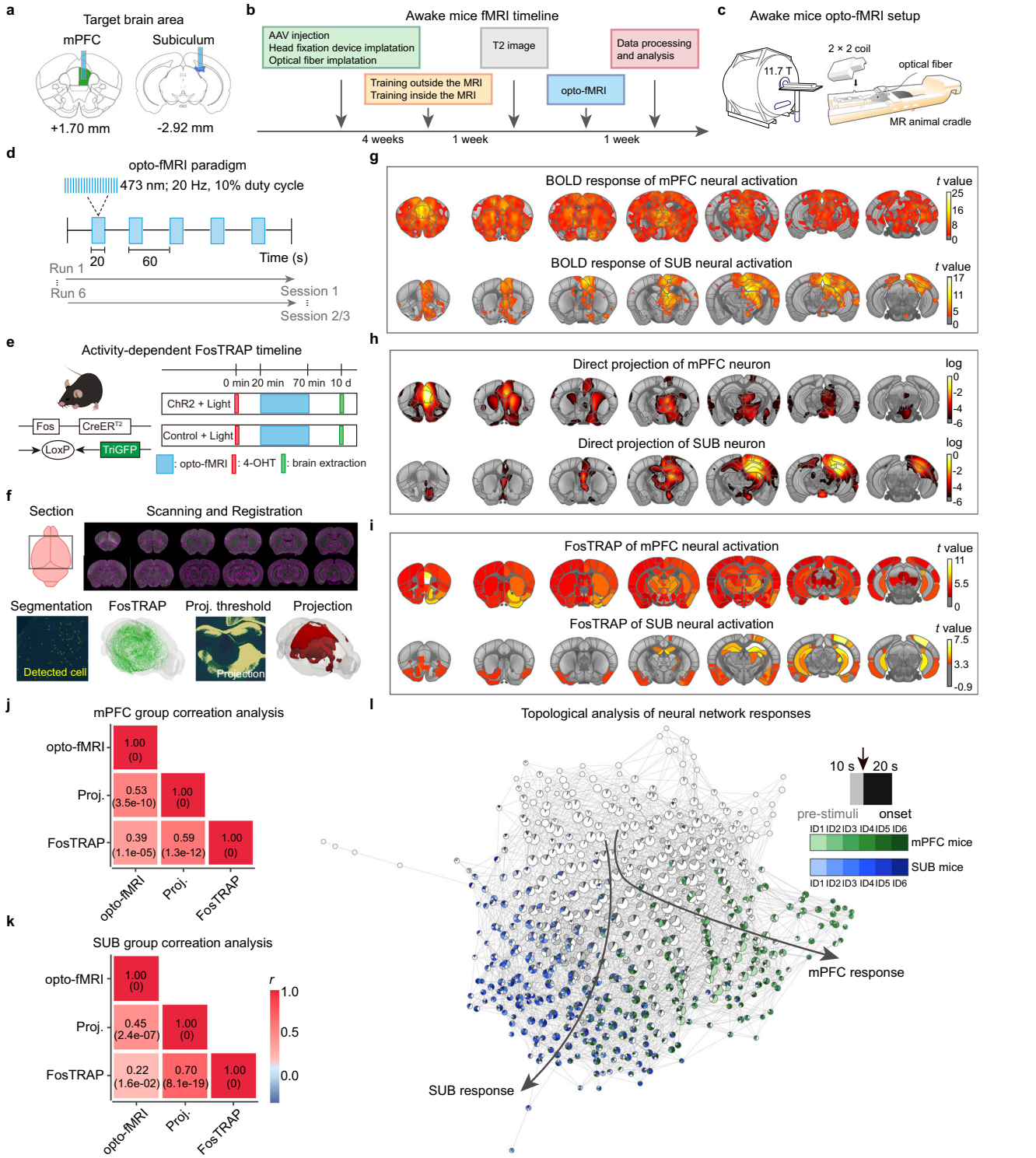


Fig. 1 | Anatomic projection and cFos-activation partially contribute to mPFC and SUB optogenetic BOLD responses at the whole-brain level. a Target brain areas: mPFC and SUB. **b** Awake mouse opto-fMRI timeline. **c** Awake mouse opto-fMRI setup. **d** Optogenetic stimulation paradigm. **e** Activity-dependent FosTRAP timeline. **f** Registration and cell segmentation of the projection. **g** Whole-brain BOLD responses of optogenetic activation of mPFC and SUB neurons. Voxel-wise two-sample tests, FDR-corrected p values < 0.05, cluster size > 10. **h** Whole-brain direct projections of mPFC and SUB neurons. Voxel-wise two-sample tests, FDR-corrected p values < 0.05, cluster size > 10. **i** Whole-brain FosTRAP responses of mPFC and SUB neural activation. Region-wise two-sample tests and FDR-corrected p values < 0.05 were used to determine activated regions. Pearson correlation of three-modal data at the whole-brain level for the mPFC (**j**) and SUB (**k**) groups, using two-sided tests. **l** Topological analysis (Mapper algorithm) of neural network responses is depicted. The gray areas indicate the baseline state 10 s prestimulation, and the colored areas mark the stimulation onset. Nodes and connectomes show neural responses across mice during optogenetic stimulation. Each node represents a cluster of neural activity at a specific time, with the size indicating the number of time points and the color proportion showing the percentage of mice present. See also Figs. S1–6.

https://doi.org/10.1038/s41467-026-71923-w

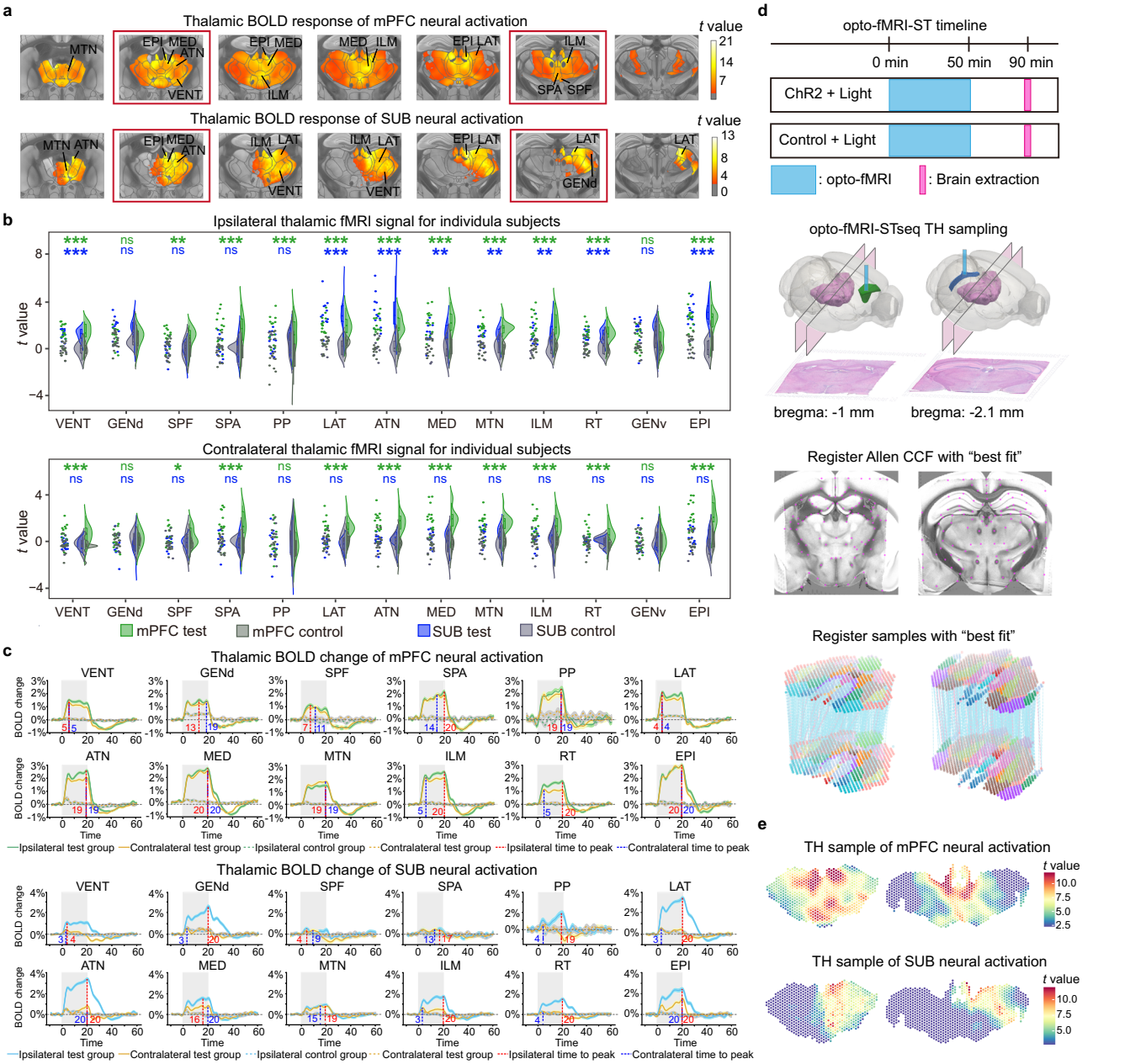


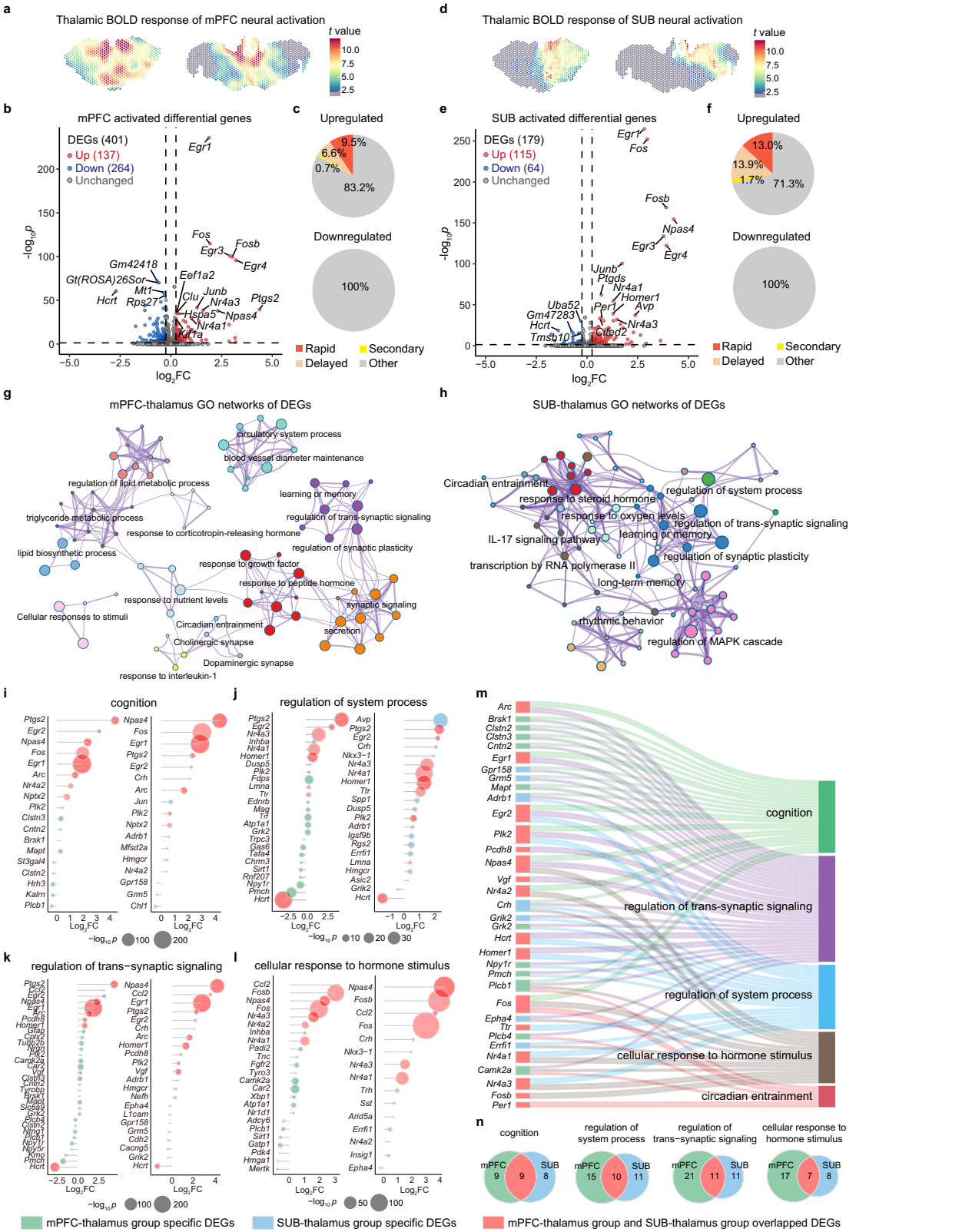
Fig. 2 | Coregistration of thalamic opto-fMRI-ST data into the same reference space. The thalamic BOLD activation spatial patterns (**a**) and intensity (**b**) caused by mPFC and SUB neural activation. The red boxes indicate the thalamic rostral and caudal slices that match the spatial transcriptomic slices (**a**). Each dot represents the mean *t* value of each thalamic subregion for each individual mouse. Statistical significance was determined using a two-tailed, two-sample *t* test. *P* values were corrected for multiple comparisons. ns: not significant; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$ (**b**). **c** The average time series caused by mPFC and SUB neural activation. The red and blue numbers represent the time to peak response in the ipsilateral and contralateral thalamic subregions, respectively. Data are presented as mean values +/− SEM. **d** Opto-ST timeline and registration strategy. **e** Thalamic spot-level BOLD activation pattern. See also Figs. S7–9. Source data are provided as a Source Data file.

asymmetry, characterized by robust BOLD responses in the ipsilateral thalamic subregions, while the contralateral responses remained minimal or near baseline levels (Fig. 2c). To enable spot-level quantitative comparison of region-specific responses, we aligned representative "best-fit" slices from rostral and caudal thalamic spatial transcriptomics datasets to the Allen Mouse Brain Common Coordinate Framework (CCFv3). Subsequently, all other samples were registered to these CCF-aligned reference slices using the SLAT algorithm. This approach enabled accurate mapping of multimodal data, including BOLD activation patterns, structural projections, and FosTRAP signals, onto spatial transcriptomics spots across different slice preparations (two independent spot-space datasets each for rostral and caudal regions), establishing a unified framework for integrative analysis (Fig. 2d, e).

Differentially expressed genes link to thalamic BOLD activation

To explore the genetic basis of BOLD responses, we analyzed differentially expressed genes in thalamic subregions showing significant BOLD activation (Fig. 3). In these activated areas of both mPFC and SUB groups, many upregulated and downregulated genes were identified (Figs. 3a–3f & Supplementary Data 1). When categorizing these genes by response speed, a few upregulated genes were immediate early genes, and none of the downregulated genes were immediate early genes (Fig. 3c, f).

https://doi.org/10.1038/s41467-026-71923-w



Comparing the mPFC and SUB groups, enriched terms related to learning, memory, cognition, and synaptic plasticity regulation were found, reflecting the cognitive functions governed by both regions (Fig. 3g, h). Circadian rhythm regulation terms were also enriched in both groups, possibly linked to arousal regulation (Fig. 3g, h). However, differences emerged: the mPFC group showed 'response to peptide hormones', suggesting a role in emotional response, while the SUB group had 'response to steroid hormones' (Fig. 3g, h). Similarly, the mPFC group exhibited 'response to interleukin-1 (IL-1)', and the SUB group showed 'interleukin-17 (IL-17) signaling pathway', indicating distinct roles in inflammatory responses (Fig. 3g, h).

Notably, terms related to the formation of BOLD activations and blood oxygen levels differed significantly between the two groups. The mPFC group included energy metabolism terms (e.g., lipid and

Fig. 3 | Identification of differentially expressed genes and genetic processes within BOLD-activated thalamic spots. Distribution of thalamic activated spots (**a**), differential gene expression (**b**), and DEG classification (**c**) induced by mPFC neural activation. Differential expression in (**b**) was assessed using a two-sided moderated *t*-test with Benjamini-Hochberg FDR correction; genes with FDR < 0.05 were considered significant. The upregulated genes are depicted in red, and the downregulated genes are depicted in blue in the volcano plot. The distribution of thalamic activated spots (**d**), differential gene expression (**e**), and DEG classification (**f**) induced by SUB neural activation. The same statistical criteria as in (**b**) were applied to (**e**). Gene processes enriched with DEGs in the mPFC (**g**) and SUB (**h**) groups. Common and distinct genes enriched in 'cognition' (**i**), 'regulation of system process' (**j**), 'regulation of transsynaptic signaling' (**k**) and 'cellular response to hormone stimulus' (**l**) in the mPFC and SUB groups. For **i**–**l**, enrichment significance was evaluated using a one-sided Fisher's exact test followed by FDR correction (Benjamini–Hochberg); terms with FDR < 0.05 were considered significantly enriched. Sankey (**m**) and Venn (**n**) diagrams summarizing the common and distinct genes enriched in the same terms across the mPFC and SUB groups. Differentially expressed genes are reported in Supplementary Data 1.

triglyceride metabolic processes) and blood oxygen level-related terms (e.g., blood circulation, vasodilation, and blood vessel diameter maintenance). In contrast, the SUB group mainly had terms related to 'response to oxygen levels' with few energy metabolism terms (Fig. 3g, h). These findings suggest that while fronto-thalamic and hippo-thalamic pathways may regulate similar macro-level functions, their underlying genetic mechanisms differ. This experimental paradigm, therefore, offers a means to investigate the genetic underpinnings of circuit-specific functions.

Distinct genetic profiles in mPFC-thalamus and SUB-thalamus pathways

To further characterize the genetic basis of BOLD responses, we compared gene sets enriched in both the mPFC and SUB groups for genes associated with the same functional terms (Figs. 3i-n). Our analysis revealed distinct genes involved in cognitive, arousal, and emotional regulatory functions within the mPFC-thalamus and SUB-thalamus networks. For instance, genes enriched in the 'cognition' term shared common elements between the mPFC and SUB groups, including *Arc*, *Egr1*, *Egr2*, *Fos*, *Npas4*, *Nr4a2*, *Nptx2*, *Plk2*, and *Ptgs2* (Fig. 3i). Among these, *Arc* and *Npas4* are implicated in synaptic plasticity and memory consolidation, while *Egr1*, *Egr2*, *Fos*, and *Nptx2* belong to the immediate early gene family. *Plk2* and *Ptgs2* are primarily associated with inflammatory responses. These findings suggest that these genes may play similar roles in regulating cognitive functions across different brain regions. However, region-specific genes, such as *Plcb1*, *Kalrn*, *Hrh3*, *Clstn2*, *Brsk1*, *Clstn3*, *Cntn2*, and *Mapt* in the mPFC group (Fig. 3i), might be involved in cognitive processes related to neural signal transmission and synaptic plasticity. In contrast, *Jun*, *Crh*, *Grm5*, *Gpr158*, *Nr4a2*, *Hmgcr*, and *Adrb1* in the SUB group (Fig. 3i) may be related to stress responses and neural transmission. These distinct genetic profiles between the mPFC and SUB groups reflect the specificity of cognitive functions originating from different brain regions.

Similarly, in the 'regulation of system process' term, shared common genes included *Ptgs2*, *Egr2*, *Nr4a3*, *Nr4a1*, *Plk2*, and others, which were the same as those in the 'cognition' term (Fig. 3j). These genes are either immediate early genes or related to inflammation. Additionally, the shared genes included *Homer1* and *Hcrt* (Fig. 3j). *Homer1* is involved in regulating cellular signal transduction, synapse formation, receptor clustering, and extracellular glutamate levels. *Hcrt* plays roles in regulating sleep, arousal, feeding, reward, fear, anxiety, and cognition. mPFC-specific genes, such as *Pmch*, *Npy1r*, *Nr4a2*, *Chrm3*, and *Fdps*, are involved in regulating appetite, metabolism, stress responses, and levels of arousal (Fig. 3j). The most prominent SUB-specific gene is *Avp* (Fig. 3j), also known as arginine vasopressin, which can directly bind to V1a receptors on vascular smooth muscle to regulate blood pressure and plays important roles in social behavior, stress response, circadian rhythms, memory, and attention.

Beyond these terms, we also compared the functions of genes involved in 'regulation of trans-synaptic signaling' (Fig. 3k) and 'cellular response to hormone stimulus' (Fig. 3l) to identify similarities and differences between the mPFC and SUB groups (Fig. 3m, n). These results demonstrated that both shared and distinct genes were involved in cognitive, arousal, and emotional regulatory functions associated with the mPFC–thalamus and SUB–thalamus networks.

Multicollinearity among altered gene expression patterns

To investigate the interconnections among four data modalities at high spatial resolution, we performed spatial correlation analyses between gene expression changes and opto-fMRI activation patterns (Fig. 4). In both the mPFC and SUB groups, Pearson correlation coefficients between gene expression changes and BOLD activation exhibited approximately normal distributions (Fig. 4a, d), indicating widespread transcriptomic associations with the fMRI activation patterns.

In the mPFC group, 434 genes showed significant positive correlations and 1,118 genes showed significant negative correlations with BOLD activation, as determined by 10,000 spatial permutation tests ($p < 0.05$, two-sided; Fig. 4a). The *Gm5741* gene had the strongest positive correlation with thalamic BOLD activation ($r = 0.33$, $p = 0.01$; Fig. 4b, c), which is related to signal transduction and G protein signaling. *Tac2* was the second strongest ($r = 0.32$, $p = 0.01$; Fig. 4c) and is involved in fear memory and neurotransmission. The most negatively correlated gene was *Car14* ($r = -0.28$, $p < 0.001$; Fig. 4a, c), which is involved in bicarbonate metabolism and cellular protection.

In the SUB group, 224 genes were positively correlated and 195 genes were negatively correlated with BOLD activation based on the same permutation-based statistical framework (Fig. 4d). *Npas4* had the strongest positive correlation ($r = 0.46$, $p < 0.001$; Fig. 4e, f), indicating its role in neural circuit development and learning. The most negatively correlated gene was *Acy3* ($r = -0.18$, $p < 0.001$; Fig. 4d, f), which is primarily involved in amino acid metabolism. Thalamic projections from SUB neurons showed a stronger correlation with BOLD activation patterns ($r = 0.69$, $p < 0.001$; Fig. 4e, f) than those from mPFC neurons ($r = 0.49$, $p = 0.08$; Fig. 4b, c). We note that viral expression in both SUB and mPFC groups covered broad regions, suggesting that the exact spatial extent of transfection may represent a common factor affecting functional-structural coupling estimates in both cases (Fig. S1). Notably, FosTRAP-based neural activation remained less correlated with BOLD activations than viral tracing-based projections in both the mPFC (Fig. 4b, c) and SUB groups (Fig. 4e, f), consistent with whole-brain level findings (Fig. 1j, k).

To assess the robustness of gene-wise associations under stringent multiple-comparison control, permutation-derived *p*-values were additionally corrected across all genes using the Benjamini–Hochberg false discovery rate (FDR) procedure. Only a limited number of genes survived FDR correction, reflecting the highly conservative nature of gene-wise correction in this high-dimensional transcriptomic setting (Supplementary Data 2). Importantly, the primary aim of this analysis was not genome-wide gene discovery, but rather to characterize system-level transcriptomic patterns associated with fMRI activation.

Finally, correlation heatmap of the top 10 genes associated with MRI, projections, and FosTRAP revealed significant correlations between the genes themselves (Fig. 4c, f). These results highlight the need to consider high multicollinearity among independent variables in subsequent modeling and analysis. Traditional multiple linear regression models may yield unstable and unreliable results due to

https://doi.org/10.1038/s41467-026-71923-w

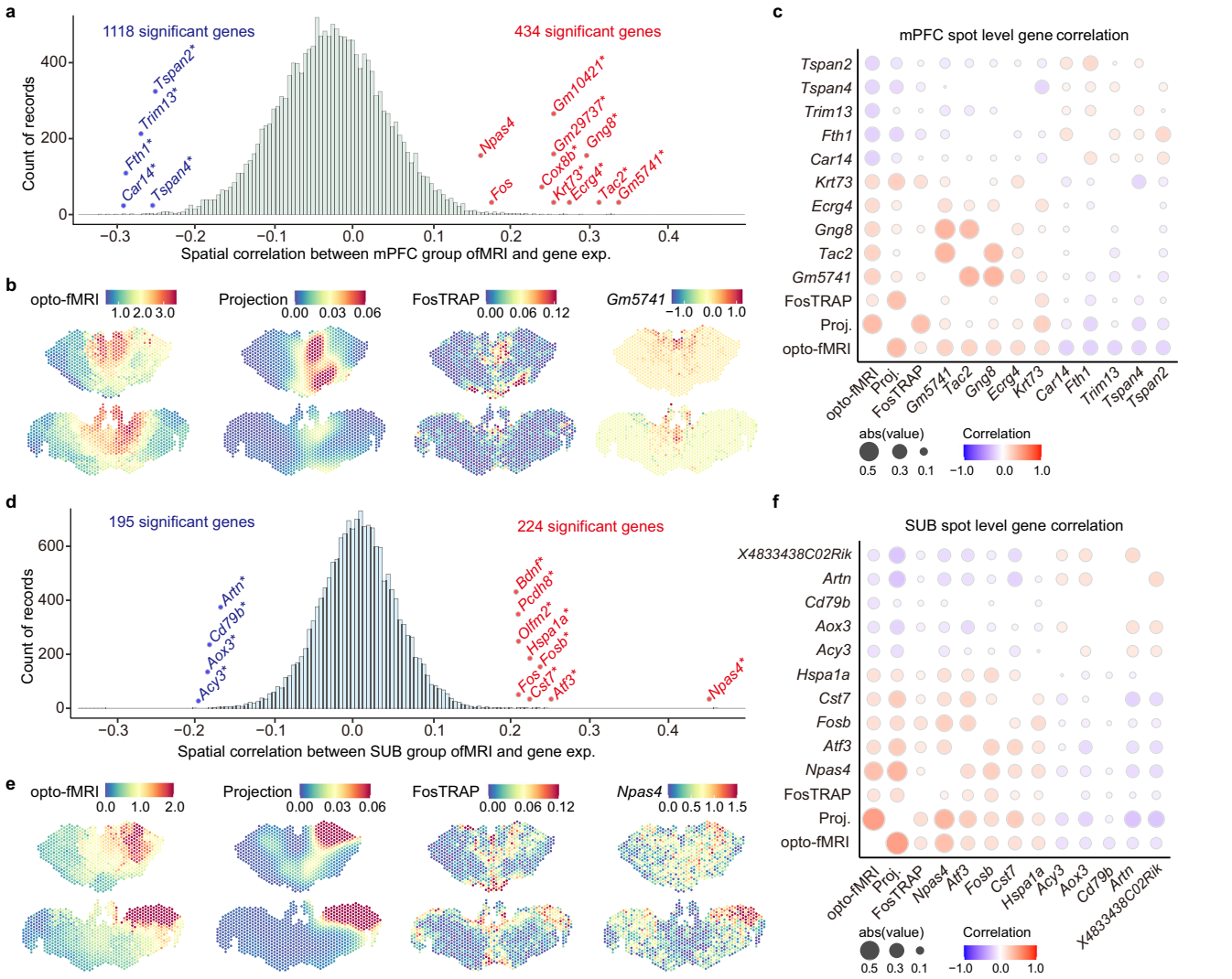


Fig. 4 | Multicollinearity among altered gene expression patterns. a Spatial correlation between the BOLD activation pattern and altered gene expression patterns following mPFC optogenetic activation of the ST subgroup. Statistical significance was assessed using 10,000 spatial permutation tests (two-sided). Genes with permutation-based $p < 0.05$ are highlighted. **b** Spot-level maps of opto-fMRI activation, projection, FosTRAP activation, and *Gm5741* expression changes in mPFC activation. **c** Pearson correlation between opto-fMRI activation, projection, FosTRAP, and the top 10 genes in mPFC activation, using two-sided tests. **d** Spatial correlation between fMRI activation patterns and gene expression patterns following SUB optogenetic activation. Statistical significance was assessed using the same two-sided permutation-based framework as in (**a**), and genes with permutation-based $p < 0.05$ are highlighted. **e** Spot-level maps of opto-fMRI activation of ST-subgroup, projection, FosTRAP activation, and *Npas4* expression changes in SUB activation. **f** Pearson correlation between opto-fMRI activation, projection, FosTRAP, and the top 10 genes in SUB activation. Permutation-derived *p*-values were additionally corrected across all genes using the Benjamini–Hochberg FDR procedure as a stringent robustness analysis; FDR-corrected results are reported in Supplementary Data 2.

inflated variances of regression coefficients, sensitivity to outliers, and the inability to capture nonlinear relationships or handle non-continuous variables[44,45], leading to issues such as unstable parameter estimation and increased standard errors.

Identifying key genes to explain and predict the thalamic BOLD response using the PLS model

The PLS model effectively handles multicollinearity by extracting latent variables that maximize covariance between independent and dependent variables, thereby reducing dimensionality, enhancing explanatory power, and providing robustness against instability caused by high intercorrelations among predictors[45,46]. Furthermore, PLS has become a mainstream and preferred method in the *imaging transcriptomics* field[10–18]. We considered gene expression alteration maps, projection maps, and FosTRAP activation maps as dependent variables, with the BOLD activation map as the independent variable (Fig. 5a). Leave-one-out cross-validation (LOOCV) was used to assess the predictive performance of both the mPFC and SUB models (Figs. S10 & S11). Based on the performance of each fold model, including explained variance, prediction accuracy, and MSE/RMSE of prediction (Fig. S10) and permutation test (Fig. S11), the optimal number of components was determined. With the optimal number of components, the mPFC model exhibited a mean predictive accuracy of 0.57, correlating empirical and predicted BOLD activation (Fig. 5f), whereas the SUB model demonstrated a mean predictive accuracy of 0.62 (Fig. 5j).

Within each fold of LOOCV, the variable importance in projection (VIP) score[47] was computed to assess the contribution of each dependent variable across different components. Variables with VIP scores above 1.5 across all folds were identified, yielding 615 genes for the mPFC final model (Fig. 5g & Supplementary Data 1) and 617 genes for the SUB final model (Fig. 5k & Supplementary Data 1). The performances of both final models are depicted in Fig. S12. The mPFC final model selected 5 components, accounting for 86.0% of BOLD

https://doi.org/10.1038/s41467-026-71923-w

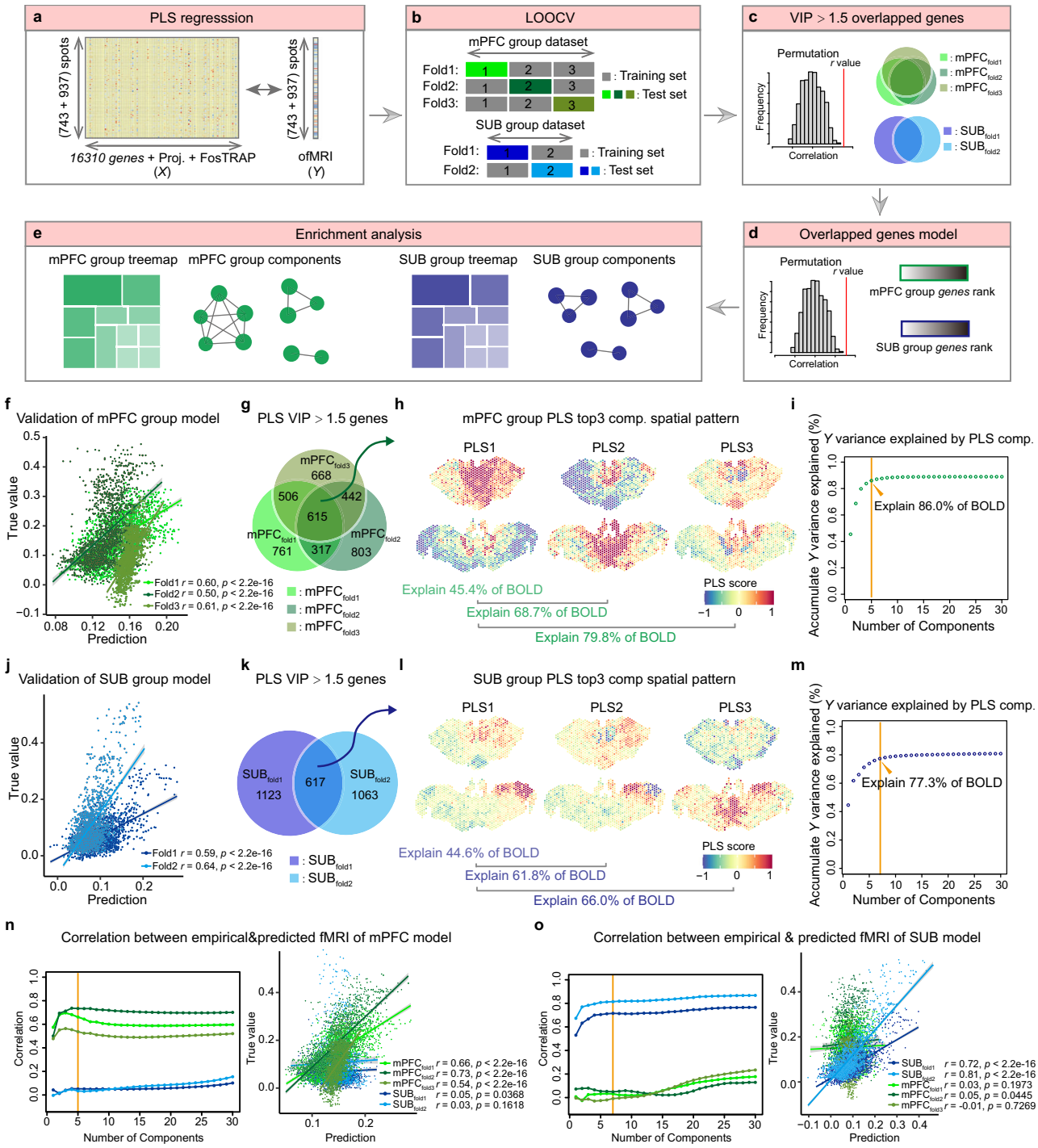


Fig. 5 | Decoding models used to predict GLM-based thalamic BOLD activation patterns based on the gene expression patterns. a–**e** Analysis flowchart. **f** Correlation analysis between the predicted and actual values for each fold of the mPFC model. The solid line indicates the linear regression fit, and the shaded area represents the 95% confidence interval. *r*, Pearson correlation coefficient. **g** Cross-fold identification of intersecting genes with VIP scores greater than 1.5 in the mPFC model. **h** Patterns of the first three components in the space constructed by the intersection genes of the mPFC model. **i** Explained variance in BOLD patterns according to the intersection gene model in the mPFC model. **j** Correlation analysis between the predicted and actual values for each fold of the SUB model. The solid line indicates the linear regression fit, and the shaded area represents the 95% confidence interval. *r*, Pearson correlation coefficient. **k** Identification of intersection genes with cross-fold VIP scores greater than 1.5 for the SUB model. **l** Patterns of the first three components in the space constructed by the intersection genes of the SUB model. **m** Correlation analysis between the predicted and actual values for each fold of the SUB model. Predictive ability of the mPFC model for mPFC and SUB group data (**n**) and predictive ability of the SUB model for mPFC and SUB group data (**o**). The solid line indicates the linear regression fit, and the shaded area represents the 95% confidence interval. *r*, Pearson correlation coefficient. See also Fig. S10–14.

activation (Fig. 5i), whereas the SUB final model chose 7 components, explaining 77.3% of BOLD activation (Fig. 5m). We also sorted the weights of the genes for each component and performed preliminary enrichment analysis on the top 100 genes with the highest weights (Fig. S14). These findings indicated that the key genes and genetic processes behind these thalamic low-dimensional networks presented some differences.

Furthermore, the final model of the mPFC had high predictive power for thalamic responses induced by mPFC activation but could not predict those caused by SUB activation (Fig. 5n). Conversely, the final SUB model was effective in predicting thalamic responses from SUB activation but not from mPFC activation (Fig. 5o). These results demonstrated the reliability and stability of our multimodal integration model, providing a solid foundation for exploring the genetic mechanisms behind complex BOLD signals.

Energy metabolism- and basic biochemistry-related genes play key roles in explaining and predicting GLM-based BOLD activation patterns

Finally, the VIP scores of 615 genes across 5 components of the mPFC final model and 617 genes across 7 components of the SUB final model were estimated (Fig. 6 & Supplementary Data 1). Interestingly, genes associated with energy metabolism play pivotal roles in BOLD activation patterns. For the mPFC group, genes such as *Ckb*, *Pkm*, *Aldoa*, *Atp5a1*, and *Atp6v0c* were identified as influential (Fig. 6b). *Ckb*, *Atp5a1*, and *Atp6v0c* are directly involved in ATP formation and maintaining the balance of phosphocreatine. Moreover, *Pkm* and *Aldoa* play key roles in the glycolytic process. In the SUB group, genes such as *Secisbp2*, *Slc52a2*, and *Zfp719* had significant impacts (Fig. 6e). *Slc52a2* is related to purine metabolism, whereas *Zfp719* is associated with lipid metabolism.

The gene enrichment analysis results indicate that the mPFC group is enriched with many terms related to energy metabolism, such as 'Metabolism of carbohydrates', 'Glycolysis', 'ATP metabolic process', and 'ADP catabolic process' (Fig. 6c). In the SUB group, fewer terms related to the regulation of energy metabolism were enriched, such as 'methylation', which affects pathways such as glycolysis; the citric acid cycle; lipid metabolism; and 'mitochondrial transcription', which affects mitochondrial function and energy metabolism (Fig. 6f). These results indicate that the genes contributing the most to the BOLD activation pattern are almost exclusively related to energy metabolism. These findings highlight the complex interplay between energy metabolism and BOLD signal generation, as evidenced by extensive literature using PET/transcription linked to resting-state fMRI[48–53]. These findings consistently revealed that areas of heightened metabolic activity correlate with increased MRI signal amplitudes or enhanced connectivity. Our results further reveal that the activation of neurons in different brain regions triggers the activation of different types of metabolic pathways in the thalamus, demonstrating that the brain's energy demands are heavily dependent on the activated circuits or networks.

Moreover, we found that the gene sets capable of explaining and predicting BOLD response patterns could no longer be enriched in macrolevel terms such as 'cognition' and 'memory', as shown in Fig. 3, but instead were enriched in the microprocesses that regulate macrofunctions, such as the DNA metabolic process, tRNA methylation, the mitochondrial RNA metabolic process, and ribosome assembly (Fig. 6c, f). A possible explanation is that the current BOLD activation patterns are derived from generalized linear model (GLM) calculations. Considering the entire spatial pattern, the greater number of primary gene expression changes in each brain region may be linearly related to the intensity of the input received. This is also the reason why the expression patterns of genes related to fundamental biochemical processes contribute strongly to GLM-based activation patterns.

Interestingly, we found that terms related to DNA damage, repair, and inflammatory responses were enriched in both the mPFC and SUB groups (Fig. 6c & f). Recent high-level research[54,55] has shown that learning and memory formation stimulate intense electrical activity in brain neurons, leading to double-strand breaks in their DNA. After DNA damage, specific signaling pathways in neurons are activated to repair DNA damage. Memory formation and consolidation occur during this process of DNA damage and repair in neurons. We speculate that the intense electrical activity and subsequent processes caused by optogenetic activation may be similar to the aforementioned memory formation process.

Notably, the VIP scores for projection (VIP = 0.32) were relatively low for the mPFC group, suggesting that their contribution to explaining and predicting BOLD activation is minimal. However, for the SUB neural activated thalamic response pattern, the SUB neural projection to the thalamus pattern provided the greatest contribution to thalamic BOLD activation (VIP = 2.6, Fig. 6e).

In summary, the pattern of expression changes in energy metabolism, fundamental biochemistry process-related genes induced by optogenetic activation in both the mPFC and SUB groups, contributes significantly to the GLM-based BOLD activation pattern.

Data-driven PCA activation mapping and conserved transcriptomic associations

To validate that the spatial patterns identified by our GLM analysis were not contingent on its specific model assumptions, we performed a complementary, data-driven analysis using Principal Component Analysis (PCA) as used in one previous opto-fMRI research[56]. This model-free approach identifies the dominant, co-varying patterns of signal variance in the fMRI data without reference to the stimulus paradigm or an assumed hemodynamic response. Strikingly, the spatial map of the first principal component (PC1) exhibited a strong qualitative resemblance to the group-level activation pattern derived from the GLM (Pearson's $r = 0.91$ for the mPFC group and $r = 0.55$ for the SUB group, Fig. S15a, b). This indicated that the core network architecture of the optogenetically evoked response is a major source of variance in the data and is robustly detectable across distinct analytical frameworks.

We then projected this PCA-derived activation pattern onto our spatial transcriptomic data using the same PLS framework (Fig. S16, S17). We found that the key gene categories were still predominantly enriched in terms related to energy metabolism, fundamental biochemical reactions, and DNA damage and repair. This convergence between a model-free and a model-driven approach strongly reinforces our central conclusion: the link between these transcriptional programs and the large-scale brain activation detected by linear methods represents a robust biological relationship, rather than an analytical artifact.

Optogenetic behavioral tests

To verify the efficacy and safety of our optogenetic activation paradigm for engaging large-scale networks, we performed a series of behavioral assays (Fig. S18). Critically, all stimulations were conducted under parameters that did not induce epileptiform activity or other overt neurological abnormalities, confirming that subsequent observations reflect physiological network engagement rather than pathological hyperexcitation.

Behavioral profiling revealed that activating either the mPFC or the SUB enhanced general exploratory activity, though through dissociable behavioral dimensions. In the open field test, mPFC stimulation specifically increased center zone exploration without affecting total locomotion, consistent with reduced anxiety-like behavior (Fig. S18a, b). In contrast, SUB stimulation decreased total locomotor distance without altering center activity (Fig. S18g, h). In cognitive tasks including spontaneous alternation (Y-maze) and novel arm

https://doi.org/10.1038/s41467-026-71923-w

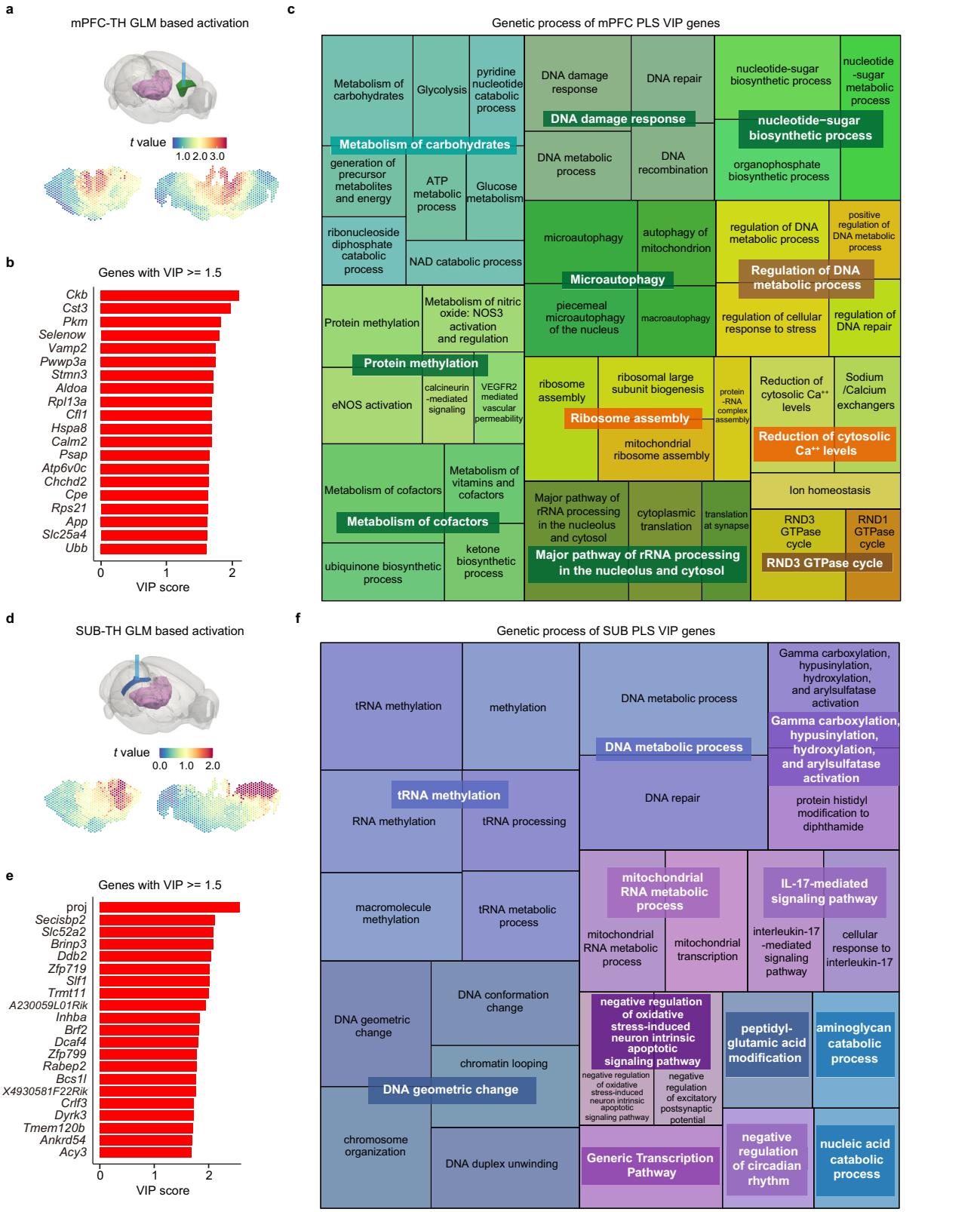


Fig. 6 | Energy metabolism- and basic biochemistry-related genes play key roles in explaining and predicting GLM-based BOLD activation patterns. a GLM-based BOLD activation pattern of mPFC neural activation. **b** Ranking of the genes of the mPFC model based on VIP scores. **c** Enrichment results for all 615 genes in the mPFC final mode, assessed using one-sided Fisher's exact test. The area of each block in the TreeMap represents the $-\log_{10}$ (adjusted p-value). Only terms with adjusted p-values < 0.05 are displayed. **d** GLM-based BOLD activation patterns of SUB neural activation. **e** Ranking of the genes of the SUB model based on VIP scores. **f** Enrichment results for all 617 genes in the SUB final model. See also Figs. S15–17. Gene VIP scores are reported in Supplementary Data 1. Source data are provided as a Source Data file.

recognition, neither manipulation altered performance, indicating preserved spatial working memory and novelty discrimination (Fig. S18c, d, i, j). In the novel object recognition test, both mPFC and SUB activation increased absolute exploration of the novel object, while the relative novelty preference remained unchanged, further supporting a general rather than memory-specific enhancement of exploratory drive (Fig. S18e, f, k, l).

Together, these results establish that our optogenetic paradigm reliably modulates macro-circuit function related to motivated exploration, with mPFC influencing anxiety-related behavioral inhibition and SUB regulating motor activity levels. This behavioral dissociation aligns with the distinct, large-scale activation patterns observed in our concurrent opto-fMRI-ST framework, thereby functionally validating the network engagement elicited by our stimulation protocol.

Discussion

We developed a systematic framework integrating opto-fMRI, opto-ST, opto-FosTRAP, and projection data to decipher the microscopic mechanisms underlying macroscopic neuroimaging phenotypes. Opto-fMRI is a potent method for mapping whole-brain responses elicited by specific cellular populations within targeted brain regions, particularly in awake rodents[57–60]. The temporal dynamics of BOLD signals, governed by neurovascular coupling (NVC), reflect a complex integration of direct neuronal projections, multitiered secondary projections, and local cellular interactions[60–62]. Disentangling the BOLD signal is equivalent to unraveling these complex neurovascular dynamics, which is essential for enhancing our comprehension of intricate neural networks and their functional correlates. The integration of BOLD responses with ST responses induced by the same optogenetic stimulation paradigm enables investigation of the neurobiological mechanisms underlying BOLD signals.

Human *imaging transcriptomics* studies, leveraging resources such as AHBA[3], have revealed putative genetic with diverse brain phenotypes, including canonical resting-state networks[63,64], functional organization[9], brain functional homotopy[65], fiber tract connectivity[66], regional maturation during brain development[67] and pathological alterations in brain disorders[10,13,68–71]. A key limitation of AHBA is that its gene expression data derive from only six healthy elderly individuals, while imaging datasets typically come from separate cohorts with different demographic characteristics. This limits the ability to perform computational cross-validation, which is essential for ascertaining the reliability of the genes identified. While direct experimental validation in humans remains challenging, animal models offer unique advantages for testing these relationships. Animal models enable precise manipulation of specific genes and circuits, combined with both imaging and molecular readouts under controlled conditions[11].

Upon reviewing the aforementioned literature, the PLS model has emerged as the predominant algorithm[10,11,13]. However, existing studies have focused predominantly on the first PLS component to elucidate and/or predict neuroimaging data, as well as to pinpoint associated genes. However, our findings indicate that a model relying solely on the first PLS component has relatively poor explanatory and predictive power for BOLD activations (Fig. 5i & m). This observation aligns with a nonhuman primate study[11], suggesting that relying on genes identified by the first PLS component alone is insufficient.

Our results reveal a critical insight: genes with the strongest individual correlations to the BOLD activations were not always the top contributors in our multivariate model, such as anatomical projection maps. While both mPFC-thalamo (mPFC-proj) and subiculum-thalamo (SUB-proj) projection patterns showed significant univariate correlations with the thalamic BOLD activation pattern, only SUB-proj emerged as a high-VIP (top-ranked) predictor in the PLS model. This result is best explained by the high multicollinearity between the mPFC-proj map and a specific module of co-varying, stimulation-induced gene expression patterns. The spatial information encoded by the mPFC projection map was largely redundant with this transcriptional module, reducing its unique contribution in the multivariate context. This suggests that the functional impact of mPFC inputs to the thalamus is tightly and directly reflected in the evoked transcriptional state of the circuit. In contrast, the SUB-proj map contributed information not fully encapsulated by the gene expression variables, implying a potentially divergent mechanism linking hippocampal activity to thalamic BOLD activations. This analysis demonstrates how multivariate modeling can disentangle overlapping versus unique contributions of anatomical and molecular features to large-scale brain activity.

A key finding is that the pattern of expression changes in energy metabolism-related genes induced by optogenetic activation, especially in the mPFC, significantly contributes to the BOLD activation pattern. This discovery underscores the notion that the generation process of BOLD signals is intricately linked to energy metabolism, as evidenced by shifts in gene expression patterns. This perspective is reinforced by previous studies that have established a close relationship between the generation of BOLD signals and energy metabolism, primarily based on resting-state fMRI of the human brain[48–52]. Tomasi et al. (2013) demonstrated a linear association between glucose metabolism and the amplitude of low-frequency MRI signal fluctuations[52]. Similarly, Deng et al. (2022) reported significant correlations between BOLD-derived variables such as the amplitude of low-frequency fluctuations (ALFF), fractional amplitude of low-frequency fluctuations (fALFF), and regional homogeneity (ReHo) with PET measurements of cerebral metabolic rates and hemodynamics[48]. Furthermore, Aiello et al. (2016) integrated PET with MRI to provide a comprehensive view of the metabolic aspects that are crucial for interpreting BOLD signals[51]. To our surprise, we found that metabolism-related and fundamental biochemistry genes were linked to the current GLM-based BOLD response patterns (Fig. 6c, 6f).

GLM-based analysis is widely used in both animal and human task-based fMRI studies, relying on it to identify brain activity patterns during active or passive tasks[72,73]. This method is valuable for modeling the hemodynamic response to neural activity and fits the same model to each voxel's time course, which is standard in fMRI analysis and well documented in the literature[72,73]. However, the use of a uniform HRF across all brain regions may introduce bias, as studies have shown variations in HRFs in the olfactory bulb, sensory cortex, and auditory cortex of mice[62]. We also found that thalamic subregions exhibited distinct and complex dynamics (Fig. 2c), and the relationship between these dynamics and the stimulation paradigm was not simply linear.

However, given the current technical limitations and the absence of mature nonlinear analytical frameworks for task-fMRI in the short term, we performed a complementary, data-driven analysis using PCA (Fig. S15–17). As a model-free approach, PCA requires no predefined stimulus paradigm or assumed hemodynamic response and has been successfully applied in prior opto-fMRI studies[56]. Reassuringly, the spatial pattern of the first principal component (PC1) closely recapitulated the GLM-derived activation map (Fig. S15), and similar key gene processes—particularly those involved in energy metabolism, basic biochemistry, and DNA damage repair—remained significantly correlated with this model-free activity pattern. This convergence suggests that the relationship between these transcriptional programs and large-scale brain activation is robust and not strictly tied to a specific analytical framework.

We acknowledge that PCA, while model-free, captures the largest sources of variance in the data, which may theoretically include structured noise or unmodeled physiological signals. Interpretation of PCA-derived activation maps and their associated gene enrichments therefore, requires caution, particularly in the absence of an explicit stimulus model. Future work incorporating region-specific HRF

modeling and more sophisticated nonlinear approaches will be essential to fully unravel the complex dynamics of large-scale brain activity.

Intriguingly, our analysis revealed an enrichment of terms associated with DNA damage, repair mechanisms, inflammatory responses, and autophagy in both the mPFC and SUB groups (Fig. 6c–f, S16, S17). While recent studies have associated similar processes with learning and memory[54,55], our behavioral data—showing enhanced general exploration but not improved memory performance (Fig. S18)—suggest a different context. We speculate that in our paradigm, these molecular signatures may predominantly reflect a state of cellular stress or homeostasis triggered by the intense and prolonged optogenetic activation, which elicits broad hemodynamic responses. Future studies are needed to dissect whether such responses are epiphenomena of strong stimulation or integral to specific cognitive processes.

Our study utilized awake mouse fMRI, a paradigm that offers critical advantages over anesthetized preparations by eliminating anesthesia-induced suppression of neural and vascular activity, avoiding the physiological complications of mechanical ventilation, and enhancing translational relevance to human studies[59,74–76]. To mitigate potential confounds of stress and motion, we implemented a robust experimental framework[59,60,62,77–79]. The effectiveness of our extended 7–14 day habituation protocol was confirmed by measuring plasma corticosterone levels, which returned to pre-restraint baseline, indicating a significant reduction in anxiety[79]. Furthermore, our advanced noise reduction strategy—incorporating regression of noise components derived from extra-cerebral signals—effectively minimized motion artifacts. This was evidenced by the correlation between global DVARS and FD, which was reduced from approximately 0.7 before denoising to below 0.2 after processing (Fig. S2), thereby substantially strengthening the validity and interpretability of our BOLD data[60,77].

Our awake optogenetic fMRI approach offers a physiologically relevant alternative to anesthesia, though the awake state itself introduces methodological considerations. Despite extensive habituation, residual stress-related neural tone remains a potential confound. We acknowledge that while habituation markedly reduces acute stress, residual arousal-related influences on neural circuits cannot be fully excluded and represent an inherent methodological limitation. Regions such as the mPFC, hippocampus, and thalamic nuclei are highly sensitive to internal state[80], and residual arousal could modulate baseline activity or response gain in these networks via neuromodulatory inputs[81]. Nevertheless, the awake preparation provides a key advantage by avoiding anesthesia-induced suppression of midline thalamic activity and decoupling of the prefrontal-hippocampal axis[82]. By assessing networks in quiet wakefulness, our approach yields a more physiologically relevant representation of brain-wide dynamics than is possible under anesthesia-related neurovascular suppression[79,83].

In addition, several non-specific factors that require careful consideration. First, minor light leakage may engage visual pathways and could, in principle, affect thalamic BOLD responses via visual thalamic nuclei, consistent with previous reports in awake optogenetic fMRI studies[77]. Second, localized intracranial heating at the fiber tip may contribute to signal changes immediately at the stimulation site, which might occur through neuronal suppression[84]. These effects were observed in control animals and appeared spatially limited.

Crucially, our experimental design mitigates these concerns in several ways. The use of a case-control paradigm and statistical inference based on direct experimental-versus-control contrasts reduces—though does not completely eliminate—the possibility that such non-specific factors drive the reported network-level and behavioral conclusions. Furthermore, characterization of light spread in fresh tissue and histological verification of viral expression (Fig. S1) indicate that optical stimulation was largely confined to the intended target. However, light propagation may extend beyond the anatomical target, and activation of nearby virally infected areas cannot be entirely excluded. Additionally, local heating may influence BOLD signals immediately at the fiber tip, as evidenced by the limited negative BOLD observed in controls (Fig. S5). We therefore interpret the resulting activation patterns as network-level responses to stimulation of the targeted neuronal populations, rather than as strictly spatially confined or monosynaptic effects.

When compared with earlier optogenetic studies[57,59,85,86], the awake preparation and high field strength (11.7 T) enhance sensitivity and enable detection of broader propagation, including previously elusive regions such as the VTA[85]. The spatially extensive BOLD activation evoked from mPFC contrasts sharply with the localized response from SUB under identical conditions. This distinction aligns with the prior work in the basal forebrain[60], where stimulation of excitatory (VGLUT2$^+$) neurons drove brain-wide activation, whereas stimulation of inhibitory populations produced focal responses. Given that mPFC consists of ~80% excitatory pyramidal neurons[87], its capacity for large-scale recruitment—similar to BF VGLUT2$^+$ neurons—supports the interpretation that the observed widespread signals reflect mPFC's role as a highly connected network hub[88,89], rather than a technical artifact of excessive light power.

Although unintended visual-circuit activation due to minor light leakage was present in control animals (Fig. S5), both experimental and control groups underwent identical stimulation protocols. All reported neural and behavioral effects were derived from direct experimental-versus-control comparisons with FDR correction, thereby removing shared sensory-driven components. The long inter-trial interval (40 s) further limits precise temporal prediction, and any residual expectation-related modulation would likely affect distributed networks rather than selectively bias visual pathways. The absence of behavioral effects in control animals reinforces that visual leakage does not account for the main findings.

The study has several technical constraints. First, achieving precise spatial correspondence between the viral transduction area, optical stimulation site, and transcriptomic sampling region remains challenging; ideal structure–function coupling requires highly specific targeting at each stage, which is rarely accomplished in current opto-fMRI studies. Second, our transcriptomic analysis was restricted to two coronal slices, limiting comprehensive integration of multimodal data. Future whole-brain spatially resolved transcriptomics will be essential to fully realize the potential of our modeling framework. Third, the single post-stimulation time point (~40 min after onset) captures robust early transcriptional waves but may miss both immediate-early gene induction and later sustained programs. Multi-time-point designs are needed to resolve the full temporal progression of activation-induced transcription. Fourth, although we validated results using a model-free PCA approach, canonical GLM-based modeling of BOLD responses may obscure nuanced region-specific relationships—particularly for gene categories related to energy metabolism and core biochemical processes. More sophisticated nonlinear analytical frameworks will help better link hemodynamic changes with molecular profiles. Finally, while our control design and contrasts minimize the influence of non-specific effects such as visual leakage and local heating, these factors cannot be fully excluded and should be considered when interpreting signal polarity at the stimulation site and the broader network response.

In conclusion, by integrating opto-fMRI, opto-ST, opto-FosTRAP, and anatomical tracing within the same experimental individuals, our work advances the field of *imaging transcriptomics* and overcomes prior limitations. Focusing on the mPFC and SUB, we identified distinct thalamic activation patterns and gene expression changes, underscoring the role of specific neural projections in modulating BOLD responses. Our findings highlight the importance of region-specific

neural activity in regulating gene expression and its correlation with BOLD signals. Moreover, the development of decoding models that spatially link multimodal data has enabled us to explore the genetic underpinnings of mPFC–thalamic and SUB–thalamic networks, enhancing our understanding of the neurogenetic basis of cognition, motivation, and emotion. This work not only deepens our knowledge of the neurogenetic landscape but also lays the groundwork for future studies to develop more complex nonlinear models. These advancements will improve the calculation of BOLD response patterns, ultimately enhancing our ability to understand and treat neurological disorders.

Methods

Mice

All animal procedures were conducted in accordance with the guidelines of the Institute of Biophysics, Chinese Academy of Sciences. Adult mice of both sexes, starting at 8 weeks, underwent surgery and under a 12-h light/dark cycle at a controlled temperature of 20–26 °C and relative humidity of 50 ± 10%. Animals were housed individually and had ad libitum access to food and water. The following mouse strains and sources were used: $Fos^{2A\text{-}iCreER}$ (also known as $Fos^{2A\text{-}iCreERT2}$ or TRAP2), Jackson catalog number IMSR_JAX:030323[40]; the Ai47 (Rosa-CAG-LoxP-triGFP) Cre reporter line was used as the reporter for the TRAP2 system; FosTRAP2×Ai47 mice refer to offspring from the crossbreeding of TRAP2 mice with Ai47 reporter mice, in which GFP expression in the cytoplasm reflects *Fos* gene expression. The experimental animals involved breeding, with C57BL/6 N background mice used for crossbreeding. Both sexes were included, and the data were not sex specific. The list of animal data used in all experiments is presented in Table 1. All procedures were approved by the Institutional Animal Care and Use Committee of the Institute of Biophysics.

Stereotaxic surgeries

Using a stereotaxic atlas, we injected viral solutions (1) AAV2/9-hSyn-hChR2(H134R)-mCherry-WPRE-pA (S0165-9; Shanghai Taitool Bioscience Co., Ltd.) and (2) AAV2/9-pAAV-SYN-MCS-mCherry-3FLAG-WPRE (OBiO Technology Corp., Ltd.) into the mPFC (AP: +1.70 mm, ML: +0.35 mm, DV: −2.79 mm) and SUB (AP: −2.92 mm, ML: +1.36 mm, DV: −1.58 mm) at a rate of 30 nL/min for 150 nL. The needles were left in place for 5–10 minutes to prevent backflow. After injection, we removed the muscle from the skull posteriorly to expose the skull for head fixation, avoiding blood vessels. The skull was kept dry and covered with 3M tissue adhesive (3 M Vetbond Tissue adhesive; PN: 1469SB; 3M). Dental resin was applied to the skull, followed by a head fixation bar. Once the cement hardened, more material was added to the edges. A fiber-optic ferrule was placed 200 µm above the injection site, and UV-curing resin was applied and cured around it. The fiber was aligned with the skull and coated with resin. The mice were placed on a heated blanket and then moved to individual cages for a 4-week recovery before acclimation training.

Habituation for awake opto-fMRI

To acclimate the mice to the fMRI environment and mitigate post-surgical stress, we implemented a seven-day habituation protocol[60,62,77]. Training involved speakers playing the recorded fMRI scanning sounds to mimic the scan acoustics, with the mouse's head and body secured. The protocol was escalated daily in duration and volume until the mice were calm under head fixation and high noise. For the last two days, the mice were exposed to the MRI room for actual scans to ensure full adaptation. The training parameters are detailed in Table 2.

Optogenetic fMRI acquisition

MRI data were acquired using a 11.7 T Bruker BioSpin scanner (500 MHz) operated with ParaVision 360 (version 1.1, Bruker BioSpin). A quadrature volume coil was used for RF excitation, and a 4-channel mouse head coil was used for signal reception. Pre-opto-fMRI, T2-weighted images were obtained using a RARE sequence: TE = 30 ms, TR = 6000 ms, average = 6, RARE factor = 8, FOV = 16 mm × 12 mm, matrix = 230 × 172, slice thickness = 0.4 mm, resolution = 0.07 mm × 0.07 mm, coronal slices = 40, duration = 12 min 36 s. For opto-fMRI, a 473 nm laser module outside the room delivered light to an optical fiber using a 6 m patch cable. An Arduino Uno-controlled optogenetic stimulation synchronized with EPI scanning using a block-design paradigm: 20 s stimulation, 40 s rest, 20 Hz frequency, and 10% duty cycle. For all experiments, we used a 200 µm core diameter, 0.73 NA optical fiber. The output power at the fiber tip was measured to be 15 mW for continuous wave light. The calculated maximum irradiance at the fiber tip was therefore approximately 478 mW/mm². However, all stimulations were performed not with continuous wave light, but with a pulsed protocol, resulting in an average irradiance of approximately 89.6 mW/mm². The EPI sequence parameters were as follows: TE = 15 ms, TR = 1000 ms, repetitions = 340, bandwidth = 652173.9 Hz, flip angle = 50°, FOV = 20.125 mm × 11 mm, matrix = 168 × 92, slice thickness = 0.5 mm, resolution = 0.12 mm × 0.12 mm, coronal slices = 25, duration = 5 min 40 s. Each mouse underwent 3 scans.

fMRI data processing

All data preprocessing was implemented using Nipype workflows, which integrated various neuroimaging tools (including DCM2NII version 24Nov2014 and ANTs version 2.3.5, AFNI version 20.3.03). Subsequent data analysis and visualization were performed using custom Python scripts (version 3.9.7) and Nilearn (version 0.9.1). For fMRI preprocessing, the following steps were applied: data conversion from Bruker's DICOM files to the NIfTI format was achieved using the DCM2NII tool; the initial 10 volumes of EPI images were discarded to allow the signal to reach a steady state; the data were then applied to automate the skull-stripping process based on a deep learning model; the data were spatially aligned to both individual anatomical spaces and the 3D Allen Mouse Brain Common Coordinate Framework, v3 (CCFv3, http://atlas.brain-map.org/) for group analysis; and light spatial smoothing (0.45 mm isotropic Gaussian kernel), nuisance regression of BOLD signals, and general linear model (GLM) region-specific optogenetic stimulation were performed.

To develop an automated skull-stripping pipeline, we first manually annotated mouse brains in both T2-weighted structural images and averaged functional images using ITK-SNAP (version 4.2.0) software. These manual annotations were used to train a 3D U-Net model for automated skull-stripping. For subsequent data processing, initial automated skull-stripping was performed using this trained model, followed by manual inspection and refinement when necessary to ensure precise brain extraction.

To achieve precise spatial normalization, EPI data underwent sequential registration steps. First, individual EPI volumes were aligned to a selected skull-stripped functional average image using FSL FLIRT (version 6.0) with rigid transformation (translation and rotation). This functional average was then registered to its corresponding high-resolution structural image, followed by registration to the T2-weighted DSURQE (Dorr–Steadman–Ullmann–Richards–Qiu–Egan) mouse MRI template, and finally to the Allen CCF v3 standard space. While rigid transformation (FSL FLIRT) was sufficient for within-subject functional alignment, non-linear registration using ANTs (Advanced Normalization Tools, http://stnava.github.io/ANTs/) SyN algorithm with mutual information metric was employed for inter-modality (functional to structural) and inter-template registrations, ensuring precise alignment across spatial scales.

To minimize the impact of scanner drift, head motion, and physiological noise, nuisance regression of BOLD signals was performed using a combination of 6 head motion parameters and their first derivatives, along with an optimal set of principal components (PCs)

Table 1 | Number of animals used in experiments

Experiments	Experimental group	*N* number	Description
Opto-fMRI	mPFC test group	15	Total 180 runs
	mPFC control group	14	Total 136 runs
	SUB test group	12	Total 114 runs
	SUB control group	9	Total 95 runs
Opto-ST	mPFC test group	3	5 rostral slides and 5 caudal slides
	mPFC control group	3	4 rostral slides and 4 caudal slides
	SUB test group	2	4 rostral slides and 4 caudal slides
	SUB control group	2	3 rostral slides and 3 caudal slides
Opto-FosTRAP	mPFC test group	9	-
	SUB test group	10	-
	mPFC control group	5	-
	SUB control group	3	-
Anterograde tracing	mPFC group	15	-
	SUB group	12	-
Opto-behavioral test	mPFC test group	10	-
	mPFC control group	9	-
	SUB test group	9	-
	SUB control group	10	-

derived from non-brain tissues (e.g., muscles) using a data-driven approach tailored for each experimental group. To address the segmentation of non-brain tissue across a large dataset, a deep learning segmentation network was trained to automatically infer muscle positions based on a subset of manually labeled muscle masks. The optimal number of PCs (Fig. S2) was determined by evaluating the Pearson correlation between framewise displacement (FD) and DVARS (temporal derivative of the time course RMS variance over voxels). The preprocessed data were used for tSNR calculation (Fig. S3). The tSNR in the thalamus was about 30 (Fig. S3c).

Each animal underwent three fMRI scanning sessions, each consisting of six runs, for a total of 18 runs per animal. Following this acquisition, all runs were subjected to rigorous, motion-based quality control performed on the fully preprocessed data (including nuisance regression). A run was excluded if the mean FD exceeded 60 μm (half the voxel size) or if the Pearson correlation between FD and DVARS was greater than 0.2. This process led to the exclusion of 33%, 46%, 47%, and 41% of runs in the mPFC test, mPFC control, SUB test, and SUB control groups, respectively. The final number of runs included in the analysis for each group is reported in Table 1.

GLM-based statistical analysis was conducted using the awake mouse-specific hemodynamic response function (HRF) from our previous study[62]. Standard first-level analysis was performed for individual EPI scans. Predictors were created by convolving the experimental stimulus events with the HRF. The observed fMRI data Y were fitted with the design matrix X to solve for the model parameters β. For the second-level analysis, all β maps obtained from the first-level analysis were further transformed to z values to standardize comparisons across different scans. The z maps for each subject were averaged and then entered a second-level group analysis, where the experimental and control groups were distinguished using binary coding (1 represents the experimental group, 0 represents the control group). A conservative, voxel-wise False Discovery Rate (FDR) correction for multiple comparisons ($p < 0.05$) was applied to the group-level statistical maps across the entire brain volume, following SPM12's standard procedures. The resultant activation maps display all voxels that survived this threshold. For the region-based analysis, the t-values reported in Fig. 2b were corrected using a respective region-wise FDR approach at the same significance level ($p < 0.05$).

Topological data analysis (TDA) visualization of BOLD activation

The Mapper algorithm, a method within Topological Data Analysis (TDA), was employed to visualize the activation patterns of mPFC and SUB using the dyneusr package (https://braindynamicslab.github.io/dyneusr/). For dimensionality reduction, we applied PCA and selected the first three principal components to transform whole-brain BOLD signals into a lower-dimensional representation. The resulting low-dimensional space was partitioned into overlapping bins, within which data points were grouped using a clustering algorithm. A topological graph was then constructed, where nodes represented clusters and edges connected nodes with overlapping data points. To enhance representation stability, BOLD time series were averaged across all runs for each mouse. The visualization highlighted pre-stimulation (10 s) and post-stimulation (20 s) periods with different color depths, with mPFC and SUB activation patterns colored in green and blue, respectively.

Optogenetic spatial transcriptomics data acquisition

The study included spatial transcriptomics analysis of 32 samples from 10 individuals (Fig. S8), utilizing 8 Visium slides with 4 capture windows each, including two technical replicates for 6 individuals.

The tissue preparation adhered to the Visium Spatial Gene Expression User Guide (CG00240). The mice were euthanized by rapid decapitation under isoflurane anesthesia 45 minutes after fMRI, followed by rapid dissection and embedding of the brains in OCT compound. The samples were flash-frozen in isopentane chilled in liquid nitrogen for 5 minutes and stored at −80 °C. All the samples were processed within four weeks.

Sectioning was guided by the Visium Spatial Gene Expression Tissue Preparation Guide (CG00240), which targets the rostral thalamus (position 275, bregma −1 mm) and caudal thalamus (position 232, bregma −2.1 mm) of the Allen P56 mouse brain atlas, with a thickness of 10 μm to prevent cell layer overlap. The sections were cryosectioned, placed in sealed boxes, and transferred onto dry ice, followed by subsequent steps on the same day.

Hematoxylin–eosin (H&E) staining and brightfield imaging were performed according to the Visium Spatial Gene Expression H&E Staining & Imaging User Guide (CG000160). Imaging was conducted on an Olympus FV3000 microscope, which captures the slide borders for corner encoding identification. Tissue permeability was optimized following the Visium Spatial Gene Expression Tissue Optimization Guide (CG00238), with an 18-minute permeability time balancing fluorescent signal intensity and diffusion (Fig. S7). Spatial transcriptomics sequencing libraries were constructed according to the Visium Spatial Gene Expression Reagent Kits User Guide (CG00239) and sequenced on the Illumina platform.

Spatial transcriptomics data processing

Spaceranger (version 1.3.0) software from 10x Genomics was used to perform process, alignment, tissue detection, and barcode/UMI counting against the mouse mm10 reference genome for each spot on the Visium spatial transcriptomic array. The raw UMI count matrix, images, spot-image coordinates, and scale factors were imported into the R Seurat package, and we analyzed the spatial transcriptome data following the Seurat vignettes of analysis, visualizatio,n and integration of the 10x Visium datasets. Initially, the regularized negative binomial regression 'SCTransform' function was applied to normalize expression values for total UMI count per spot, which is better than log normalization for counting variability in total spot RNA content.

Table 2 | Summary of habituation schedules for awake mouse optogenetic fMRI

	Day1	Day2	Day3	Day4	Day5	Day6	Day7
Duration of habituation	30 min	40 min	50 min	60 min	60 min	45 min	45 min
Acoustic noise	-	95 db	100 db	110 db	120 db	120 db	120 db
Simulation	-	-	-	-	-	+	+
Position of habituation	Simulated scan	Simulated scan	Simulated scan	Simulated scan	Simulated scan	Real scan	Real scan

"−/+" means "absence or presence."

Optogenetic FosTRAP data acquisition and processing

The FosTRAP2×Ai47 mice used in this study were created by crossing TRAP2 mice with Ai47 reporter mice to label active neurons within a defined time frame. 4-Hydroxytamoxifen (4-OHT) served as the inducer[40]. The 4-OHT solution was prepared by dissolving 15 mg of solution in 500 µL of anhydrous ethanol and shaking at 37 °C and 200 rpm for 30 minutes. Then, 1 mL of corn oil was added, and the mixture was shaken for 50 minutes to extract the 4-OHT. The mixture was left in a fume hood for 45 minutes to evaporate the ethanol.

The mice were intraperitoneally injected with 4-OHT (100 mg/kg) 20 minutes before opto-fMRI to induce genetic recombination effectively and with minimal side effects. Post-opto-fMRI, the mice were returned to their home cages and left undisturbed for three days. Tissue processing and analysis were performed 10 days postinjection.

Continuous sections were obtained from the olfactory bulb to the cerebellum using a vibratome with a thickness of 50 µm. Every fourth section was stained with DAPI (4′,6-diamidino-2-phenylindole) for nuclear visualization at evenly spaced intervals. The tissue sections were scanned using a high-throughput slide scanner (Olympus VS200), with imaging performed using a 10× objective lens.

The quantification of triGFP-labeled cell bodies was achieved using a machine learning-based cell segmentation model trained with the ilastik software platform[90] (Fig. 1f). Custom Python scripts enabled the counting of Fos-positive cells within atlas-defined regions on high-resolution images, preserving spatial data integrity. Aggregated slice counts yielded total Fos-positive cell numbers per brain region, which were normalized by region area to derive cell density, enabling direct regional and subject comparisons. FDR corrections were applied to *p*-values when comparing the FosTRAP data to BOLD activation maps, integrating the molecular and functional data to identify neural activation patterns.

Anterograde axon tracing data acquisition and processing

Guided by the prior projection patterns from Allen Brain Atlas, we performed threshold-based binarization on individual mouse anterograde axon tracing data, retaining the top 4% of voxels as active projections. Visual inspection confirmed that this threshold effectively minimized background noise and nonspecific signals (Fig. 1f). For group-level analyses, projection patterns were first normalized within each mouse to account for scanner variability, then averaged across mice within mPFC and SUB groups, respectively. To address slice correspondence discrepancies with the CCF v3 template across mice, spatial smoothing with an FWHM of 0.3 mm was applied to ensure data continuity and consistency.

Registration of tissue slices to the Allen brain atlas

We developed a semi-automated pipeline for registering mouse brain slices to the Allen Brain Atlas template for histological analysis, FosTRAP quantification, and projection mapping. The registration process began with rotating the Allen template to match the approximate angle of experimental sections[91,92]. We then performed manual one-to-one matching between experimental slices and the Allen template, using prominent anatomical structures such as the anterior commissure, hippocampus, and amygdala as reference points. After establishing these slice correspondences, we applied the ANTs SyN algorithm for both linear and nonlinear registration to optimize alignment and compensate for morphological variations between samples and the template. To address potential registration artifacts in challenging cases, we implemented landmark-based registration using a combination of SIFT-detected feature points and manually verified landmarks, followed by targeted manual refinements using the BigWarp plugin in FIJI for regions with residual misalignment.

Spot-level registration of multimodal data

The two representative 10-µm Visium H&E-stained slices (rostral and caudal) were selected as the "best fit" based on optimal anatomical preservation and image quality. These slices were registered to the Allen CCF using anatomical landmarks to establish a reference framework. Subsequent alignment of additional spatial transcriptomic slices to this reference was performed using the SLAT algorithm, which leverages spatial graph convolution and adversarial matching for robust, unsupervised alignment of heterogeneous samples, thereby minimizing registration errors across datasets[93].

The scSLAT package (version 0.2.1) was used for all slice-to-slice registrations. The Cal_Spatial_Net function constructed a spatial neighborhood network for each spot using KNN with a cutoff of 10, followed by model building and training via run_SLAT using an SVD-based strategy. The spatial_match function then computed a probabilistic matching matrix, with Euclidean distances filtered at a threshold of 3000 to achieve one-to-one spot mapping. This process enabled the integration of gene expression, FosTRAP, MRI, and structural projection data into a unified spatial framework, establishing a consistent foundation for multimodal analysis.

Identification of common differentially expressed genes

The differentially expressed genes between the control and test samples from each spot were computed with the Seurat function FindMarkers (https://satijalab.org/seurat/reference/findmarkers). Only genes with adjusted *p*-values smaller than 0.05 were considered differentially expressed genes. The DEGs were classified into the following four categories: rapid primary response genes, delayed primary response genes, secondary response genes, and other genes[38].

Spatial correlation analysis between gene expression and fMRI activation patterns

To quantify the association between transcriptomic changes and functional activation patterns, we performed spatial correlation analyses between group-level gene expression change maps and group-level opto-fMRI activation contrast maps in the mPFC and SUB groups.

For each brain region (mPFC or SUB), analyses were restricted to animals belonging to the ST subgroup, consistent with all fMRI analyses presented in the main figures. Group-level gene expression change patterns were constructed as contrast maps, calculated as spot-level gene expression differences between the ST subgroup and the corresponding control condition, yielding one spatial expression map per gene. Similarly, group-level fMRI activation patterns were derived as optogenetically specific contrast maps, computed by subtracting the matched control condition from the ST subgroup at each

voxel, thereby isolating stimulation-specific BOLD responses. These contrast maps were projected onto the same spatial framework as the gene expression maps.

Spatial similarity between each gene expression map and the corresponding fMRI activation contrast map was quantified using Pearson's correlation coefficient, computed across all shared spatial locations. This resulted in one correlation coefficient per gene for the mPFC and SUB groups, respectively.

To assess statistical significance while accounting for spatial autocorrelation, nonparametric permutation testing was employed. Specifically, 10,000 spatial surrogate fMRI activation maps were generated for each region using a spatial permutation procedure that preserves the intrinsic spatial structure of the original activation contrast map. For each gene, correlation coefficients were recomputed between the gene expression map and each surrogate activation map to construct a null distribution. Two-sided permutation *p*-values were calculated as the proportion of surrogate correlations whose absolute values exceeded that of the observed correlation.

Permutation-based *p*-values ($p < 0.05$, two-sided) were used to identify genes exhibiting significant spatial correspondence with fMRI activation patterns and to generate Fig. 4a and 4d. To further assess robustness under stringent multiple-comparison control, permutation-derived *p*-values were additionally corrected across all genes using the Benjamini–Hochberg false discovery rate (FDR) procedure, and FDR-corrected results are reported in Supplementary Data 2.

PLS modeling

We utilized partial least squares (PLS) regression via the *plsr* function in the R package pls (version 2.8.2) to explore the correlation and impact between BOLD activation and spatial gene expression. The BOLD activation map served as the independent variable, with differential gene expression maps, direct projection maps, and the FosTRAP map as the dependent variables. Data normalization was performed using decimal scaling prior to PLS regression.

Leave-one-out cross-validation (LOOCV) was employed to evaluate the predictive performance of our model. The predictive accuracy across all iterations was averaged to yield the final accuracy estimate. To ascertain the optimal number of PLS components, a permutation test with 10,000 iterations, incorporating spatial-autocorrelation correction, was conducted. Within each fold of LOOCV, the variable importance in projection (VIP)[47] score was computed to assess the contribution of each variable—encompassing genes, FosTRAP, and projection data—across different components. Variables with VIP scores greater than 1.5 were deemed significant and were retained for further analysis.

The intersection of variables with VIP scores exceeding 1.5 across all folds was identified, and these variables were designated the predictive markers in the final model. A subsequent 10,000 permutation test was performed to determine the optimal number of PLS components for the final model. The gene weights derived from each PLS component were subsequently utilized for gene enrichment analysis. We also explored the incorporation of overlapping genes (mPFC: 1868 genes, SUB: 1507 genes) resulting from lowering the VIP score threshold to 1 into the final model and found that the spatial patterns of the PLS components were similar to the current patterns (Fig. S13).

Gene enrichment analysis

Gene enrichment analysis was performed using Metascape (https://metascape.org)[94] and with Benjamini–Hochberg adjusted *p* values (FDRs) < 0.05. Top weighted genes in PLS components analyzed and visualized by ClueGO[95], $p < 0.05$ after FDR correction.

Data-driven activation mapping using principal component analysis (PCA)

To complement the hypothesis-driven GLM analysis, a data-driven PCA was performed on the preprocessed fMRI data from all optogenetic stimulation runs in the test groups. For each subject, the voxel-wise time series from all runs were concatenated and decomposed using PCA implemented in *Scikit-learn*. The first principal component (PC1), which captures the dominant pattern of co-varying signal variance across the brain, was extracted for each animal. These individual PC1 maps were then normalized and used in a second-level group analysis to create a group-average PCA activation map (Supplementary Fig. S15). This model-free map was subsequently used in the PLS analysis alongside the GLM-derived maps to establish its spatial relationship with the transcriptomic data.

Optogenetic behavioral tests

Behavioral assessments included the open field test, Y-maze spontaneous alternation test, Y-maze novel-arm test, and novel object recognition test. All experiments were performed under optogenetic stimulation, with stimulation parameters matched to those used in the opto-fMRI experiments. Mice were assigned to either the test or control group.

In the open field test, mice were placed in a 40 × 40 cm arena and allowed to explore freely for 10 min. The center zone was defined as a 20 × 20 cm square. In the Y-maze spontaneous alternation test, mice freely explored the maze for 10 min. In the novel-arm test, one arm of the Y-maze was blocked, and mice were allowed to explore for 5 min. Two hours later, the barrier was removed, and the mice were reintroduced from the same starting arm for an additional 5 min exploration.

In the novel object recognition test, mice first freely explored the open field for 10 min. Two identical objects were then placed in the apparatus (10 cm from each side wall), and mice were introduced facing away from the objects and allowed to explore for 5 min. After 1 h, one of the objects was replaced with a novel object differing in both shape and color; mice were again introduced facing away from the objects and allowed to explore for 5 min. Exploration events were manually counted and defined as instances in which the mouse oriented its head toward the object within 1 cm and made contact or investigative movements. The novel index was calculated as the proportion of novel-object exploration events relative to the total number of exploration events.

Locomotor trajectories were recorded using Smart v3.0, and parameters including total distance traveled, center-zone distance, spontaneous alternation rate, and the proportion of time spent in the novel arm were extracted. Statistical analyses were performed in GraphPad Prism 10.4.1 using unpaired two-sample *t*-tests for group comparisons.

Measurement of light spread in fresh brain tissue

Light transmission measurements were performed following established protocols[96,97], using acute brain slices obtained from C57BL/6 N mice (N = 2). Coronal brain slices with thicknesses ranging from 200 μm to 1 mm were prepared using a vibratome (VT1200S, Leica Microsystems). Each brain slice was carefully placed on a thin glass coverslip positioned directly above the sensor of an optical power meter (Thorlabs). A 200-μm-diameter optical fiber (NA = 0.73) coupled to a 473-nm blue laser was positioned perpendicular to the tissue surface and aligned coaxially with the detector. The fiber tip was placed directly above the cortical tissue. Blue laser light (15 mW output at the fiber tip) was delivered through the optical fiber, and the total transmitted optical power was recorded by the power meter. For each brain slice, two independent measurements were obtained. For each tissue thickness, measurements were performed on three different slices. To estimate the illuminated volume in the optogenetic configuration, a simplified geometrical model was applied, approximating light propagation as a conical frustum based on the optical fiber numerical aperture. Using this model, the light intensity ratio as a function of tissue depth was calculated. Light transmission through

the brain tissue was subsequently modeled using an exponential attenuation function.

Reporting summary

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

Data availability

The spatial transcriptomics data generated in this study have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE284625. The fMRI data generated in this study are under active use by the reporting laboratory; all the raw data that support the findings of this study are available from the corresponding authors upon request. The data for individual mouse thalamic BOLD activation intensities, average time series associated with mPFC and SUB neural activation, and gene enrichment analysis results associated with GLM- and PCA-based BOLD activation patterns in the main figures and Supplementary Information are provided with this paper in the Source Data file. Source data are provided with this paper.

Code availability

All original code has been deposited at GitHub (https://github.com/wangxiaoqun-lab/opto_fMRI_ST.git) and is publicly available.

References

1. Fornito, A., Arnatkevičiūtė, A. & Fulcher, B. D. Bridging the Gap between Connectome and Transcriptome. *Trends Cogn. Sci.* **23**, 34–50 (2019).
2. Bazinet, V., Hansen, J. Y. & Misic, B. Towards a biologically annotated brain connectome. *Nat. Rev. Neurosci.* **24**, 747–760 (2023).
3. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
4. Arnatkeviciute, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353–367 (2019).
5. Larivière, S. et al. BrainStat: A toolbox for brain-wide statistics and multimodal feature associations. *NeuroImage* **266**, 119807 (2023).
6. Markello, R. D. et al. Standardizing workflows in imaging transcriptomics with the abagen toolbox. *eLife* **10**, e72129 (2021).
7. Rizzo, G., Veronese, M., Expert, P., Turkheimer, F. E. & Bertoldo, A. MENGA: A new comprehensive tool for the integration of neuroimaging data and the allen human brain Transcriptome Atlas. *PLOS ONE* **11**, e0148744 (2016).
8. Krienen, F. M., Yeo, B. T. T., Ge, T., Buckner, R. L. & Sherwood, C. C. Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. *Proc. Natl. Acad. Sci. USA* **113**, E469–E478 (2016).
9. Zhang, X.-H. et al. The cell-type underpinnings of the human functional cortical connectome. *Nat. Neurosci.* 1–11 (2024).
10. Yu, H. et al. Morphological connectivity differences in Alzheimer's disease correlate with gene transcription and cell-type. *Hum. Brain Mapp.* **44**, 6364–6374 (2023).
11. Wang, W. et al. Noncoding transcripts are linked to brain resting-state activity in non-human primates. *Cell Rep.* **42**, 112652 (2023).
12. Bo, T. et al. Brain-wide and cell-specific transcriptomic insights into MRI-derived cortical morphology in macaque monkeys. *Nat. Commun.* **14**, 1499 (2023).
13. Xia, M. et al. Connectome gradient dysfunction in major depression and its association with gene expression profiles and treatment outcomes. *Mol. Psychiatry* **27**, 1384–1393 (2022).
14. Thomas, G. E. C. et al. Regional brain iron and gene expression provide insights into neurodegeneration in Parkinson's disease. *Brain* **144**, 1787–1798 (2021).
15. Vogel, J. W. et al. A molecular gradient along the longitudinal axis of the human hippocampus informs large-scale behavioral systems. *Nat. Commun.* **11**, 960 (2020).
16. Marshel, J. H. et al. Cortical layer–specific critical dynamics triggering perception. *Science* **365**, eaaw5202 (2019).
17. Vértes, 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).
18. Christie, I. N. et al. fMRI response to blue light delivery in the naïve brain: Implications for combined optogenetic fMRI studies. *NeuroImage* **66**, 634–641 (2013).
19. Seidlitz, J. et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. *Nat. Commun.* **11**, 3358 (2020).
20. Roy, D. S., Zhang, Y., Halassa, M. M. & Feng, G. Thalamic subnetworks as units of function. *Nat. Neurosci.* **25**, 140–153 (2022).
21. Shine, J. M., Lewis, L. D., Garrett, D. D. & Hwang, K. The impact of the human thalamus on brain-wide information processing. *Nat. Rev. Neurosci.* **24**, 416–430 (2023).
22. Kapustina, M. et al. The cell-type-specific spatial organization of the anterior thalamic nuclei of the mouse brain. *Cell Rep.* **43**, 113842 (2024).
23. Guillery, R. W. & Sherman, S. M. Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system. *Neuron* **33**, 163–175 (2002).
24. Aggleton, J. P. & O'Mara, S. M. The anterior thalamic nuclei: core components of a tripartite episodic memory system. *Nat. Rev. Neurosci.* **23**, 505–516 (2022).
25. Toader, A. C. et al. Anteromedial thalamus gates the selection and stabilization of long-term memories. *Cell* **186**, 1369–1381.e17 (2023).
26. Vetere, G. et al. An inhibitory hippocampal–thalamic pathway modulates remote memory retrieval. *Nat. Neurosci.* **24**, 685–693 (2021).
27. Ma, J. et al. Convergent direct and indirect cortical streams shape avoidance decisions in mice via the midline thalamus. *Nat. Commun.* **15**, 6598 (2024).
28. Aquino-Miranda, G. et al. Functional properties of corticothalamic circuits targeting paraventricular thalamic neurons. *Neuron* **112**, 4060–4080.e7 (2024).
29. Choi, E. A. et al. A Corticothalamic circuit trades off speed for safety during decision-making under motivational conflict. *J. Neurosci.* **42**, 3473–3483 (2022).
30. Giannotti, G., Barry, S. M., Siemsen, B. M., Peters, J. & McGinty, J. F. Divergent Prelimbic Cortical Pathways Interact with BDNF to Regulate Cocaine-seeking. *J. Neurosci.* **38**, 8956–8966 (2018).
31. Otis, J. M. et al. Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. *Nature* **543**, 103–107 (2017).
32. Do-Monte, F. H., Quiñones-Laracuente, K. & Quirk, G. J. A temporal shift in the circuits mediating retrieval of fear memory. *Nature* **519**, 460–463 (2015).
33. Chen, X., Cramer, S. R., Chan, D. C. Y., Han, X. & Zhang, N. Sequential Deactivation Across the Hippocampus-Thalamus-mPFC Pathway During Loss of Consciousness. *Adv. Sci.* **11**, 2406320 (2024).
34. Belleau, E. L., Treadway, M. T. & Pizzagalli, D. A. The impact of stress and major depressive disorder on hippocampal and medial prefrontal cortex morphology. *Biol. Psychiatry* **85**, 443–453 (2019).
35. Roland, A. V. et al. Alcohol Dependence Modifies Brain Networks Activated During Withdrawal and Reaccess: A c-Fos–Based Analysis in Mice. *Biol. Psychiatry* **94**, 393–404 (2023).
36. Lu, H. et al. Spatial correlations of laminar BOLD and CBV responses to rat whisker stimulation with neuronal activity localized by Fos expression. *Magn. Reson. Med.* **52**, 1060–1068 (2004).

37. Yang, P.-F. et al. Comparison of fMRI BOLD Response Patterns by Electrical Stimulation of the Ventroposterior Complex and Medial Thalamus of the Rat. *PLOS ONE* **8**, e66821 (2013).
38. Tyssowski, K. M. et al. Different Neuronal Activity Patterns Induce Different Gene Expression Programs. *Neuron* **98**, 530–546.e11 (2018).
39. Yap, E.-L. & Greenberg, M. E. Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron* **100**, 330–348 (2018).
40. Guenthner, C. J., Miyamichi, K., Yang, H. H., Heller, H. C. & Luo, L. Permanent genetic access to transiently active neurons via TRAP: targeted recombination in active populations. *Neuron* **78**, 773–784 (2013).
41. Nelson, E. D. et al. Activity-regulated gene expression across cell types of the mouse hippocampus. *Hippocampus* **33**, 1009–1027 (2023).
42. Marco, A. et al. Mapping the epigenomic and transcriptomic interplay during memory formation and recall in the hippocampal engram ensemble. *Nat. Neurosci.* **23**, 1606–1617 (2020).
43. Hu, Y. et al. Comparative brain-wide mapping of ketamine- and isoflurane-activated nuclei and functional networks in the mouse brain. *eLife* **12**, RP88420 (2024).
44. Shrestha, N. Detecting Multicollinearity in Regression Analysis. *Am. J. Appl. Math. Stat.* **8**, 39–42 (2020).
45. Chan, J. Y.-L. et al. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics* **10**, 1283 (2022).
46. Streukens, S. & Leroi-Werelds, S. Multicollinearity: An Overview and Introduction of Ridge PLS-SEM Estimation. *Methodol. Issues Applications* 183–207 (2023).
47. Wold, S., Sjöström, M. & Eriksson, L. Partial Least Squares Projections to Latent Structures (PLS) in Chemistry. In *Encyclopedia of Computational Chemistry* (John Wiley & Sons, Ltd, 2002).
48. Deng, S. et al. Hemodynamic and metabolic correspondence of resting-state voxel-based physiological metrics in healthy adults. *Neuroimage* **250**, 118923 (2022).
49. Fu, C. et al. A combined study of 18F-FDG PET-CT and fMRI for assessing resting cerebral function in patients with major depressive disorder. *Exp. Ther. Med.* **16**, 1873–1881 (2018).
50. Bernier, M., Croteau, E., Castellano, C.-A., Cunnane, S. C. & Whittingstall, K. Spatial distribution of resting-state BOLD regional homogeneity as a predictor of brain glucose uptake: A study in healthy aging. *NeuroImage* **150**, 14–22 (2017).
51. Aiello, M., Cavaliere, C. & Salvatore, M. Hybrid PET/MR Imaging and Brain Connectivity. *Front. Neurosci.* **10**, (2016).
52. Tomasi, D., Wang, G.-J. & Volkow, N. D. Energetic cost of brain functional connectivity. *Proc. Natl. Acad. Sci. USA* **110**, 13642–13647 (2013).
53. Fulcher, B. D. & Fornito, A. A transcriptional signature of hub connectivity in the mouse connectome. *Proc. Natl. Acad. Sci.* **113**, 1435–1440 (2016).
54. Kelvington, B. A. & Abel, T. Innate immunity in neurons makes memories persist. *Nature* **628**, 40–42 (2024).
55. Jovasevic, V. et al. Formation of memory assemblies through the DNA-sensing TLR9 pathway. *Nature* **628**, 145–153 (2024).
56. Decot, H. K. et al. Coordination of Brain-Wide Activity Dynamics by Dopaminergic Neurons. *Neuropsychopharmacol* **42**, 615–627 (2017).
57. Ferenczi, E. A. et al. Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science* **351**, aac9698 (2016).
58. Cover, C. G. et al. Whole brain dynamics during optogenetic self-stimulation of the medial prefrontal cortex in mice. *Commun. Biol.* **4**, 66 (2021).
59. Liang, Z. et al. Mapping the functional network of medial prefrontal cortex by combining optogenetics and fMRI in awake rats. *NeuroImage* **117**, 114–123 (2015).
60. Zou, Y. et al. Cell-type-specific optogenetic fMRI on basal forebrain reveals functional network basis of behavioral preference. *Neuron* **112**, 1342–1357.e6 (2024).
61. Schaeffer, S. & Iadecola, C. Revisiting the neurovascular unit. *Nat. Neurosci.* **24**, 1198–1209 (2021).
62. Chen, X. et al. Sensory evoked fMRI paradigms in awake mice. *Neuroimage* **204**, 116242 (2020).
63. Richiardi, J. et al. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241 (2015).
64. Forest, M. et al. Gene networks show associations with seed region connectivity. *Hum. Brain Mapp.* **38**, 3126–3140 (2017).
65. Zhao, H. et al. Genetic mechanisms underlying brain functional homotopy: a combined transcriptome and resting-state functional MRI study. *Cereb. Cortex* **33**, 3387–3400 (2023).
66. Goel, P., Kuceyeski, A., LoCastro, E. & Raj, A. Spatial patterns of genome-wide expression profiles reflect anatomic and fiber connectivity architecture of healthy human brain. *Hum. Brain Mapp.* **35**, 4204–4218 (2014).
67. Whitaker, K. J. et al. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proc. Natl. Acad. Sci. USA* **113**, 9105–9110 (2016).
68. Kang, X. et al. Convergent Neuroimaging and Molecular Signatures in Mild Cognitive Impairment and Alzheimer's Disease: A Data-Driven Meta-Analysis with N = 3,118. *Neurosci. Bull.* **40**, 1274–1286 (2024).
69. Romme, I. A. C., de Reus, M. A., Ophoff, R. A., Kahn, R. S. & van den Heuvel, M. P. Connectome Disconnectivity and Cortical Gene Expression in Patients With Schizophrenia. *Biol. Psychiatry* **81**, 495–502 (2017).
70. McColgan, P. et al. Brain Regions Showing White Matter Loss in Huntington's Disease Are Enriched for Synaptic and Metabolic Genes. *Biol. Psychiatry* **83**, 456–465 (2018).
71. Romero-Garcia, R., Warrier, V., Bullmore, E. T., Baron-Cohen, S. & Bethlehem, R. A. I. Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. *Mol. Psychiatry* **24**, 1053 (2018).
72. Esteban, O. et al. Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nat. Protoc.* **15**, 2186–2202 (2020).
73. Monti, M. M. Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. *Front. Hum. Neurosci.* **5**, (2011).
74. Gao, Y.-R. et al. Time to wake up: Studying neurovascular coupling and brain-wide circuit function in the un-anesthetized animal. *Neuroimage* **153**, 382–398 (2017).
75. Shim, H.-J. et al. Mouse fMRI under ketamine and xylazine anesthesia: Robust contralateral somatosensory cortex activation in response to forepaw stimulation. *NeuroImage* **177**, 30–44 (2018).
76. Franceschini, M. A. et al. The effect of different anesthetics on neurovascular coupling. *NeuroImage* **51**, 1367–1377 (2010).
77. Yu, Y. et al. Sleep fMRI with simultaneous electrophysiology at 9.4 T in male mice. *Nat. Commun.* **14**, 1651 (2023).
78. Han, Z. et al. Awake and behaving mouse fMRI during Go/No-Go task. *Neuroimage* **188**, 733–742 (2019).
79. Xu, W. et al. A systematically optimized awake mouse fMRI paradigm. *bioRxiv* 2022.11.16.516376 https://doi.org/10.1101/2022.11.16.516376 (2022).
80. Liang, Z., Li, T., King, J. & Zhang, N. Mapping thalamocortical networks in rat brain using resting-state functional connectivity. *Neuroimage* **83**, 237–244 (2013).
81. Tsurugizawa, T. & Yoshimaru, D. Impact of anesthesia on static and dynamic functional connectivity in mice. *NeuroImage* **241**, 118413 (2021).

82. Paasonen, J., Stenroos, P., Salo, R. A., Kiviniemi, V. & Gröhn, O. Functional connectivity under six anesthesia protocols and the awake condition in rat brain. *NeuroImage* **172**, 9–20 (2018).
83. King, J. A. et al. Procedure for minimizing stress for fMRI studies in conscious rats. *J. Neurosci. Methods* **148**, 154–160 (2005).
84. Owen, S. F., Liu, M. H. & Kreitzer, A. C. Thermal constraints on in vivo optogenetic manipulations. *Nat. Neurosci.* **22**, 1061–1065 (2019).
85. Yang, C. et al. Medial prefrontal cortex and anteromedial thalamus interaction regulates goal-directed behavior and dopaminergic neuron activity. *Nat. Commun.* **13**, 1386 (2022).
86. Liu, J. et al. Astrocyte dysfunction drives abnormal resting-state functional connectivity in depression. *Sci. Adv.* **8**, eabo2098 (2022).
87. Riedemann, T. Diversity and Function of Somatostatin-Expressing Interneurons in the Cerebral Cortex. *Int J. Mol. Sci.* **20**, 2952 (2019).
88. Zingg, B. et al. Neural networks of the mouse neocortex. *Cell* **156**, 1096–1111 (2014).
89. Vertes, R. P. Differential projections of the infralimbic and prelimbic cortex in the rat. *Synapse* **51**, 32–58 (2004).
90. Berg, S. et al. ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
91. Gong, R., Xu, S., Hermundstad, A., Yu, Y. & Sternson, S. M. Hindbrain Double-Negative Feedback Mediates Palatability-Guided Food and Water Consumption. *Cell* **182**, 1589–1605.e22 (2020).
92. Xiong, J., Ren, J., Luo, L. & Horowitz, M. Mapping Histological Slice Sequences to the Allen Mouse Brain Atlas Without 3D Reconstruction. *Front Neuroinform.* **12**, 93 (2018).
93. Xia, C.-R., Cao, Z.-J., Tu, X.-M. & Gao, G. Spatial-linked alignment tool (SLAT) for aligning heterogenous slices. *Nat. Commun.* **14**, 7236 (2023).
94. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
95. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
96. Leong, A. T. L. et al. Optogenetic fMRI interrogation of brain-wide central vestibular pathways. *Proc. Natl. Acad. Sci. USA* **116**, 10122–10129 (2019).
97. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K.Optogenetics in neural systems. *Neuron***71**,9–34 (2011).

Acknowledgements

We thank the Beijing Magnetic Resonance Imaging Center at the Institute of Biophysics for technical assistance. This work was supported by Science and Technology Innovation 2030—Brain Science and Brain-inspired Intelligence Project of China (2022ZD0211900, 2022ZD0207700, 2021ZD0200102), the National Natural Science Foundation of China (NSFC) (32192411, 82501714), CAS Project for Young Scientists in Basic Research (YSBR-013, YSBR-097), the New Cornerstone Science Foundation, Changping Laboratory, the Fundamental Research Funds for the Central Universities.

Author contributions

X.W. conceived and supervised the project and designed the experiments. X.W., Z.W., A.L., L.S., and Q.W. discussed the results. C.Z. carried out the virus injection and animal training. Y.Z. and C.Z performed the MRI experiments. Y.Z. and X.L. analyzed MRI data. S.Z. and M.L. performed spatial RNA-seq. L.Z., M.W., and B.Z. analyzed the data. W.W. and C.Z. maintain the animals. C.Z., C.D., and X.Z. analyzed Fos images. Y.Z. worked on the PLS modeling. Y.Z., C.Z., Q.W., and X.W. wrote the manuscript, and C.Z., Y.Z., L.Z., X.L., S.Z., C.Y., M.L., M.W., W.W., X.Z., B.Z., C.D., L.S., Z.W., A.L., Q.W., and X.W. edited and proofed the manuscript. C.Z., Y.Z., L.Z., X.L., and S.Z. are co-first authors and contributed equally to this work.

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/s41467-026-71923-w.

Correspondence and requests for materials should be addressed to Le Sun, Zheng Wang, Ang Li, Qian Wu or Xiaoqun Wang.

Peer review information *Nature Communications* thanks Ed Wu, who co-reviewed with Linshan Xie, and Ruiqi Wu for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions 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.

[1]State Key Laboratory of Cognitive Science and Mental Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. [2]Changping Laboratory, Beijing, China. [3]University of Chinese Academy of Sciences, Beijing, China. [4]IDG/McGovern Institute for Brain Research, New Cornerstone Science Laboratory, Beijing Normal University, Beijing, China. [5]Beijing Institute of Brain Disorders, Capital Medical University, Beijing, China. [6]School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, State Key Laboratory of General Artificial Intelligence, IDG/McGovern Institute for Brain Research, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [7]These authors contributed equally: Changjiang Zhang, Yijuan Zou, Liangchen Zhuo, Xiya Liu, Suijuan Zhong. ✉e-mail: sunle@ibp.ac.cn; zheng.wang@pku.edu.cn; al@ibp.ac.cn; qianwu@bnu.edu.cn; xiaoqunwang@ibp.ac.cn