Article

Developmental xenocortication using human-derived organoids in mice

https://doi.org/10.1038/s41586-026-11032-2

Received: 27 June 2025

Accepted: 12 August 2026

Published online: 16 September 2026

Open access

Check for updates

Konstantin Kaganovsky[1,2,9], Kevin W. Kelley[1,2,9], Tilo Gschwind[3,9], Paul M. Harary[1,2,9], John Kochalka[1], Alexander D. White[4], Garikoitz Lerma-Usabiaga[5,6], Xiaoyu Chen[1,2], Omer Revah[1,2], Felicity Gore[1,4], Ayano Aoyama[1], Jennifer L. Shadrach[3], Se-Jin Yoon[1,2], Alfredo Valencia[1,2], Satoe Ogawa[1,2], Noah Reis[1,2], Hannes Vogel[7], Brian Wandell[8], Julia A. Kaltschmidt[3], Ivan Soltesz[3], Karl Deisseroth[1,4] & Sergiu P. Pașca[1,2] ✉

The inaccessibility of human brain tissue limits the study of human development and function, a challenge that human stem-cell-derived neural models are beginning to address[1,2]. Transplantation of neural organoids into rodent hosts enables the in vivo study of aspects of human neurodevelopment and circuit function, alongside behavioural phenotyping of the host animals. However, spatial limitations and competition with host circuits constrain the integration of neural organoids, which is critical for studying disease. Here we establish a transplantation platform using a genetic strategy to effectively deplete glutamatergic neurons from mouse neocortex and hippocampus (apallial) and neonatally engraft the cortical cavity with human stem-cell-derived cortical organoids (hCO) to generate xenocortical mice. This leads to robust graft growth with hCOs occupying most of the cortical volume and generating a diversity of human cortical cell types, including layer 5 extratelencephalic projection neurons. Human cortical neurons integrate with the mouse nervous system, and in vivo cortical graft-wide calcium imaging and electrophysiological analyses revealed patterns of organized activity resembling developing circuits. Behavioural analyses of apallial and xenocortical mice revealed broadly preserved locomotion alongside selective differences in limb coordination and altered organization of spontaneous behaviour. Lastly, this platform enabled behavioural readouts in a model of injury to developing human cortical cells. We envision that xenocortication will be useful for obtaining circuit- and behaviour-level readouts using human neurons to study neurodevelopment, model disease and develop therapeutics.

Human stem-cell-based models of the nervous system hold promise to examine previously inaccessible aspects of human brain development, reveal mechanisms of disease and identify evolutionary differences between species. For example, guiding human induced pluripotent stem (hiPS) cells to generate three-dimensional (3D) hCOs recapitulates aspects of neurogenesis and gliogenesis of the cerebral cortex in vitro, including the sequential generation of specialized neuronal cell types[1,2]. Transplantation of hCOs (t-hCO) into the newborn rat cortex results in integration and enhanced maturation of human neurons[3], but physical space constraints within the intracranial cavity and competition with the host cortex for establishing connectivity limit their potential. While transplanted hCOs can occupy up to a third of a rat's cortical hemisphere[3], the human-derived neurons develop in the context of extant rat connections. Rodent neurons develop faster than human neurons, so the rapidly maturing host environment may constrain graft proliferation and neural projections. This limitation is particularly salient when modelling neurodevelopmental disorders associated with large-scale distributed circuit dysregulation and altered behavioural output or when testing therapeutics on human neurons in vivo.

To overcome these constraints, we generated an immunodeficient apallial mouse in which the dorsal and medial pallium is genetically depleted at early stages of development. Anatomical studies, magnetic resonance imaging (MRI) and whole-brain single-nucleus RNA-sequencing (snRNA-seq) analysis confirmed that apallial mice lack dorsal and medial pallium derivatives, including neocortex and hippocampus. Transplantation into apallial mice, termed xenocortication, resulted in the growth of a large volume of human-derived neural tissue that recapitulates the cellular diversity of the developing human cerebral cortex. In these xenocortical (XCX) mice, hCOs integrate within the host central nervous system and exhibit spontaneous electrical activity. Unsupervised machine learning applied to spontaneous mouse behaviour using motion sequencing (MoSeq) revealed

[1]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [2]Stanford Brain Organogenesis Program, Wu Tsai Neurosciences Institute & Bio-X, Stanford, CA, USA. [3]Department of Neurosurgery, Stanford University, Stanford, CA, USA. [4]Department of Bioengineering, Stanford University, Stanford, CA, USA. [5]BCBL—Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain. [6]IKERBASQUE—Basque Foundation for Science, Bilbao, Spain. [7]Department of Pathology, Stanford University, Stanford, CA, USA. [8]Department of Psychology, Stanford University, Stanford, CA, USA. [9]These authors contributed equally: Konstantin Kaganovsky, Kevin W. Kelley, Tilo Gschwind, Paul M. Harary. ✉e-mail: spasca@stanford.edu

distinct behavioural repertoires in control and apallial mice, with the XCX group positioned intermediate between, yet distinct from, both the control and apallial groups. In support of the MoSeq data, targeted behavioural tasks demonstrated minimal gross impairments in the apallial and XCX conditions. However, apallial mice showed deficits on tasks requiring neocortex and hippocampus (fine motor coordination in the CatWalk and working memory in the Y maze), a pattern that was only partially observed in XCX mice. Moreover, hypoxia induced a cellular response along with deficits in the CatWalk procedure in XCX mice, underscoring the potential of this in vivo platform to facilitate disease modelling and therapeutic discovery using large human-derived cortical grafts and behavioural outputs.

Here we used a conditional-knockout strategy to deplete the majority of the mouse cerebral cortex through targeted cell death. We deleted the sister chromatid cohesion factor *Esco2* specifically in cells expressing the dorsal and medial pallium marker *Emx1* in an immunocompromised SCID background[4–6] (apallial mice: *Emx1-cre*$^{+/-}$*Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$; Fig. 1a and Extended Data Fig. 1a). A previous study using *Emx1-cre*$^{+/-}$*Esco2*$^{fl/fl}$ mice showed that early postnatal apallial mice emit largely intact ultrasonic vocalizations[5], yet viability and behavioural alterations beyond this early timepoint were not explored. To enable functional and behavioural studies in adulthood after engraftment and integration of human cells, we implemented several strategies to optimize mouse viability: reducing the potential for germline recombination in offspring through exclusion of male *Emx1-cre*$^{+/-}$ breeders[7], reducing competition for resources within the litter through culling the majority of pups with intact cortex, delaying weaning of apallial pups, supplementing the breeding and offspring cages with high-caloric food and introducing an 'aunting' procedure with Swiss Webster active dams. Our updated breeding and rearing protocol yielded pups with similar body weight (Extended Data Fig. 1c) and 23.4% apallial offspring (n = 91 litters, 46 breeding cages; Extended Data Fig. 1a,b), consistent with the 25% phenotypic ratio predicted by Mendelian inheritance. While the survival rate was not different between groups, both breeders and apallial mice demonstrated a trend towards lower survival at around 2 months of age (Extended Data Fig. 1d). Closer morphological analysis, MRI and immunostaining demonstrated that subcortical structures formed the dorsal surface of the brain and cortical structures were severely depleted (Fig. 1b,c and Extended Data Fig. 1e,f). Indeed, whole-brain MRI revealed an approximately 50% decrease in total brain tissue relative to the control *Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$ mice (Fig. 1d and Extended Data Fig. 1f), consistent with estimates that mouse cortex makes up about 42% of brain mass[8]. To comprehensively characterize the brain-wide cell-type composition, we performed whole-brain snRNA-seq analysis of apallial mice and matched controls at 5 months of age (n = 2 apallial; n = 2 controls). After stringent quality filtering, we obtained 880,149 single-nucleus profiles. We annotated snRNA-seq profiles by transcriptomic mapping to the mouse Allen Brain Cell Atlas[9], which yielded consistent annotations across animals (Fig. 1e,f and Extended Data Figs. 1i–k and 2a,c). We observed a sevenfold depletion of all dorsal pallial-derived glutamatergic neuronal cell classes and preservation of olfactory bulb neuronal classes and other non-pallial derived cell classes (Fig. 1f,g, Extended Data Fig. 1i and Supplementary Table 1). At a higher subclass resolution within depleted dorsal pallial-derived glutamatergic neuronal cell classes, we observed broad depletion of glutamatergic neuronal subclasses from the neocortex and hippocampus (Fig. 1h,i, Extended Data Fig. 2b and Supplementary Table 2). Consistent with previous reports[5,6], ventral/lateral glutamatergic pallial neuronal subclasses were preserved, including those from amygdala and piriform cortex (palaeocortex; Fig. 1h,i and Extended Data Fig. 1j). We nonetheless retain the term apallial as an operational designation for this model, recognizing that it does not denote complete elimination of all pallial derivatives. Cortical GABAergic neuron classes showed the next largest reductions (Fig. 1g and Extended Data Fig. 2d). Given the reported dependence on local pyramidal neuron signalling for survival[10,11], we performed differential expression analysis across cortical GABAergic subclasses and found minimal transcriptional changes (Extended Data Fig. 2e), suggesting that surviving cortical GABAergic neurons do not substantially alter their fate in the apallial model. These results were robust across animals and broadly consistent with annotations obtained using a distinct transcriptomic mapping algorithm and alternative mouse brain atlas[12] (Extended Data Fig. 2b).

Having established effective cortical depletion in the apallial mouse model, we next tested whether hCOs would grow and integrate within the cortical cavity after transplantation. We transplanted hCOs into the apallial mouse, generating a large cortex-wide XCX graft. We optimized our previously established transplantation protocol for engraftment of four hCOs into early postnatal apallial mice[13] (Fig. 2a,b). This approach resulted in a graft success rate of 86.2%, as measured in 29 mice with three hiPS cell lines (Fig. 2c). This success rate is consistent with our previously reported t-hCO protocol[3], suggesting that hCO survival does not rely on an intact host cortex. MRI-based quantification of XCX graft volume demonstrated an approximately 4.7-fold growth between 2 and 3 months after transplantation (n = 14 mice, 2 hiPS cell lines; Fig. 2d). At 3 months after transplantation, the xenocortical graft constituted 91.9% of combined cortical tissue volume (n = 7 mice; Fig. 2e), averaging around 32,000 neurons per mm^3 (Extended Data Fig. 7b,c).

We next examined the integration of the hCOs into the host nervous system through anterograde tracing, rabies-based retrograde tracing and diffusion-weighted MRI (DWI). For anterograde tracing, we engrafted hCOs infected with lentiviral constructs encoding fluorophores driven by the human *SYN1* (h*SYN1*) promoter. Through transplantation of one hCO expressing eYFP and one hCO expressing oScarlet into each hemisphere (i.e. 4 hCOs transplanted per mouse), we observed organized axonal tracts coursing through a shared pathway in subcortical structures and travelling dorsally to the superior colliculus (Fig. 2f and Extended Data Fig. 7d). For retrograde tracing, we pre-infected hCOs with a *G*-deleted rabies virus encoding GFP and an AAV encoding G protein, restricting labelling to a single retrograde step from the organoid. We transplanted the pre-infected organoids into apallial mice and, 3 weeks later, we observed consistent labelling of host input neurons in the palaeocortex, thalamus and pallidum (Extended Data Fig. 3a–g). Although the hCO graft directly borders the caudate putamen (CP; Fig. 2f and Extended Data Figs. 3f,i and 7a), sparse ectopic CP-to-graft projections were only observed in one mouse. Most host inputs to the graft originated from the palaeocortex, followed by the thalamus and pallidum, with similar input patterns between mice (Extended Data Fig. 3d,e). Consistent with thalamocortical inputs from host[14], netrin-G1$^+$ projections were present within the XCX graft (Extended Data Fig. 3h).

To investigate structural connectivity in XCX mice, we acquired DWI data and applied tractography methods (Fig. 2g) to provide a quantitative description of the directionality of local water diffusion and, therefore, the organization of tissue microarchitecture. Quantifying the principal direction of water diffusion (Fig. 2h) demonstrated predominantly dorsoventral (DV) orientations, with secondary/tertiary components in the mediolateral (ML) and anteroposterior (AP) directions (Extended Data Fig. 3i–k). To quantify the variability of tissue microarchitecture between grafts, we calculated the root mean squared deviation (r.m.s.d.) of the control (0.1445) and XCX groups (0.0944), and normalized these values to the average magnitude of the vectors (control = 0.5104, XCX = 0.3663). This yielded a normalized r.m.s.d. of 0.28 (control) and 0.26 (XCX). Together, the distribution of host inputs to the graft from rabies tracing along with DV microstructural organization from DWI support some level of consistency in XCX connectivity patterns.

To characterize the cellular composition of the XCX graft, we performed snRNA-seq analysis at 5 months after differentiation. After stringent quality filtering, transcriptomic reference mapping with

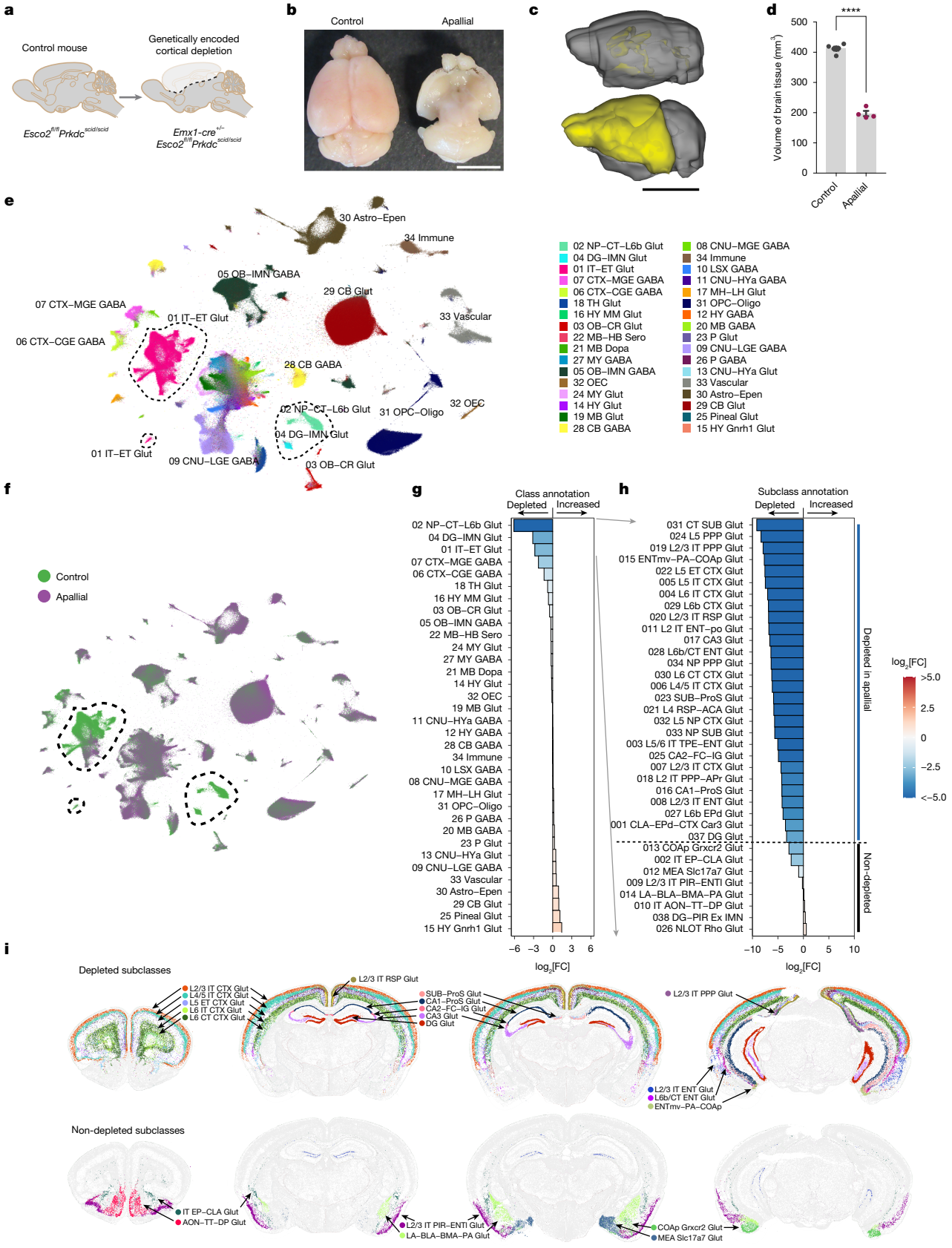


Fig. 1 | See next page for caption.

human fetal and adult cortical cell types identified clusters of major cortical cell classes, including both deep and superficial layer glutamatergic neurons, cycling progenitors, oligodendrocyte progenitor cells and astrocyte lineage cells (Fig. 2i, Extended Data Fig. 4a–f and Supplementary Table 3), consistent with our previous characterization of hCOs transplanted into the rat somatosensory cortex[3]. Notably, we

Fig. 1 | Characterization of the apallial mouse model. a, The strategy for genetic depletion of mouse cortex. **b**, Morphology of the whole brain from control (*Esco2*$^{fl/fl}$; left) and apallial (*Emx1-cre*$^{+/-}$*Esco2*$^{fl/fl}$; right) mice. All mice were homozygous for *Prkdc*^{scid}. Scale bar, 5 mm. **c**, 3D rendering of control (top) and apallial (bottom) brain segmentation from T2-weighted MRI (grey, brain tissue; yellow, CSF). Scale bar, 5 mm. **d**, Quantification of total grey and white matter volume ($n = 5$ (control) and $n = 4$ (apallial) mice; unpaired t-test, $t = 20.25$, d.f. = 7, $P = 1.8 \times 10^{-7}$). **e**, Whole-brain snRNA-seq ($n = 880{,}149$ nuclei): uniform manifold approximation and projection (UMAP) dimensional reduction visualization of high-quality nuclei annotated to cell classes using MapMyCells to the Allen Whole Brain Cell Atlas[9]. $n = 2$ (control) and $n = 2$ (apallial) mice. Dashed lines denote pallial-derived glutamatergic neuronal cell classes, excluding olfactory bulb neuronal classes. **f**, UMAP visualization of samples from control and apallial mice. **g**,**h**, The log-transformed proportion changes between control and apallial cell classes (**g**) and subclasses (**h**). The dashed line separates cell subclasses meeting an exploratory criterion for depletion (negative binomial generalized linear model with FDR < 0.1) from those not meeting this criterion. **i**, MERFISH sections from the Allen Brain Cell Atlas, highlighting selected depleted and non-depleted cell subclasses. Data are mean ± s.e.m. All statistical tests were two-tailed; ****$P < 0.0001$. CT, corticothalamic; ex., excitatory; IT, intratelencephalic; NP, near projecting. Cell class and subclass annotation acronyms are defined by the mouse Allen Whole Brain Cell Atlas (taxonomy identifier: CCN20230722; https://brain-map.org/; https://alleninstitute.github.io/abc_atlas_access/_static/WMB-taxonomy/20230830/cluster.html).

observed a small but variable population of human GABAergic neurons (Extended Data Fig. 4c). As expected for this earlier timepoint relative to our previous t-hCO dataset, 1.8 ± 0.6% of the cells were cycling, which is comparable to second-trimester primary human cortical data (1.3 ± 0.2% cycling progenitors)[15]. To address whether human-enriched progenitor populations emerge in the XCX graft, we mapped our snRNA-seq data to a recent annotated atlas of human neocortical development[15]. Within XCX progenitor cells, we identified signatures consistent with the major human cortical progenitor subtypes (Extended Data Fig. 5f,g), including outer radial glia, truncated radial glia, ventricular radial glia and the recently described tripotential intermediate progenitor cells[15]. To further benchmark the maturation stage of XCX cells, we performed transcriptomic comparisons with developing human cortical snRNA-seq datasets[16,17]. We performed principal component analysis (PCA) of pseudobulk profiles spanning primary human cortical development, demonstrating that PC1 is strongly correlated with sample age (Extended Data Fig. 4g). Using this maturation signature, we constructed an age predictor and found that XCX glutamatergic neuron samples at 24 weeks after differentiation exhibited a maturation state equivalent to approximately late second trimester, suggesting that XCX glutamatergic neurons are consistent with mid-gestation human fetal cortex development. We performed further immunostaining validation and found human astroglia (GFAP$^+$HNA$^+$) and host-derived microglia (IBA1$^+$HNA$^-$) and astrocytes (GFAP$^+$HNA$^-$) without evidence of reactive inflammatory changes (Extended Data Fig. 7e–h). Notably, staining against GAD65/67$^+$ revealed a small population of both human and host-derived GABAergic neurons within the XCX graft (Extended Data Fig. 7i–l), consistent with our snRNA-seq characterization.

To further verify the cell type composition and to comprehensively address the question of cytoarchitectural organization, we performed spatial transcriptomics (multiplexed error-robust fluorescence in situ hybridization (MERFISH); Extended Data Fig. 6) in XCX mice. We found a broad agreement between cell type proportions quantified by MERFISH and snRNA-seq ($r^2 = 0.66$, $P = 0.0008$; Extended Data Fig. 6d). As expected, based on the presence of multiple ventricular-like zones in organoids without consistent orientation, we did not observe canonical cortical layering throughout the graft (Extended Data Fig. 6e,f). However, we examined whether local cytoarchitectural organization might emerge within the XCX graft. We found regions within the graft that were predominantly occupied by either upper-layer intratelencephalic or deep-layer non-intratelencephalic neuron subclasses (Extended Data Fig. 6i). While these regions do not form a continuous laminar structure across the graft, nearest-neighbour analysis reveals that transcriptomically defined glutamatergic subclasses that normally reside in similar cortical layers are found in closer spatial proximity than expected by chance, consistent with local self-organization (Extended Data Fig. 6g,h). We next investigated whether the XCX graft exhibits any features of cortical arealization. We reanalysed our snRNA-seq dataset alongside primary fetal prefrontal cortex and primary visual cortex data[15] for markers of cortical arealization identified from primary fetal human cortical studies[18,19]. We found some evidence of increased expression of prefrontal-cortex-associated markers in XCX glutamatergic neurons, although this was not as strong as seen in primary human cortical data (Extended Data Fig. 5a–e). Furthermore, spatial transcriptomics analysis of the XCX mice did not reveal a clear gradient along the AP axis as defined by available cortical arealization genes present on our MERFISH gene panel (Extended Data Fig. 6j and Supplementary Table 4). This suggests that the mismatch between the developmental age of the graft and host may not support distinct cortical arealization at this timepoint.

Notably, the XCX graft contained a distinct cluster of glutamatergic neurons mapping to layer 5 extratelencephalic (L5-ET) projecting neurons (Fig. 2i–k). These transplanted human L5-ET neurons expressed both general subclass markers (*FEZF2* and *POU3F1*) and region-specific markers (*GABRQ* and *BMP3*) characteristic of human frontal insula L5-ET neurons that have been associated with human von Economo neurons (VENs)[20]. To examine the extent to which L5-ET neurons are present in various in vitro cortical organoid models, we analysed 213,753 dorsal telencephalic glutamatergic neurons from the Human Neural Organoid Cell Atlas (HNOCA)[21], which integrates single-cell RNA-seq (scRNA-seq) data across organoid protocols. We scored each neuron for L5-ET identity using UCell[22] with a gene signature derived from the adult human M1 cortex. UCell is a rank-based method in which the scores only rely on within-cell gene rankings, making them comparable across datasets. We found that essentially no in vitro organoids showed enrichment for the L5-ET program (3 out of 213,753 neurons had a L5-ET UCell score greater than 0.1), whereas 85% of XCX L5-ET neurons and 100% of adult M1 L5-ET neurons exceeded this threshold (Extended Data Fig. 5l). We next directly benchmarked cortical organoid transplantation procedures with respect to L5-ET neuron generation. We analysed snRNA-seq datasets from transplanted human cortical organoids, including our previous neonatal rat somatosensory cortex transplantation approach[3] and an adult mouse cortical cavity transplantation approach[23]. Applying the same transcriptomic mapping procedure to each dataset, we found that the XCX graft contained more than a threefold greater proportion of L5-ET neurons compared with previous approaches (Extended Data Fig. 5j). Beyond generating a greater proportion of L5-ET neurons, we used the XCX mouse platform to characterize the molecular identity of these cells in greater detail. Gene set enrichment analysis of XCX L5-ET neurons revealed enrichment for human-specific L5-ET genes identified in the primary motor cortex[24], as well as genes characteristic of L5-ET neurons in human frontoinsular cortex[20], which contains VENs—a specialized population implicated in social cognition and neuropsychiatric disease (Extended Data Fig. 5h,i).

VENs are a morphologically distinct neuronal population with large somas that are selectively present in large-brain mammals, but not rodents, and have been implicated in neuropsychiatric disorders[25–27]. To investigate whether the XCX graft contained VEN-like cells, we performed an unbiased analysis of soma size in grafts sparsely labelled with a lentivirus encoding GCaMP8s or eYFP driven by the h*SYN1* promoter (Fig. 2l,m and Extended Data Fig. 7p–r). We measured a median soma size of 9.251 µm, but the size distribution showed a long right

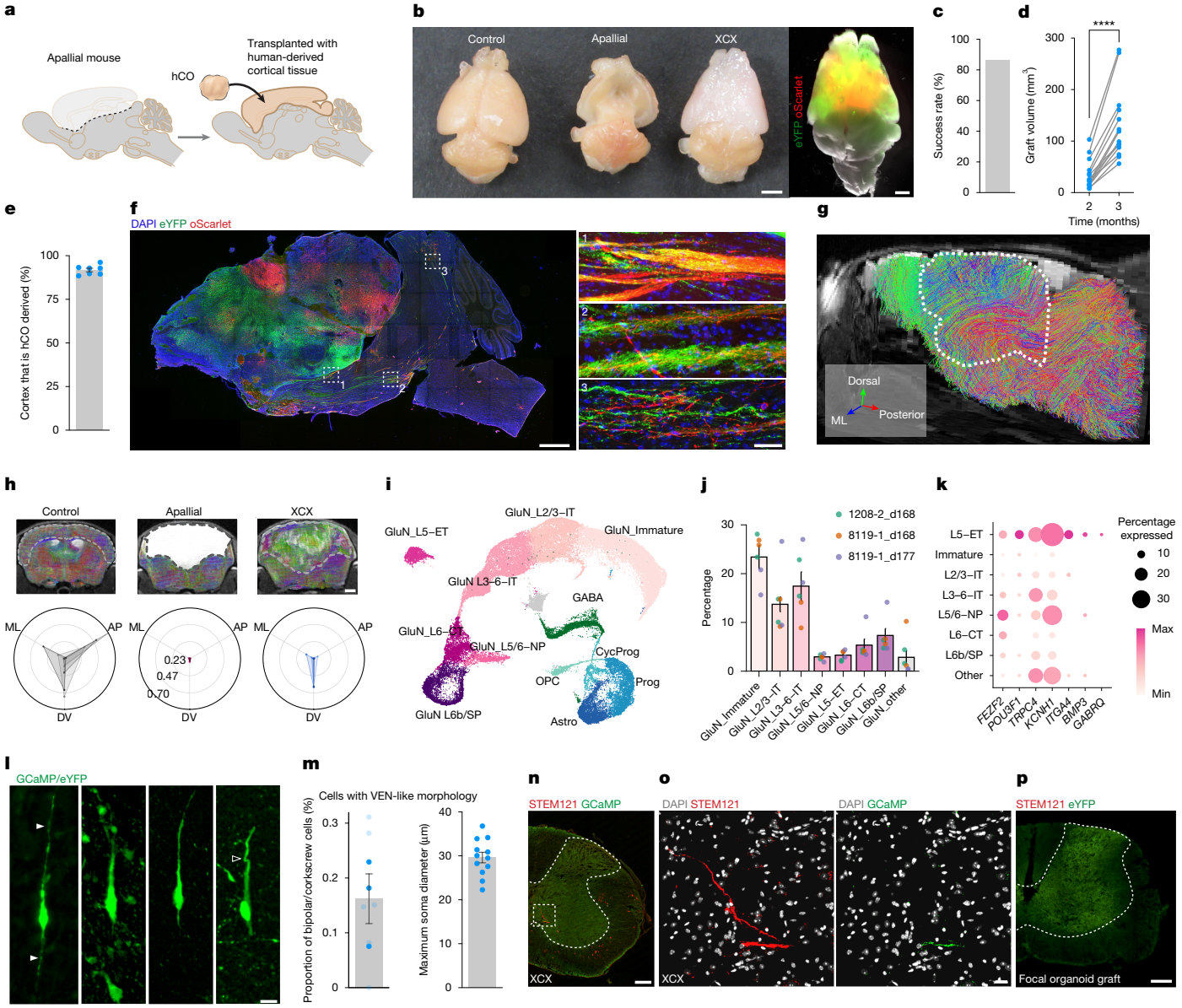


Fig. 2 | See next page for caption.

tail, prompting us to examine large cells with the distinctive bipolar or corkscrew morphology characteristic of VEN morphology[28,29]. We identified cells with VEN-like morphology in all sampled XCX mice (3 out of 3), comprising 0.16% of fluorescently labelled somas (Fig. 2l,m and Extended Data Fig. 7p–r). The median soma size of these cells is 29.8 µm (n = 3 XCX mice, 6 sections; Fig. 2m).

The presence of L5-ET neurons in XCX mice further suggested the potential for corticospinal projections, as corticospinal neurons belong to this neuronal class. To test whether the enlarged cortical space in apallial mice allowed for such long-distance human-derived projections, we examined XCX graft connectivity with the host spinal cord. Anterograde tracing from hCOs pre-infected with hSYN1-GCaMP8s consistently revealed sparse projections in the host cervical spinal cord in all mice examined (Fig. 2n,o and Extended Data Fig. 7m; n = 3 mice, 2 hiPS cell lines). Co-localization of GCaMP8s signal with immunostaining against STEM121, an antibody that detects a human-specific cytoplasmic antigen, confirmed the presence of human-derived neuronal projections in the mouse spinal cord (Fig. 2n,o). High-magnification imaging revealed the presence of GCaMP8s^{+}STEM121^{+} double-positive puncta surrounding spinal cord nuclei (Extended Data Fig. 7n). Critically, we did not find eYFP or STEM121 signal in either the spinal cord of a control *Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$ mouse with an hCO transplanted into the intact cortex (focal organoid graft; Fig. 2p) or the spinal cord of a wild-type (WT) control mouse (Extended Data Fig. 7o), both of which were concurrently stained and imaged with the XCX mouse spinal cord. Together, these data imply that the genetic depletion of the host cortex may enable the enrichment of L5-ET neurons, including corticospinal neurons.

Building on the diversity of cell types and robust integration of the hCO within the mouse central nervous system, we next examined whether the XCX graft exhibited spontaneous network activity in vivo. To test this, we imaged population-level neural calcium dynamics using widefield optoencephalography (OEG; Fig. 3a). To selectively record from graft-derived neurons, we infected hCOs with a lentivirus encoding GCaMP8s driven by the h*SYN1* promoter before transplantation (Fig. 3b). When the transplanted hCO reached 4 months after differentiation, the skulls of XCX mice were optically cleared to enable cortex-wide calcium imaging. As a control for movement and blood flow-related artefacts, excitation at 405 nm was interleaved in alternating frames. We observed large synchronous calcium events punctuated by periods

Fig. 2 | Transplantation of human cortical organoids into the apallial mouse. a, Schematic of engraftment of an apallial mouse with hCOs. **b**, Dorsal view of brains extracted from control, apallial and apallial engrafted with hCOs (left). Right, dorsal view of mouse engrafted with two organoids expressing *hSYN1-eYFP* and 2 organoids expressing *hSYN1-oScarlet*. Scale bars, 2 mm. **c**, The percentage of transplanted mice demonstrating graft survival based on MRI 2 months after transplantation. $n = 29$ mice; 1208-2, 8119-1 and 0524-1 cell lines. **d**, Increase in graft volume from 2 to 3 months after transplantation. $n = 14$ mice; 8119-1 and 0524-1 cell lines; each dot represents a single mouse and each line connects the same mouse across timepoints). Paired *t*-test, $t = 8.219$, d.f. = 13, $P = 1.66 \times 10^{-6}$. **e**, The percentage of total cortical tissue of an engrafted apallial mouse that is hCO-derived at 3 months after transplantation, defined by MRI. $n = 7$ mice, 2 hiPS cell lines. **f**, Representative sagittal view of graft-derived axonal projections 4 months after differentiation. Scale bars, 1 mm (overview) and 50 µm (insets). **g**, Whole-brain tractogram with 1 million streamlines (10,000 shown) of a representative XCX mouse, overlaid on a T2-weighted (T2W) image. The green streamlines indicate a DV preference for water diffusion; red indicates an AP preference; and blue indicates a ML preference. The dashed white lines show the location of the XCX graft. **h**, Coronal T2W MRI scans in control, apallial and XCX (2 hiPS cell lines) mice at 4–5 months (organoid age after differentiation and animal age) with fibre orientation distributions (FODs) estimated from superimposed DWI (top). FODs are coded with the same criteria as in **g**. The cortex or graft margin is indicated by a dashed line. The lack of an average diffusion direction in the apallial cavity is consistent with the absence of tissue and coherent microstructure. Scale bar, 1 mm. Bottom, radar plots depict the principal directions of water diffusion for control, apallial and XCX mice (8119-1 and 0524-1 cell lines). The dashed lines delineate isocortex, apallial cavity and graft in control, apallial and XCX groups, respectively. Each plot depicts $n = 4$ mice per condition, differentiated by the shape of the points and shade of the respective colour. The full legend is shown in Extended Data Fig. 3k. **i**, UMAP dimensionality reduction of all clustered high-quality human nuclei. $n = 3$ mice, 6 XCX samples; 1208-2 and 8119-1 cell lines. GluN, glutamatergic neuron; SP, subplate; CycProg, cycling progenitor; Prog, progenitor; Astro, astroglia. **j**, Glutamatergic neuron cluster proportions across $n = 3$ mice, 6 XCX samples and 2 hiPS cell lines. Data are mean ± s.e.m. **k**, L5-ET marker gene expression in select glutamatergic neuron clusters. **l**, Representative bipolar/corkscrew neurons. Bipolar dendrites (solid arrows) and a characteristic corkscrew (open arrow) are indicated. Scale bar, 20 µm. **m**, Left, the mean per-mouse proportion of neurons with bipolar or corkscrew morphology as the percentage of all *hSYN1-GCaMP/eYFP*-expressing neurons (the transparent circles indicate data from individual histological sections; the opaque circles indicate average values per XCX mouse; $n = 3$ mice, 6 sections; 8119-1 and 0524-1 cell lines, 18,504 cells). Right, the distribution of maximum soma diameter (long axis) of cells with VEN-like morphology ($n = 3$ mice; 8119-1 and 0524-1 cell lines). **n**, Mouse cervical spinal cord from the XCX mouse group demonstrates GCaMP8s$^+$ and STEM121$^+$ projections. Scale bar, 200 µm. **o**, High-magnification fields of view showing representative GCaMP8s$^+$ and STEM121$^+$ neuronal projections from graft into spinal cord. Scale bar, 20 µm. **p**, Mouse cervical spinal cord from *Esco2*$^{fl/fl}$ *Prkdc*$^{scid/scid}$ mouse with an organoid transplanted into cortex (focal organoid graft) does not demonstrate eYFP$^+$ and STEM121$^+$ projections. Scale bar, 200 µm. Data are mean ± s.e.m. All statistical tests were two-tailed. ****$P < 0.0001$.

of quiescence (Fig. 3c–f and Supplementary Video 1). On a shorter timescale, these calcium bursts were composed of multiple events (Fig. 3f) with distinct event initiation profiles across the graft. Bursts lasted tens of seconds and repeated every few minutes (Fig. 3g) with individual events spaced by approximately 1 s (Fig. 3h). Further characterization of the spatiotemporal profile of calcium activity revealed that events had a localized origin (Fig. 3a,c) and propagated across the entire dorsal surface of the XCX graft on the order of 100 ms (Fig. 3a,c). To probe the relationship between XCX activity and behavioural output, we calculated the motion energy from high-speed behavioural videos and plotted the calcium activity relative to motion onset. This demonstrated a close correlation between orofacial movement and XCX calcium activity (Fig. 3i,j). To test whether the cortex-wide calcium bursts reflected bona fide population-level electrophysiological events, we performed depth-resolved local-field-potential (LFP) recordings (Fig. 3k; $n = 3$ mice) using a 32-channel linear silicon probe inserted into the graft at locations previously identified by calcium imaging as origins of spontaneous activity (Fig. 3c). Post hoc histology confirmed that the DiI-labelled probe track was entirely confined to the XCX tissue (Fig. 3k). Recordings revealed discrete LFP events that organized into clusters—hereafter termed LFP bursts (Fig. 3l). The temporal pattern of these bursts closely resembled that observed with calcium imaging (Fig. 3m,n). Furthermore, bursts were detected synchronously along the full length of the shank, indicating coherent network-wide activity rather than focal discharges. Taken together, these data indicate that the human neural graft develops into an electrically active, functionally integrated network.

We next characterized the behavioural consequences of neocortical and hippocampal ablation, with or without transplantation. To begin, we compared the overall condition of control (*Esco2*$^{fl/fl}$ *Prkdc*$^{scid/scid}$), apallial and XCX mice (Extended Data Fig. 7s,t and Supplementary Video 2). Consistent with alternative cortical depletion models in which either *Pals1* or *Chd8* was deleted from *Emx1*$^+$ cells[30,31], apallial mice weigh less than control mice, whereas XCX mice demonstrate intermediate body weights between control and apallial groups (Extended Data Fig. 7t). Next, control, apallial and XCX mice performed two 60 min open-field sessions at 2 and 3 months (mouse age), while an overhead depth camera captured their full 3D pose dynamics (Fig. 4a). Analysis using MoSeq[32,33] decomposed the depth videos into 60 stereotyped, subsecond syllables that account for 99% of the variance. As previously described[34], male and female mice were modelled separately. Combining syllable usage with basic kinematics produced a compact behavioural fingerprint for each mouse and session (Fig. 4b (rows)).

To determine whether these fingerprints encode group identity and to assess the effect of transplantation, we performed a suite of complementary analyses. Linear classifiers trained on MoSeq syllables outperformed speed- or position-based models, reliably identifying control sessions but separating apallial from XCX mice only weakly (Extended Data Fig. 8a,b). Leave-one-class-out analysis reinforced this pattern: when an entire group-session pair was withheld, control recordings were assigned to the other control-session class, whereas apallial—and especially XCX mice—sessions were reassigned to non-control classes, underscoring their overlapping signatures (Extended Data Fig. 8c,d). Syllable-level *F*-tests revealed several deviations from control but few features unique to either experimental group (Extended Data Fig. 8e,f). Linear-discriminant analysis placed the three groups at roughly equidistant centroids; XCX recordings clustered nearer apallial and exhibited the greatest session-to-session drift (Fig. 4c,d and Extended Data Fig. 8g–k). Thus, pallial depletion shifts—and transplantation partially repositions but does not normalize—the behavioural state space.

To assess the stability of the MoSeq-defined behavioural repertoire across multiple timescales, we analysed divergence and dispersion of syllable usage at the level of individual mice, groups and minutes within a session. The Kullback–Leibler divergence (KLD) between sessions of the same mouse was lowest in controls and progressively higher in apallial and XCX cohorts (Fig. 4e and Extended Data Fig. 8l). Intra-mouse coefficients of variation (CV) for individual syllables followed the same ranking (Extended Data Fig. 8m,n). Minute-scale KLD and CV was likewise elevated, reflecting frequent sudden alterations in syllable usages (Fig. 4f,g and Extended Data Fig. 8o–q). Collectively, our findings show that spontaneous behavioural patterns in apallial and XCX mice are defined by a concerted reweighting of multiple behavioural syllables, accompanied by pronounced dispersion of syllable usage within and across sessions, yielding a distributed and labile repertoire rather than a discrete, syllable-specific defect.

To complement the unbiased behavioural analysis, we probed predefined behavioural domains using targeted tasks. Contrary to

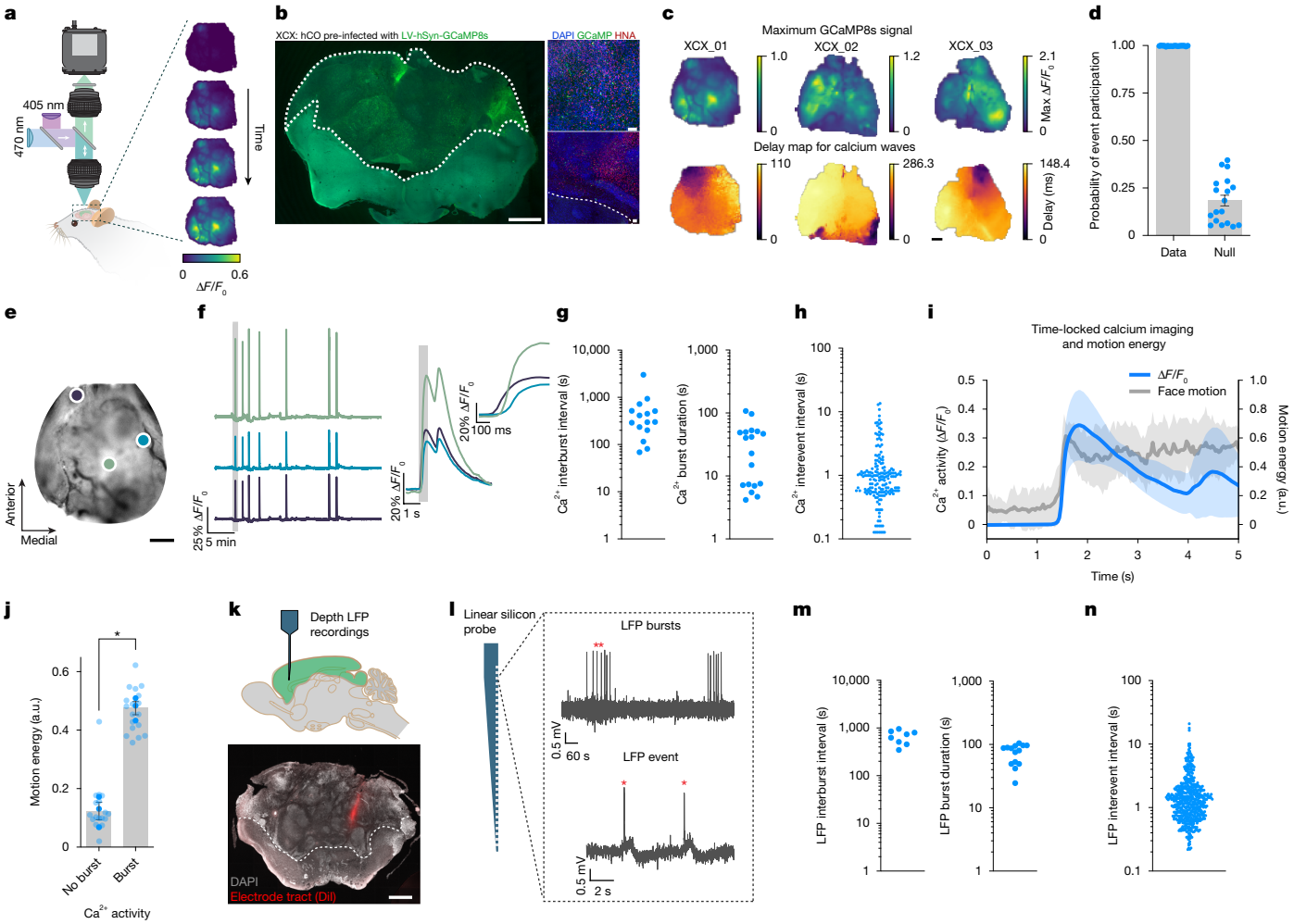


Fig. 3 | Spontaneous in vivo XCX network activity. a, Schematic of the tandem lens OEG macroscope (left). Right, representative spread of a calcium wave in the XCX graft. **b**, Representative coronal section of an XCX mouse brain used for OEG recording showing GCaMP expression confined to the graft. High-magnification insets were acquired from an adjacent coronal section co-labelled with HNA. Scale bars, 1 mm (overview) and 50 µm (insets). **c**, Maximum projections of the $\Delta F/F_0$ activity from single events in each subject (top). Bottom, maps of the time delay from the start of the event to each pixel reaching 10% of its maximum fluorescence show that the activity propagates throughout the graft at 4 months after differentiation. Scale bar, 1 mm. **d**, Pixelwise participation in calcium activity bursts compared with a null distribution with randomly permuted time indices. $n = 3$ mice, 18 bursts. **e**, Representative macroscope image showing GCaMP fluorescence. Scale bar, 1 mm. **f**, Calcium activity corresponding to the three locations in **e**; the signal was calculated using the change in fluorescence relative to the mean fluorescence ($\Delta F/F_0$) and linearly regressing out the 405 nm signal (left). Right, magnified time traces of activity corresponding to a single activity burst. **g**, Ca^{2+} interburst interval (left). Right, Ca^{2+} burst duration. $n = 3$ mice. **h**, Ca^{2+} interevent interval for events within bursts. $n = 3$ mice. Here an event is defined with the constrained non-negative deconvolution of calcium dynamics[52]. **i**, Calcium activity and facial motion energy aligned to the onset of motion and averaged across seven bursts for one mouse. The solid lines are the average of activity bursts and error bars show the s.d. **j**, Quantification of motion energy during calcium activity bursts compared with periods of low calcium activity. The transparent circles indicate data from individual bursts; the opaque circles indicate average values per XCX mouse. $n = 3$ mice, 18 bursts. Statistical analysis was performed using paired t-test; $t = 6.749$, d.f. = 2, $P = 0.021$. **k**, Schematic of depth LFP recordings (top). Bottom, representative coronal section of a XCX mouse brain showing the probe track within the graft, visualized with DiI. Scale bar, 1 mm. **l**, Representative raw LFP trace illustrating LFP bursts (top), temporal clusters of individual LFP events (bottom). Red asterisks indicate the same two selected LFP events in both magnified views. **m**, LFP interburst interval (left). Right, LFP burst duration. $n = 3$ mice. **n**, LFP interevent interval for events within bursts. $n = 3$ mice. All experiments were performed using the 8119-1 and 0524-1 cell lines. For **d** and **j**, data are represented as mean ± s.e.m. All statistical tests were two-tailed; $^*P < 0.05$. a.u., arbitrary units.

alternative cortical depletion models, we did not find gross motor changes: control, apallial and XCX mice moved at similar rates in the activity chamber and CatWalk (Fig. 4i,j, Extended Data Fig. 9a and Supplementary Video 2). Despite equivalent locomotor output, XCX mice spent significantly less time in the centre of the dark activity chamber compared with either the control or apallial cohort (Extended Data Fig. 9b), and the XCX mice spent less time rearing. A detailed spatiotemporal analysis of gait in the CatWalk (Fig. 4h and Extended Data Fig. 9d–f) revealed a shorter swing duration in apallial mice compared with the controls, but no difference between the control and XCX mice (Fig. 4k). Consistent with the general shift in behavioural repertoire found using MoSeq analysis, we also observed a shift in the overall pattern of motor behaviour during the CatWalk task. Control mice placed diagonal pairs of paws on the ground at nearly the same time (Fig. 4l,m). However, this stereotyped pattern of paw synchronization was not observed in the apallial and XCX groups. Similarly, the patterns of simultaneous paw support and step sequences differed across groups (Fig. 4l–o). Overall, the CatWalk task revealed two patterns of changes: (1) discrete changes in gait such as shorter swing duration and reduced hind-paw base of support (Extended Data Fig. 9d) in apallial

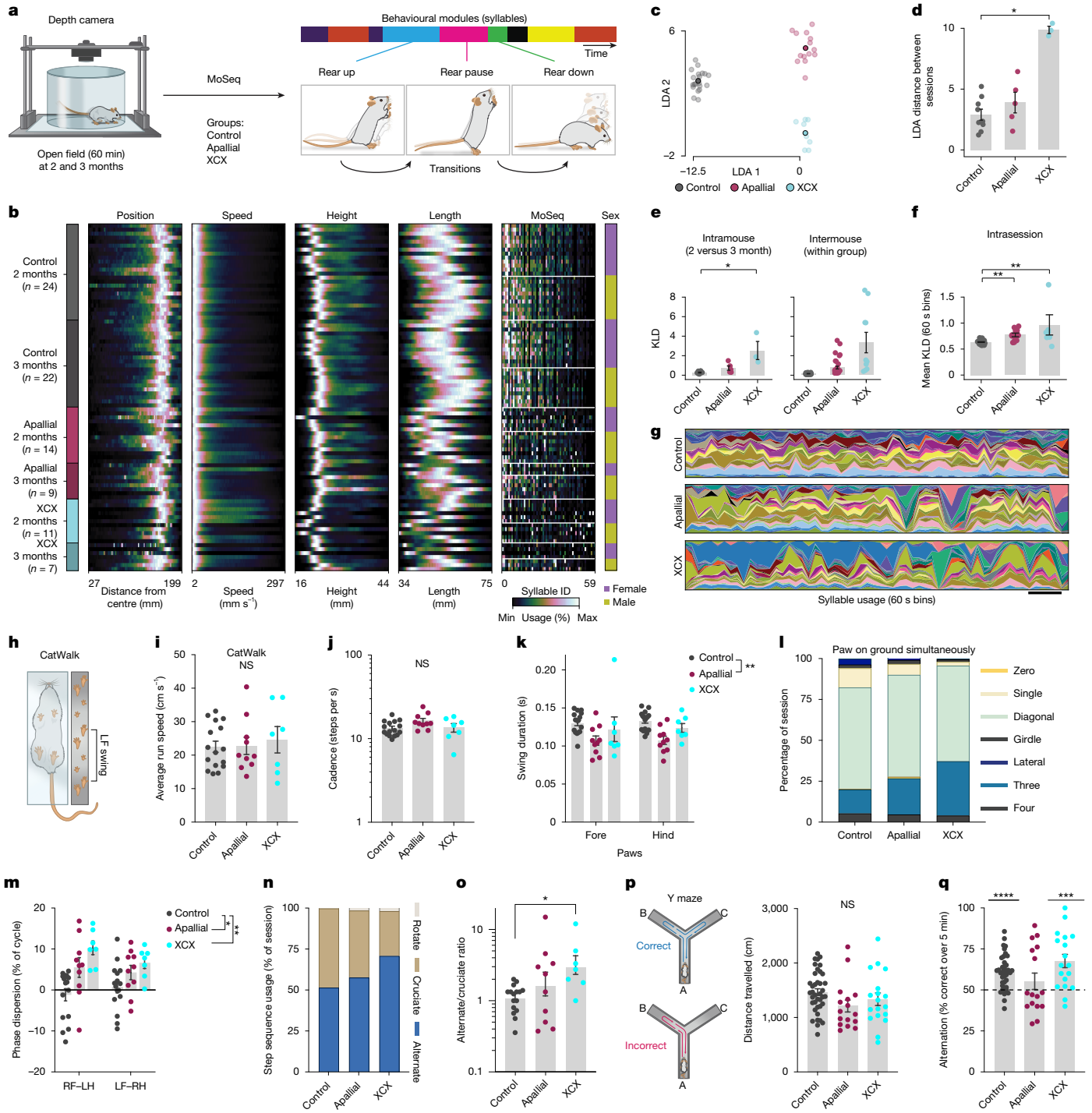


Fig. 4 | See next page for caption.

mice; and (2) higher-order changes in the pattern of paw placement in apallial and XCX mice.

The largely intact motor function made it possible to assess performance on behavioural tasks that require motor output. Given the central role of the hippocampus and neocortex in memory formation[35,36], we assayed working memory in the spontaneous alternation Y-maze task (Fig. 4p). This procedure relies on the mouse's native preference to explore novel environments, with the assumption that the mice must maintain a memory trace of the previously explored environments. Indeed, control and XCX, but not apallial, mice performed above chance levels (Fig. 4q). This effect was reproduced in two independent cohorts tested months apart (Extended Data Fig. 9g). Shifting to longer-term memory, we used an aversive Pavlovian trace conditioning procedure and explicitly defined the unconditioned stimulus (mild foot shock) and conditioned stimulus (tone) while measuring the conditioned response (freezing). Compared with the control mice, apallial and XCX mice froze less during tone presentation in the training phase as well as during tone presentation 1 day later in an altered context in the absence of foot shock. This suggests an impairment in the acquisition of aversive trace conditioning (Extended Data Fig. 9h–j), as previously reported for pre-training lesion studies of the anterior cingulate cortex[35]. Testing in the original conditioning context 1 day after training also revealed a deficit in contextual freezing, as previously reported for pre-training lesion studies of the dorsal hippocampus[37]. This deficit

Fig. 4 | Behavioural characterization of apallial and XCX mice. **a**, Unsupervised behavioural analysis of spontaneous mouse behaviour. Mouse 3D pose dynamics in an open-field assay during two 60 min sessions were captured with a depth camera and analysed using MoSeq. **b**–**g**, Control (session 1, $n = 24$; session 2, $n = 22$), apallial (session 1, $n = 14$; session 2, $n = 9$) and XCX (session 1, $n = 11$; session 2, $n = 7$) mice were recorded at 2 and 3 months old. **b**, Behavioural summary for different groups and ages. From left to right, the position (distance to centre position), speed, height and length, and MoSeq-identified syllable usage, were computed for each mouse (rows). In accordance with best practice, male and female mice were modelled separately and explained variance was set to 99% to include 60 syllables for each model. Syllable ID and colour bar (normalized syllable usage) relate to the respective male and female datasets. **c**, Linear discriminant analysis (LDA) plot indicating the similarity between the mean behavioural summaries of sessions in male mice. Points represent individual sessions; darker circles mark group centroids. **d**, Similarity of mean behavioural summaries of male mice within group-session pairs. The Euclidean distance in LDA space between the first and second session for each group is shown. A Kruskal–Wallis test showed a significant group effect ($H_2 = 7.88$, $P = 0.019$), with a significant post hoc comparison for XCX versus the control (Bonferroni-corrected $P = 0.027$). **e**, KLD of syllable-usage distributions in male mice. Left, the intramouse KLD between the first (aged 2 months) and second session (aged 3 months) of the same animal. Right, the intermouse mean KLD for all mouse–mouse pairs within a group (sessions first averaged per mouse). Dots represent single mice (control, $n = 9$; apallial, $n = 5$; XCX, $n = 3$) or mouse–pair values for visualization (control, $n = 11$; apallial, $n = 9$; XCX, $n = 5$ mice); intermouse inference was performed by permuting whole-mouse group labels. Kruskal–Wallis tests revealed significant group effects (intramouse, $H_2 = 9.93$, $P = 6.96 \times 10^{-3}$). Post hoc two-sided Mann–Whitney *U*-tests with Bonferroni correction; intramouse: XCX versus control ($P = 0.027$); apallial versus control ($P = 0.087$), and apallial versus XCX ($P = 0.21$) were not different. The intermouse permutation test did not reveal a significant group effect ($P = 0.091$). **f**, KLD between syllable usages of the entire session and 60 s windows with 30 s overlap in male mice (averaged per session). The dots represent sessions for visualization; *n* denotes mice (control, $n = 11$; apallial, $n = 9$; XCX, $n = 5$). Data are mean ± s.e.m. of the mouse-level means. Kruskal–Wallis tests revealed significant group effects intrasession ($H_2 = 14.29$, $P = 7.87 \times 10^{-4}$). Post hoc two-sided Mann–Whitney *U*-tests with Bonferroni correction; XCX versus control ($P = 0.0014$) and apallial versus control ($P = 0.0071$); apallial and XCX ($P = 1.0$) were not different. **g**, The usage of each syllable (assigned a unique colour on the *y* axis) over time (*x* axis). The syllable usages in 60 s bins are shown for three representative example recordings of male mice. Note the variations in the pattern of averaged used syllables within the 1 min windows across a session. Scale bar, 5 min. **h**, Schematic of gait analysis in the CatWalk task for control ($n = 16$), apallial ($n = 10$) and XCX ($n = 7$) mice in **i**–**o**. **i**, Average run speed; analysis of variance (ANOVA), $F_{2,30} = 0.1673$ ($P = 0.8467$). **j**, Steps per second. ANOVA after log transformation, $F_{2,30} = 2.482$ ($P = 0.1005$). **k**, Swing duration in the CatWalk. Two-way ANOVA with paw as a within-subject factor: group main effect, $F_{2,30} = 7.259$ ($P = 0.0027$); paw main effect, $F_{1,30} = 0.1501$ ($P = 0.7011$); interaction, $F_{2,30} = 0.03734$ ($P = 0.9634$); Holm–Šidák-corrected post hoc tests: control versus apallial ($P = 0.0019$); control versus XCX ($P = 0.2303$); XCX versus apallial ($P = 0.1084$) (Supplementary Table 5). **l**, The proportion of time that mice used the depicted paw support strategies. **m**, Phase dispersion of diagonal paw pairs in the CatWalk, quantifying the relative timing of the target paw with respect to an anchor paw, expressed as a percentage of the step cycle. Two-way ANOVA with paw-pair as a within-subject factor: group main effect, $F_{2,30} = 8.599$ ($P = 0.0011$); paw-pair main effect, $F_{1,30} = 0.7741$ ($P = 0.3859$); interaction, $F_{2,30} = 2.169$ ($P = 0.1319$); Holm–Šidák-corrected post hoc tests: control versus apallial ($P = 0.0290$), control versus XCX ($P = 0.0014$), XCX versus apallial ($P = 0.1458$). LH, left hind; RF, right front. **n**, The proportion of alternate, cruciate or rotate step sequences used at the group level. **o**, The ratio of alternate to cruciate step sequences. Welch's ANOVA after log transformation, $W_{2.000,11.76} = 4.960$ ($P = 0.0274$). Holm–Šidák-corrected post hoc tests: control versus apallial ($P = 0.4108$); control versus XCX ($P = 0.0406$), XCX versus apallial ($P = 0.4108$). **p**, The Y-maze task for control ($n = 39$), apallial ($n = 16$) and XCX ($n = 18$) mice (left). Right, distance travelled in the Y maze. ANOVA after log transformation: $F_{2,70} = 2.690$ ($P = 0.0749$) (Supplementary Table 5). **q**, Alternation in the Y-maze task. One sample *t*-test against chance performance (50%): control, $t = 6.319$, d.f. = 38, $P = 2.08 \times 10^{-7}$; apallial, $t = 1.070$, d.f. = 15, $P = 0.3016$; XCX, $t = 4.343$, d.f. = 17, $P = 0.0004$; the same group sizes as indicated in **p**. All experiments used the 8119-1 and 0524-1 cell lines. Data are mean ± s.e.m. NS, not significant; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

did not reflect an inability to sense foot shocks, nor an inability to learn a foot shock contingency. In an active avoidance task requiring mice to exit a chamber upon cue presentation, apallial mice outperformed controls (Extended Data Fig. 9k,l). This result aligns with previous findings whereby hippocampal-lesioned mice exhibited a change in avoidance strategy: a reduction in passive avoidance and an increase in active avoidance[35]. Finally, we assessed social preference and mechanical and thermal sensitivity. Using a three-chamber sociability assay, we observed a preference for a novel mouse over a novel object that did not differ across groups, although total investigation time was lower in apallial than control mice (Extended Data Fig. 9m). Notably, no group displayed a preference for interaction with a novel mouse over a familiar mouse (Extended Data Fig. 9n), consistent with previous reports of social deficits in *Prkdc*^{scid} mice[38]. Lastly, no group differences were detected in the von Frey (VF) test or hot plate assays of mechanical and thermal sensation (Extended Data Fig. 9o,p). The behavioural analyses reveal intact gross locomotor output with subtle differences in limb coordination, altered higher-order organization of behaviour and restricted changes in targeted behavioural assays.

We next wanted to investigate whether XCX mice could be used to model disease-related phenotypes. Specifically, we applied the xenocortication platform to model hypoxic injury in developing human cortical tissue in vivo, an injury that is associated with motor deficits[39]. We exposed control and XCX mice to 5% O_2 for 5 h in a BioSpherix chamber (including a 1 h graded descent from atmospheric oxygen to 5% O_2, as described in the Methods). To assess the hypoxic response with cellular and species specificity, we used immunohistochemistry to evaluate HIF1α together with human nuclear antigen (HNA). In 3 out of 3 XCX mice exposed to 5% O_2 for 5 h, HIF1α immunoreactivity was detected in the human graft, whereas HIF1α signal was not appreciable in the adjacent mouse palaeocortex (Fig. 5a,b) or in control mice exposed to the same hypoxia protocol (Extended Data Fig. 10c). Thus, our analysis localizes the most prominent HIF1α immunoreactivity to the human graft. Previous bulk protein immunoblotting studies provide complementary evidence that HIF1α can also be elevated in the adult mouse brain after a milder hypoxia protocol[40]. In a new cohort of mice, we performed susceptibility-weighted imaging (SWI) to obtain a tissue-level correlate of hypoxia[41,42]. This MRI method is highly sensitive to intravascular deoxyhaemoglobin and extravascular blood product deposits[43]. A comparison of SWI sequences of the same mice at baseline and 10 days after injury revealed enhanced hypointense vasculature in 3 out of 3 XCX mice exposed to 5% O_2 for 5 h (Fig. 5c). Critically, we did not observe HIF1α expression or an increase in SWI hypointensities in XCX mice exposed to control atmospheric oxygen conditions nor in control mice exposed to 5% O_2 for 5 h (Extended Data Fig. 10a–d). Finally, we found that XCX mice exposed to 5% O_2 for 5 h exhibited an increased density and complexity of intragraft microglia and astroglia (Fig. 5d–i), a shared cellular response to brain insults[44–46].

We then tested whether XCX mice could reveal functional behavioural readouts of injury. Control, apallial and XCX mice performed the CatWalk procedure 1 day before and 2 days after injury (Fig. 5j). Although all groups were exposed to the same hypoxic injury protocol, XCX mice showed the greatest increase from baseline in three- to four-paw support, adopting a more stable position than the more common two-paw position[47] (Fig. 5k). This was potentially driven by longer hindpaw contact duration, larger contact area and greater maximum print intensity on the glass surface (Fig. 5l,m and Extended Data Fig. 10h). Critically, this phenotype was not associated with changes in overall running speed, step cadence or a non-specific increase of variability in the XCX group (Fig. 5n and Extended Data Fig. 10i–k).

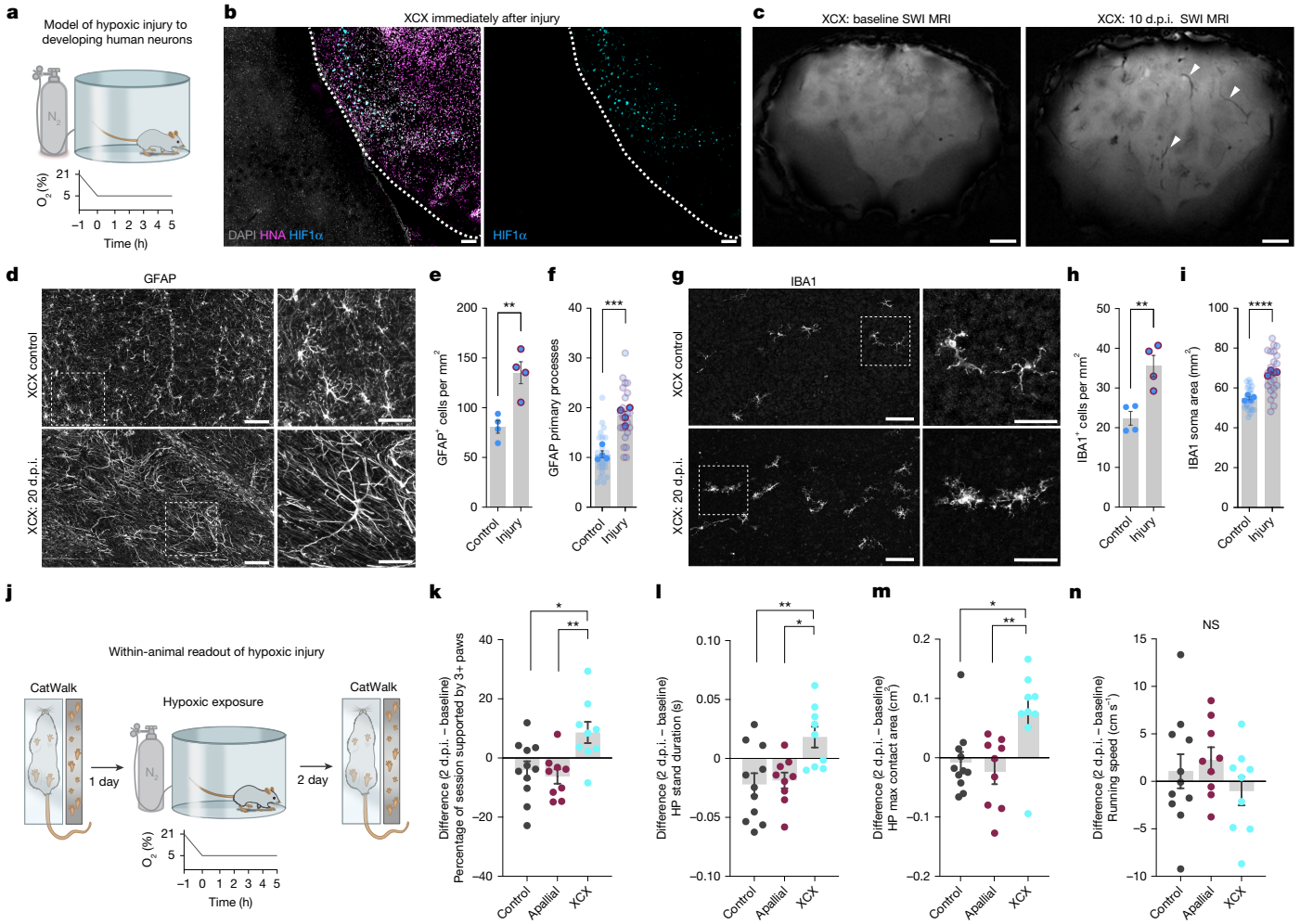


Fig. 5 | Hypoxic injury to developing human cortical tissue in XCX mice. **a**, Induction of hypoxic injury using N_2 displacement. Mice were exposed to 5% O_2 for 5 h preceded by a 1 h run-in period during which the oxygen percentage was gradually decreased to the 5% target. **b**, Representative immunostaining for HIF1α in an XCX graft collected immediately after hypoxic injury. The dashed line indicates the graft margin, delineated based on HNA signal. Scale bars, 100 µm. **c**, SWI sequence of the same XCX mouse before injury (left) and 10 days post-injury (10 d.p.i., right). The white arrows indicate selected examples of tortuous hypointensities that were new or more prominent post-hypoxic injury. The increased hypointensities on SWI are consistent with a larger ratio of deoxy- to oxyhaemoglobin. Scale bars, 1 mm. **d**, Representative immunostaining of intragraft GFAP in an XCX graft 20 days after injury (bottom left) and an age-matched normoxia XCX graft (top left) highlighting the density. Right, high-magnification insets from the region indicated with dashed squares highlighting the morphology. Scale bars, 100 µm (overview) and 50 µm (insets). **e**, Quantification of intragraft GFAP$^+$ cellular density. $n = 4$ (control XCX) and $n = 4$ (post-injury XCX) mice. Unpaired t-test, $t = 4.270$, d.f. = 6, $P = 0.0053$. **f**, Quantification of intragraft GFAP$^+$ primary processes ($n = 4$ control XCX mice, 23 cells; 4 post-hypoxia XCX mice, 24 cells; unpaired t-test, $t = 7.23$, d.f. = 6, $P = 0.0004$). The transparent circles indicate data from individual cells; the opaque circles indicate average values per XCX mouse. **g**, The same as **d**, but for intragraft IBA1. Scale bars, 100 µm (overview) and 50 µm (insets). **h**, Quantification of intragraft IBA1$^+$ cellular density. $n = 4$ (control XCX) and $n = 4$ (post-injury XCX) mice. Unpaired t-test, $t = 4.119$, d.f. = 6, $P = 0.0062$. **i**, Quantification of intragraft IBA1$^+$ soma area ($n = 4$ control XCX mice, 23 cells; 4 post-hypoxia XCX mice, 24 cells; unpaired t-test, $t = 13.99$, d.f. = 6, $P = 8.3 \times 10^{-6}$). The transparent circles indicate data from individual cells; the opaque circles indicate average values per XCX mouse. **j**, Schematic of the CatWalk procedure to quantify within-animal (that is, before versus after manipulation) behavioural readouts of hypoxic injury. **k**–**n**, CatWalk data for control ($n = 6$ female, 5 male), apallial ($n = 9$ female) and XCX ($n = 4$ female, 5 male) mice. Each point represents the within-animal difference between performance at 2 days after injury and at the baseline; that is, values above zero indicate that the measurement was increased after injury. **k**, The difference in the percentage of the session supported by 3 or 4 paws. ANOVA, $F_{2,26} = 6.634$, $P = 0.0047$. Holm–Šidák-corrected post hoc tests: control versus apallial ($P = 0.6144$), control versus XCX ($P = 0.0120$), XCX versus apallial (P = 0.0076). **l**, Difference in hindpaw (HP) stand duration per step cycle. ANOVA, $F_{2,26} = 6.543$, $P = 0.0050$. Holm–Šidák-corrected post hoc tests: control versus apallial ($P = 0.7842$), control versus XCX ($P = 0.0075$), XCX versus apallial ($P = 0.0140$). **m**, Difference in the maximum hindpaw print area. ANOVA, $F_{2,26} = 6.657$, $P = 0.0046$. Holm–Šidák-corrected post hoc tests: control versus apallial ($P = 0.5553$), control versus XCX ($P = 0.0134$), XCX versus apallial ($P = 0.0069$). **n**, Difference in the average speed. ANOVA, $F_{2,26} = 0.9919$, $P = 0.3845$. All experiments were performed using the 8119-1 and 0524-1 cell lines. Data are mean ± s.e.m. All statistical tests were two-tailed. NS, not significant; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

Discussion

Understanding human neurodevelopment and disease has been limited by the inaccessibility of human brain tissue for experimental manipulation. The transplantation of neural organoids into rodent hosts offers a promising in vivo platform for investigating human neural function and developing therapeutics. For example, using t-hCO, we identified cellular phenotypes in neurons derived from patients with a severe neurodevelopmental disorder and developed a therapeutic that rescued these defects in vivo[48]. However, current approaches that engraft human

cells into intact host tissues face substantial spatial constraints and competition with endogenous circuits, limiting their ability to model cortex-wide alterations after environmental or genetic perturbations. Here we genetically ablated the majority of the mouse cortex to allow integration of an electrophysiologically active, self-organizing graft of human-derived neural cells into the mouse brain. The XCX graft developed extensive connectivity throughout the host brain and extended projections into the spinal cord. This method also generated a defined cortical L5-ET cell-type reminiscent of VENs, a population that has been challenging to generate in vitro and is implicated in multiple neuropsychiatric diseases[25–27]. Furthermore, this platform enables cortex-wide recording of human neurons in an awake preparation and, importantly for disease modelling, the assessment of behavioural readouts associated with human-derived cortical neurons. For example, XCX mice display gait deficits following hypoxic injury, highlighting the potential for modelling perinatal hypoxia and associated cerebral palsy[49] as well as human neurodevelopmental disorders more broadly.

These applications are made possible by the finding that extensive depletion of the mouse cortex early in development preserves many behavioural functions. Although the genetic manipulation depleted half of total brain volume, apallial mice are viable, albeit with reduced body weight compared with the control mice. With the conditions and sample sizes tested, we did not detect deficits in gross locomotion, heat sensitivity or mechanical sensitivity. Alternative approaches to deplete forebrain structures in rodents have used conditional deletion of *Pals1*, a key apical complex protein[30], in *Emx1*-expressing cells. *Pals1* conditional knockout (cKO) results in premature cell cycle withdrawal and excessive production of early-born neurons, whereas conditional deletion of *Esco2* leads to cell death in dividing cells. *Pals1*-cKO mice, although underweight and hypoexploratory, retain basic reflexes and normal rotarod performance yet display irregular gait, weak wire-hang endurance and centre-bound open-field behaviour[30]. Furthermore, *Pals1*-cKO mice can learn an elaborate spatial memory assay, albeit slower than control mice[50], and *Chd8*-cKO mice demonstrate enhanced novel-object recognition memory, probably through enhanced whisker somatosensation[31]. Our behavioural experiments combined with whole-brain cell-composition profiling support more extensive cortical depletion in the *Esco2*-cKO model and demonstrate that the apallial approach enables behavioural phenotyping of XCX mice with minimal locomotor confounds.

This neural transplantation paradigm has several advantages in comparison to previous in vitro stem-cell-based or in vivo transplantation models. First, XCX mice capture cellular and circuit phenotypes induced by genetic and environmental perturbations in engrafted human neural cells, together with associated behavioural outcomes, with less interference from the host rodent cortex than in other transplantation models. Second, the presence of L5-ET neurons enables functional studies of this population, including specialized subtypes such as VENs, which are restricted to a few large-brained species thereby limiting investigation of their disease susceptibility. More broadly, xenocortication could enable insights into the biology and evolution of human cortical cells. Third, the system enables therapeutic testing across a large, integrated volume of developing human cortical tissue, together with assessment of effects on broader circuit activity. Fourth, it could provide a foundation for the preclinical development and evaluation of cell therapies for microcephaly or severe ischaemic injury during early development.

This transplantation approach has several limitations and will require further optimization. Although the present study focused on hCOs, transplanting non-cortical organoids into the apallial cavity will help to distinguish cortical lineage-specific effects from general effects of engraftment. The observed behavioural outputs probably reflect interactions between the human graft and preserved mouse subcortical circuits, but whether graft-derived neurons are necessary or sufficient for specific behaviours remains unclear. Although the graft forms anatomical connections with host tissue in the brain and spinal cord, pathway-specific recordings and manipulations will be needed to determine the extent of functional integration. The present experiments used organoids up to 177 days after differentiation, an early developmental stage marked by synchronized, high-amplitude calcium waves reminiscent of early network oscillations[51]. By comparison, sensory-evoked graft responses were detected at a later stage in a previous rat t-hCO study[3]. Although graft activity was strongly correlated with orofacial movement in XCX mice, whether this synchronized activity propagates beyond the graft remains unclear. At the timepoints examined, the XCX graft lacked several features of mature cortical circuitry, including complete areal patterning, a well-defined laminar cortical plate, full representation of GABAergic interneurons and a mature transcriptional profile. Ultimately, the mismatch between the rapid maturation of the rodent nervous system and the prolonged developmental trajectory of the human graft may limit integration.

Generating XCX mice with more mature and canonical cortical organization will require early and proactive engagement to establish ethical guidance. Relevant considerations may include studies at later stages of maturation, cortical layering or gyrification, physiological representation of excitatory and inhibitory neurons, and mature network properties, such as sparse activity with experience-dependent adaptation. These considerations may become especially important if the developmental mismatch between the graft and host is minimized, for example, by transplanting organoids into cortically depleted embryonic rodent hosts or non-human primates. Close collaboration between ethicists, researchers, patient representatives and regulatory bodies will be essential for the responsible advancement of transplantation models and support their application to investigating human biology and developing therapeutics.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-026-11032-2.

1. Pașca, S. P. et al. A nomenclature consensus for nervous system organoids and assembloids. *Nature* **609**, 907–910 (2022).
2. Pașca, S. P. et al. A framework for neural organoids, assembloids and transplantation studies. *Nature* **639**, 315–320 (2025).
3. Revah, O. et al. Maturation and circuit integration of transplanted human cortical organoids. *Nature* **610**, 319–326 (2022).
4. Whelan, G. et al. Cohesin acetyltransferase Esco2 is a cell viability factor and is required for cohesion in pericentric heterochromatin. *EMBO J.* **31**, 71 (2011).
5. Hammerschmidt, K., Whelan, G., Eichele, G. & Fischer, J. Mice lacking the cerebral cortex develop normal song: insights into the foundations of vocal learning. *Sci. Rep.* **5**, 8808 (2015).
6. Gorski, J. A. et al. Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).
7. Luo, L. et al. Optimizing nervous system-specific gene targeting with Cre driver lines: prevalence of germline recombination and influencing factors. *Neuron* **106**, 37–65 (2020).
8. Herculano-Houzel, S. The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.* **3**, 857 (2009).
9. Yao, Z. et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317–332 (2023).
10. Wong, F. K. et al. Pyramidal cell regulation of interneuron survival sculpts cortical networks. *Nature* **557**, 668–673 (2018).
11. Wu, S. J. et al. Pyramidal neurons proportionately alter cortical interneuron subtypes. *Nature* **651**, 421–428 (2026).
12. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
13. Kelley, K. W. et al. Host circuit engagement of human cortical organoids transplanted in rodents. *Nat. Protoc.* **19**, 3542–3567 (2024).
14. Lin, J. C., Ho, W.-H., Gurney, A. & Rosenthal, A. The netrin-G1 ligand NGL-1 promotes the outgrowth of thalamocortical axons. *Nat. Neurosci.* **6**, 1270–1276 (2003).
15. Wang, L. et al. Molecular and cellular dynamics of the developing human neocortex. *Nature* **647**, 169–178 (2025).
16. Velmeshev, D. et al. Single-cell analysis of prenatal and postnatal human cortical development. *Science* **382**, eadf0834 (2023).

17. Herring, C. A. et al. Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428–4447 (2022).
18. Bhaduri, A. et al. An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).
19. Qian, X. et al. Spatial transcriptomics reveals human cortical layer and area specification. *Nature* **644**, 153–163 (2025).
20. Hodge, R. D. et al. Transcriptomic evidence that von Economo neurons are regionally specialized extratelencephalic-projecting excitatory neurons. *Nat. Commun.* **11**, 1172 (2020).
21. He, Z. et al. An integrated transcriptomic cell atlas of human neural organoids. *Nature* **635**, 690–698 (2024).
22. Andreatta, M. & Carmona, S. J. UCell and pyUCell: single-cell gene signature scoring for R and Python. *Bioinformatics* **42**, btag055 (2026).
23. Wang, M. et al. Morphological diversification and functional maturation of human astrocytes in glia-enriched cortical organoid transplanted in mouse brain. *Nat. Biotechnol.* **43**, 52–62 (2025).
24. Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
25. Brüne, M. et al. Von Economo neuron density in the anterior cingulate cortex is reduced in early onset schizophrenia. *Acta Neuropathol.* **119**, 771–778 (2010).
26. Santos, M. et al. Von Economo neurons in autism: a stereologic study of the frontoinsular cortex in children. *Brain Res.* **1380**, 206–217 (2011).
27. Gefen, T. et al. Von Economo neurons of the anterior cingulate across the lifespan and in Alzheimer's disease. *Cortex* **99**, 69–77 (2018).
28. Banovac, I. et al. Somato-dendritic morphology and axon origin site specify von Economo neurons as a subclass of modified pyramidal neurons in the human anterior cingulate cortex. *J. Anat.* **235**, 651–669 (2019).
29. von Economo, C. F. & Koskinas, G. N. *Die Cytoarchitektonik Der Hirnrinde Des Erwachsenen Menschen* (Springer, 1925).
30. Kim, S. et al. The apical complex couples cell fate and cell survival to cerebral cortical development. *Neuron* **66**, 69–84 (2010).
31. Kweon, H. et al. Excitatory neuronal CHD8 in the regulation of neocortical development and sensory-motor behaviors. *Cell Rep.* **34**, 108780 (2021).
32. Wiltschko, A. B. et al. Mapping sub-second structure in mouse behavior. *Neuron* **88**, 1121–1135 (2015).
33. Gschwind, T. et al. Hidden behavioral fingerprints in epilepsy. *Neuron* **111**, 1440–1452 (2023).
34. Levy, D. R. et al. Mouse spontaneous behavior reflects individual variation rather than estrous state. *Curr. Biol.* **33**, 1358–1364 (2023).
35. Han, C. et al. Trace but not delay fear conditioning requires attention and the anterior cingulate cortex. *Proc. Natl Acad. Sci. USA* **100**, 13087–13092 (2003).
36. Squire, L. R. Memory and brain systems: 1969–2009. *J. Neurosci.* **29**, 12711–12716 (2009).
37. Burman, M. A., Simmons, C. A., Hughes, M. & Lei, L. Developing and validating trace fear conditioning protocols in C57BL/6 mice. *J. Neurosci. Methods* **222**, 111–117 (2014).
38. Filiano, A. J. et al. Unexpected role of interferon-γ in regulating neuronal connectivity and social behaviour. *Nature* **535**, 425–429 (2016).
39. Perez, A. et al. Long-term neurodevelopmental outcome with hypoxic-ischemic encephalopathy. *J. Pediatr.* **163**, 454–459 (2013).
40. Bernaudin, M. et al. Normobaric hypoxia induces tolerance to focal permanent cerebral ischemia in association with an increased expression of hypoxia-inducible factor-1 and its target genes, erythropoietin and VEGF, in the adult mouse brain. *J Cereb. Blood Flow Metab.* **22**, 393–403 (2002).
41. Messina, S. A. et al. Early predictive value of susceptibility weighted imaging (SWI) in pediatric hypoxic-ischemic injury. *J. Neuroimaging* **24**, 528–530 (2014).
42. Kitamura, G. et al. Hypoxic-ischemic injury: utility of susceptibility-weighted imaging. *Pediatr. Neurol.* **45**, 220–224 (2011).
43. Tong, K. A. et al. Susceptibility-weighted MR imaging: a review of clinical applications in children. *Am. J. Neuroradiol.* **29**, 9–17 (2008).
44. Hong, J. et al. Lipopolysaccharide administration for a mouse model of cerebellar ataxia with neuroinflammation. *Sci. Rep.* **10**, 13337 (2020).
45. Nespoli, E. et al. Glial cells react to closed head injury in a distinct and spatiotemporally orchestrated manner. *Sci. Rep.* **14**, 2441 (2024).
46. Chung, J. E. et al. Experience and safety of intraoperative Neuropixels: a case series of 56 patients. *J. Neurosurg.* **144**, 63–73 (2025).
47. Perez, H. et al. A novel, ataxic mouse model of ataxia telangiectasia caused by a clinically relevant nonsense mutation. *eLife* **10**, e64695 (2021).
48. Chen, X. et al. Antisense oligonucleotide therapeutic approach for Timothy syndrome. *Nature* **628**, 818–825 (2024).
49. Zhang, S., Li, B., Zhang, X., Zhu, C. & Wang, X. Birth asphyxia is associated with increased risk of cerebral palsy: a meta-analysis. *Front. Neurol.* **11**, 704 (2020).
50. Zheng, J., Guimaraes, R., Hu, J., Perona, P. & Meister, M. Mice in the Manhattan maze: rapid learning, flexible routing and generalization, with and without cortex. *ccneuro.org* https://2024.ccneuro.org/pdf/320_Paper_authored_CCN_2024_wtitle.pdf (2024).
51. Wu, M. W., Kourdougli, N. & Portera-Cailliau, C. Network state transitions during cortical development. *Nat. Rev. Neurosci.* **25**, 535–552 (2024).
52. Pnevmatikakis, E. A. et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

Methods

Genetically encoded cortical depletion

Emx1-cre (JAX, *B6.129S2-Emx1tm1(cre)Krj/J*), *Esco2*$^{fl/fl}$ (JAX, *B6N.129S-Esco2tm1.1Ge/J*) and C57BL SCID (JAX, *B6.Cg-Prkdcscid/SzJ*) mice were used to generate the breeding scheme: male *Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$ × female *Emx1-cre*$^{+/-}$*Esco2*$^{fl/+}$*Prkdc*$^{scid/scid}$. Breeder and offspring cages were supplemented with breeder chow (Envigo Teklad 2919) and DietGel 76A (with plant protein) to help to increase pup survival. Some breeders were more prone to parental infanticide. In these cases, we implemented an aunting strategy with Swiss Webster active dams from Charles River. To reduce competition for maternal care and milk, a subset of pups was euthanized during the second postnatal week.

hiPS cells, generation of hCOs and viral infection

We generated and cultured hCOs from hiPS cells as previously described[53,54]. In brief, hiPS cells were treated with Accutase (Innovate Cell Technologies, AT-104) at 37 °C for 7 min to dissociate them into single cells. The resulting single-cell suspension was collected in a 50-ml Falcon tube, and a cell pellet was obtained by centrifugation at 200*g* for 4 min. After resuspending the cell pellets, cell counting was performed. Approximately 3×10^6 cells in Essential 8 medium (Life Technologies, A1517001) supplemented with the ROCK inhibitor Y-27632 (10 μM; Selleckchem, S1049) were added to each well of the AggreWell 800 plate (StemCell Technologies, 34815). The plates were then centrifuged at 100*g* for 3 min to capture the cells in the microwells and incubated at 37 °C with 5% CO_2 (day −1). Then, 24 h after cell aggregation (day 0), spheroids were collected from each microwell by gently pipetting the medium up and down using a cut P1000 tip and transferring them into ultra-low attachment plastic dishes (Corning, 3262) in Essential 6 medium (Life Technologies, A1516401) supplemented with dorsomorphin (2.5 μM; Sigma-Aldrich, P5499) and SB-431542 (10 μM; Tocris, 1614). From days 2 to 5, the Essential 6 medium was changed daily and supplemented with dorsomorphin and SB-431542. On the sixth day in suspension, the neural spheroids were transferred to neural medium composed of Neurobasal A (Life Technologies, 10888), B-27 supplement without vitamin A (Life Technologies, 12587), GlutaMax (1:100, Life Technologies, 35050) and 10 U ml^{-1} penicillin–streptomycin (Gibco, 15140122). From days 6 to 24, the neural medium was supplemented with 20 ng ml^{-1} epidermal growth factor (EGF; R&D Systems, 236-EG) and 20 ng ml^{-1} basic fibroblast growth factor (FGF; R&D Systems, 233-FB), with medium changes occurring daily from days 6 to 15 and every other day until day 24. From days 25 to 42, the neural medium included 20 ng ml^{-1} brain-derived neurotrophic factor (BDNF; Peprotech, 450-02) and 20 ng ml^{-1} NT-3 (Peprotech, 450-03), with medium changes every other day. Starting from day 43, the hCOs were cultured in neural medium without growth factors, with medium changes every 4 days.

For viral labelling, all organoids were infected at least 1 week before transplant, most commonly between day 35 and day 40 of differentiation. Organoids were transferred to a 1.5 ml microcentrifuge Eppendorf tube containing 100 μl neural medium with lentivirus and incubated overnight. The next day, neural organoids were transferred into fresh neural medium in 24-well ultra-low-attachment plates with fresh medium changed daily to wash out any ambient virus particles. All lentiviruses were generated by VectorBuilder with previously used vector maps and published sequences: pLV-hSYN1-GCaMP8s, pLV-hSYN1-oScarlet and pLV-hSYN1-eYFP[55,56]. For rabies tracing experiments, day 47 hCOs were co-infected with AAV-8-CAG-FLExOFF-RabiesG (AAV-G) from the Stanford viral vector core and G-Deleted Rabies-eGFP (RVΔG-GFP) from Salk Institute. Each organoid was incubated overnight with 0.5 μl of 1:10 diluted RVΔG-eGFP virus and 0.5 μl of AAV-G into 200 μl of the neural medium described above. The next day, 800 μl of the neural media was added to each organoid and, the day after that, the organoids were washed with fresh neural medium three times and transferred into low attachment dishes (6 or 24 well). All cell lines tested negative for mycoplasma contamination, and genomic integrity was assessed using the single-nucleotide polymorphism microarray GSAMD-24v2--0.

Organoid transplantation

All procedures were performed in accordance with NIH guidelines and with previous approval of the Stanford University Administrative Panel on Laboratory Animal Care (APLAC) and in accordance with a previously published protocol[13], but optimized for the apallial mouse neuroanatomy. In brief, organoids at 30–60 days in vitro were transplanted into 5–17-day old apallial pups (median, 10; interquartile range, 9–13). Importantly, this is before the critical period of activity-dependent establishment of neural connectivity in the mouse[57]. Mice were anaesthetized with isoflurane (5% induction, 2–3% maintenance) and received Ethiqa-XR (0.65 mg per kg) or 10 mg per kg carprofen + 0.25–0.5% bupivacaine local infiltration below the incision site before placement into a stereotaxic frame with a neonatal insert (RWD) and heating pad. Switching from Ethiqa-XR to bupivacaine + carprofen helped to reduce the rate of infanticide. While maintaining an intact dura, a craniotomy was performed on each hemisphere at AP, +1 mm; ML ± 1 mm. For each hemisphere, two hCOs were loaded into the Hamilton syringe and the contents of the syringe were deposited at a rate of 10 μl min^{-1} at a depth of 1.5 mm below the surface of the dura. This yielded four transplanted hCOs per mouse. As the surgeries occurred at such a young age, there was not enough time to reliably obtain genotyping results before the surgery. Therefore, the presence or absence of the neocortex was confirmed through visual inspection after an initial incision.

MRI monitoring of graft development

MRI scans were performed on the Bruker BioSpec 70/40 small-animal scanner (7.0 Tesla, 300 MHz; Bruker) equipped with a BGA-12S gradient insert (660 mT m^{-1}, 4,570 T m^{-1} s^{-1}) and interfaced to ParaVision 360 V 3.6. A 4-element receive-only mouse CryoProbe and 86 mm volume coil were used for MR imaging. Animals were anaesthetized with 3% isoflurane mixed in O_2 and maintained with 1–1.5% isoflurane during the imaging procedure. Core body temperature was maintained at 37 °C using a warm water circulator pad and a temperature controller (Thermo Fisher Scientific). Mouse respiratory rates were monitored through a pneumatic pad placed underneath the animal (SA instruments).

Two-dimensional T2W Turbo rapid acquisition with relaxation enhancement (RARE) sequence was used to obtain transverse structural MRI for volume measurements (TR/TE = 4,500/41 ms, FOV = 18 × 18 mm^2, matrix = 258 × 258 yielding 70 × 70 μm^2 in-plane resolution, slice thickness = 0.5 mm, RARE Factor = 8, NEX = 2, scan time = 4 min 3 s). For some animals, 3D RAREvfl sequence was used (TR/TE = 1,800/90.26 ms, FOV = 15 × 15 × 8 mm^3, matrix = 150 × 150 × 80 yielding 0.100 mm isotropic resolution, RARE Factor = 43, NEX = 1.6, scan time = 10 min 33 s with compressed sensing) and data are reconstructed using the Smart Noise Reduction option.

Two-dimensional SWI data were acquired (TR/TE = 858.8/18 ms, FOV = 15 × 15 mm^2, matrix = 280 × 280 yielding 54 × 54 μm^2 in-plane resolution, slice thickness = 0.5 mm, NEX = 6, scan time 21 min 54 s). The volume of brain tissue was determined by segmenting the whole brain and subtracting the volume of cerebrospinal fluid. All analyses were performed with Slicer3D.

For high-resolution anatomical T2W images for DWI co-registration, two RARE sequences were acquired per animal: (1) a low-resolution scan (200 × 200 × 500 μm^3) for co-registration reference; and (2) a high-resolution scan (100 × 100 × 300 μm^3; RARE factor = 4, TR/TE = 4,000/30 ms, 2 averages, flip angle = 180°) used as the primary anatomical reference for visualization and registration. DWI data were acquired with a four-segment diffusion-weighted EPI sequence (Bruker:DtiEpi) using the following parameters: b = 1,000 s/mm^2, 40 non-collinear

diffusion-encoding directions plus 5 b = 0 images (45 volumes total), diffusion timing δ/Δ = 4.5/10.3 ms, TR/TE = 3,500/21 ms, in-plane resolution 150 × 150 μm^2, slice thickness 500 μm, FOV = 16 × 15 mm^2, matrix = 107 × 100, bandwidth = 250 kHz, second-order B_0 map shimming, three FOV saturation bands. The B_0 map was obtained before the sequence and first- and second-order shims were applied using mapshim. NiFti formatted data were converted from Bruker ParaVision format using dcm2niix[58].

Diffusion-weighted image processing

All diffusion and anatomical data were preprocessed using MRtrix3 (v.3.0.3)[59] and ANTs (v.2.3.5)[60,61]. The processing pipeline was applied identically to all 12 mouse brains. The other analyses were implemented in Python (version ≥3.10) using NumPy[62], SciPy[63], scikit-learn[64], nibabel[65] and matplotlib[66].

Orientation standardization. Raw images were reoriented to match the Allen Common Coordinate Framework (Allen CCF) through a two-step transformation (we called this step flip): (1) axis permutation (axes 2, 1, 0) to map scanner axes to AP, DV and ML axes; and (2) a flip along axis 1 (Y) to align the DV direction with the atlas convention. The same transformation was applied to the 4D DWI data (axes 2, 1, 0, 3) and the high-resolution T2W image.

Preprocessing. We preprocessed all mice with the same protocol using the following steps. (1) Gradient verification: the diffusion gradient table was checked and corrected for orientation consistency using MRtrix3's dwigradcheck. (2) Denoising: thermal noise was removed using MP-PCA denoising[67] with a 5 × 5 × 5 voxel kernel and the Exp2 estimator. (3) Brain mask estimation: an initial brain mask was created from the denoised DWI data using dwi2mask, with all automatically generated masks then requiring manual editing. A refined brain mask was manually created from the registered DWI data for use in all downstream analyses. (4) Bias field correction: B_1 field inhomogeneity was corrected using the N4 algorithm from ANTs (B-spline fitting distance = 200, convergence [10, 1e-6], sigma = 4). (5) DWI-to-T2W registration: the bias-corrected mean b = 0 image was rigidly registered to the high-resolution T2W anatomical using ANTs. The resulting affine transform was converted to MRtrix format and applied to the full 4D DWI volume, bringing all diffusion data into T2W anatomical space.

ROI definition. For control mice, the isocortex region of interest (ROI) was derived from the Allen CCF atlas (Allen Institute for Brain Science). The Allen average template was nonlinearly registered to each mouse's T2W image using ANTs SyN, and the warp was applied to the Allen annotation volume using nearest-neighbour interpolation. ROIs were extracted by collecting all descendant structure IDs under the following parent structures: isocortex (ID 315). Each ROI was split into left and right hemispheres along the midline of axis 2.

For apallial and XCX mouse groups, the apallial cavity or graft, respectively, was manually segmented in 3D Slicer on the high-resolution T2W images. Manual ROIs were transformed with the same flip as the rest of the MRI images and resampled to FOD space using nearest-neighbour interpolation.

FOD estimation. For cross-subject comparability, response functions were averaged across all 12 subjects to create group-level response functions. FODs were then estimated per mouse using multi-shell multi-tissue constrained spherical deconvolution (MSMT-CSD)[68] with the group-averaged WM and CSF response functions. Intensity normalization was performed using mtnormalise[69] with WM and CSF compartments only (three-tissue normalization including GM was unstable for mouse brain data). The resulting group-normalized WM FOD images were used for all subsequent analyses. As a negative control, we confirmed the lack of directional FODs in the apallial cavity and excluded the apallial group from statistical comparisons between groups.

Fixel-based analysis. FODs were decomposed into discrete fixels using fod2fixel, yielding per-fixel apparent fibre density and peak amplitude maps. The voxel-wise total apparent fibre density was computed by summing all fixel contributions through fixel2voxel. Directional fibre density within each ROI was decomposed into AP, DV and ML components for ROI-level microstructural characterization.

Slicewise fixel direction analysis. To characterize the spatial organization of fibre orientations within the seed ROI, a slicewise fixel analysis was performed. For each coronal slice through the seed ROI (isocortex for control, graft for XCX), fixel peak direction vectors were extracted from the group-normalized FOD using fod2fixel followed by fixel2peaks. Fixel directions were clustered into two populations using k-means clustering (k = 2) applied to the absolute values of the direction vectors (to handle antipodal symmetry). For each cluster, the mean direction was computed using the scatter-matrix eigenvector method, which correctly averages axial data by finding the principal eigenvector of the covariance matrix, $S = \sum_i v_i v_i^T$. Cluster directions were scaled by the mean FOD peak amplitude to produce amplitude-weighted direction vectors.

Voxel-wise dominant diffusion orientation. For each voxel, the primary direction of restricted water diffusion (defined as the fixel corresponding to the largest FOD lobe) was identified. The spatial orientation of this dominant diffusion component was categorized by the axis with the largest absolute component of its direction vector (AP, DV or ML). The percentage of voxels dominated by each spatial direction was calculated per subject and averaged per group.

Direction vector r.m.s.d. analysis. Within-group consistency of fibre orientations was quantified by the r.m.s.d. of direction vectors across mice, computed per slice and overall:

$$\text{r.m.s.d.} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{v}_i - \bar{\mathbf{v}}\|^2},$$

where $\mathbf{v}_i$ is the k-means cluster 1 direction vector (scaled by FOD peak amplitude) for subject i, and $\bar{\mathbf{v}}$ is the group mean vector. The mean vector norm $\|\bar{\mathbf{v}}\|$ and the ratio $\frac{\text{r.m.s.d.}}{\|\bar{\mathbf{v}}\|}$ were also computed to contextualize variability relative to signal magnitude.

Radar plots. Fixel direction vectors were visualized as radar (polar) plots with three axes corresponding to the AP, DV and ML anatomical directions. Each radar plot displays one coronal slice through the seed ROI. A thin radial grid with three concentric rings provides the amplitude scale. The amplitude axis was unified across all groups and slices for direct comparison.

Whole-brain tractography. Whole-brain tractograms were generated using the iFOD2 probabilistic algorithm[70] seeded from the brain mask. Tracking parameters were as follows: step size = 0.02 mm, maximum angle between successive steps = 60°, FOD amplitude cutoff = 0.1, minimum streamline length = 2 mm, maximum streamline length = 15 mm. 1 million streamlines were generated per subject.

Rabies-virus-based monosynaptic retrograde tracing from organoid graft

hCOs were pre-infected with Rb-GFP and AAV-G as described above, washed multiple times with fresh medium followed by daily medium change for 3 days and transplanted as described above. Then, at

3 weeks after transplantation, the mice were transcardially perfused as described above. Deletion of the rabies-virus *G* gene restricts retrograde transport to cells carrying the G protein (that is, one retrograde step from the organoid). The entire forebrain was sectioned (100 µm) into four series (300 µm space between sections) and individual series were mounted. The series of sections used for analysis were also immunostained with HNA and only $HNA^{-}GFP^{+}$ cells were included for analysis. Sections consecutive to the analysed series (that is, the next series) were immunostained for TH and PV to reveal the CP and reticular thalamus (respectively) to aid identification of regions. Then, using the Allen Institute Brain Atlas, bright-field images of the sections revealing white matter tracts, previous MRIs of apallial mice and the TH and PV immunostaining, we localized each section for analysis to the relevant AP slice in the brain atlas. Then, using the BigWarp plugin for FIJI/ImageJ[71], we warped the atlas to relevant anatomical landmarks. Finally, we manually counted $HNA^{-}GFP^{+}$ cells using the CellCounter plugin within the defined brain regions. Owing to the altered anatomy in the apallial mouse, here we combined subregions into Allen Institute's classification of parent structures. Cells in regions that could not be reliably delineated, as well as a small number of midbrain inputs, were included in the total input count but not shown as separate regional categories. Accordingly, the regions presented accounted for 98.1–100% of labelled inputs per animal, and percentages do not necessarily sum to 100%.

MoSeq analysis

Behavioural data acquisition was performed similar to previous descriptions. Mice were tested in a circular open-field arena (18 inch diameter, 15 inch height; US Plastics) under dim red illumination. The opaque enclosure was spray-painted black (Acryli-Quik Ultra Flat Black) to avoid image artefacts. Each animal was placed into the centre of the arena and allowed to explore freely for 60 min. Between trials, the enclosure was wiped sequentially with 10% bleach, 1% Alconox and 70% ethanol to remove odour cues. All sessions were run during the dark phase, when mice are naturally active. 3D tracking was obtained using the Microsoft Kinect v2 depth sensor mounted approximately 65 cm above the arena to give a top-down view. Depth frames were streamed at 30 Hz over USB 3.0 to a dedicated acquisition computer (Intel i5-6400 CPU, 16 GB RAM, NVIDIA GT1030 GPU) and written to disk in raw binary format via a custom C#/C++ interface.

Analysis of MoSeq

Preprocessing and modelling with MoSeq. All depth videos were processed with the open-source MoSeq pipeline (Python) running either on local workstations or on the Stanford cluster computer (Sherlock 2). Each raw frame was first resampled and background-subtracted using a median depth image built from the initial 1,000 frames. Negative values and pixels higher than the defined ceiling were zeroed, yielding a clean relief map of which the intensities represent the height above the arena floor. The mouse contour was located, its centre of mass and yaw angle were computed, and an 80 × 80-pixel patch centred and aligned to the nose–tail axis was extracted for every frame. The resulting 30 Hz image stack constitutes the aligned behavioural video.

For modelling, every stack was flattened and reduced by PCA. The principal component time series from all animals of a sex were then whitened to remove cross-component covariance. Separate autoregressive hierarchical Dirichlet-process HMMs (AR-HMMs) were fitted to the male and female datasets. Each model was truncated after the first 60 latent states (syllables), which together accounted for at least 99% of the total variance, and these state sequences formed the basis of all downstream analyses.

Behavioural summaries. Session-level behavioural fingerprints were assembled by concatenating four kinematic distributions with a truncated syllable-usage vector. As previously described, the distance of the animal's centre to the arena centroid and its planar speed were each binned into 100 equiprobable intervals, whereas body height and body length were summarized with 45-bin histograms. The syllable component comprised the relative occupancy of the first 60-syllable labels produced by sex-specific AR-HMMs. Histograms were normalized such that every column represented a probability distribution. Scalar channels were additionally scaled to a common 0–1 range using the global minimum and maximum estimated across sessions of the same sex. For downstream analyses, the four scalar histograms and the syllable vector were stacked to yield a 350-element feature vector per recording session ($2 \times 100 + 2 \times 45 + 60$). The resulting matrix (sessions × features) constitutes the behavioural summary used for all classification and statistical tests.

Classification and confusion matrices. Linear classifiers were trained independently on each feature modality—distance to centre, planar velocity, body height, body length and the 60-element MoSeq vector—using the session-level fingerprints described above. For a given modality, the relevant column(s) were extracted from the fingerprint matrix, concatenated into X, and paired with the session label y (CON_SES1 and so on). A one-versus-rest logistic-regression model (liblinear optimizer) served as the base estimator. The penalization constant C was optimized once, before cross-validation, by a grid search over 50 logarithmically spaced values between 10^{-6} and 10^{3}.

Performance was assessed with 500 repeated stratified shuffle–split resamples that preserved class proportions (80% training, 20% test per fold). Each split was processed in parallel across CPU cores; within every split, the model was fit to the training set and evaluated on the withheld samples. To obtain a chance-level baseline, the entire procedure was repeated 100 times with the class labels randomly permuted before fitting. The resulting collections comprise, for every modality, the vectors of true and predicted labels, the per-split coefficient vectors and overall accuracies for both the real and the shuffled runs; they were serialized for reproducibility. The displayed matrices and accuracy labels summarize the original 500 session-level splits, with C selected on the full dataset and without grouping by mouse. An additional sensitivity analysis used 500 80:20 splits of mice, stratified by experimental group, with all sessions from each mouse kept together. Within each training set, C was selected by threefold mouse-grouped cross-validation over the same grid, keeping the cohort-derived features fixed. Pooled held-out accuracies for MoSeq, speed and position were 0.590, 0.351 and 0.311 in males and 0.631, 0.490 and 0.440 in females, respectively.

Aggregated predictions were summarized as row-normalized confusion matrices (i,j) whose entry gives the fraction of samples from true class i assigned to class j. For each modality, the real and shuffled matrices were displayed side by side, rendered on a 0–1 greyscale.

Held-out confusion analysis. To assess pairwise similarity between behavioural conditions without relying on any within-class information, we performed a leave-one-class-out evaluation on the syllable-usage fingerprints. For each sex, the 60-dimensional MoSeq vector of every session was retained, row-normalized to unit sum and associated with its class label (CON_SES1 and so on). The procedure iterated once over every label: at each iteration all sessions belonging to the target class were withheld as the test set, a linear logistic-regression model with L2 regularization (inverse-variance class weighting, regularization constant = 100, liblinear optimizer) was fitted to the remaining data and immediately applied to the unseen samples. Concatenating the predictions from all iterations yielded a square confusion matrix (i,j) of which the entry recorded the proportion of samples of true class i classified as j after row normalization. The matrix was visualized with a perceptually uniform greyscale and the standard group/session colour palette, providing an intuitive map of interclass similarity that is independent of cross-validation performance metrics.

***F*-statistics for syllable-level discrimination.** As previously described, we identified the syllables that contributed most strongly to group differences by applying a univariate *F* test to the row-normalized MoSeq usage matrix. For each sex, the 60-element usage vector of every session was first normalized to sum to one. Two contrasts were then evaluated independently. In the one-versus-rest contrast, every class was compared with the pooled remainder; in the session-matched control contrast, each treatment session (APA_SES1, APA_SES2, XCX_SES1, XCX_SES2) was tested against the control session with the same session number (CON_SES1 or CON_SES2). For a given contrast, the labels were converted to a binary vector and, for every syllable, an *F* statistic and corresponding *P* value were obtained with the f_classif implementation in scikit-learn. Syllables of whose tail probability multiplied by the number of tested syllables fell below 0.01 were deemed to be significant. For every class, the full vector of *F* values, the associated *P* values and the indices of significant syllables were saved. These outputs were visualized as horizontal barcode plots in which each strip shows the *F* statistic normalized to its class-specific maximum, and the number of significant syllables is indicated in the legend.

Linear-discriminant embeddings and distance metrics. Syllable-level embeddings were derived with LDA. For each sex, the per-session MoSeq fingerprint was reduced to the 60-dimensional, row-normalized syllable-usage vector. These vectors were assembled into a sessions × syllables matrix and supplied, together with the categorical label denoting either the experimental group (control, apallial, XCX) or the full group-by-session identifier (for example, apallial-2), to an LDA model with two components for the experimental-group labels and three components for the group-by-session labels, using the eigensolver with automatic Ledoit–Wolf shrinkage. The resulting coordinates were visualized as 2D and 3D scatter plots coloured according to the palettes used throughout the study.

Session-pair distances were computed in LDA space as the Euclidean norm between the embeddings of the first and second recording session of each mouse. Only mice contributing both usable sessions were included. Distances were summarized by grey bars showing the mean ± s.e.m. overlaid with jittered points for individual animals. Group differences were evaluated using a Kruskal–Wallis test followed, where appropriate, by pairwise two-sided Mann–Whitney *U*-tests with Bonferroni correction.

To relate treatment trajectories to the baseline behaviour, we calculated, for every XCX mouse session, the distance in LDA space to the combined centroid of control mouse sessions and to the combined centroid of apallial mouse sessions. Repeated sessions were first averaged within each reference mouse, and the reference centroid was then calculated across mice so that every reference mouse contributed equally. The two distances for each XCX mouse and session were compared using a two-sided Wilcoxon signed-rank test, and the resulting *P* values were Bonferroni corrected across the two embedding metrics and two sessions within each sex.

KLD analysis. Behavioural-variability metrics were adapted from a previous study[34]. As in that study, syllable-usage distributions were summarized using KLD and CV. As experimental condition was assigned at the level of the mouse, repeated sessions and derived syllable or mouse-pair values were not treated as independent biological replicates in inferential analyses.

To quantify the within- and between-animal differences in syllable usage, we calculated the KLD between normalized syllable-occupancy distributions. For every mouse, we assembled two distributions—one from the first recording session (SES1) and one from the second (SES2)—each defined over the first 60 syllable labels and regularized with a 1×10^{-10} pseudocount to avoid undefined logarithms. The within-mouse divergence DKL(SES1||SES2) was obtained by summing $\mathbf{p} \log(\mathbf{p}/\mathbf{q})$, where $\mathbf{p}$ and $\mathbf{q}$ are the respective occupancy vectors.

Between-animal differences in syllable usage were assessed by first averaging the two sessions of every animal to a single subject-level distribution and then computing the KLD for every unique pair of animals within the same experimental group. To remove dependence on mouse ordering, inter-mouse KLD was symmetrized as $D_{\mathrm{sym}}(\mathbf{p},\mathbf{q}) = [D_{\mathrm{KL}}(\mathbf{p}\|\mathbf{q}) + D_{\mathrm{KL}}(\mathbf{q}\|\mathbf{p})]/2$. Pairwise values were retained for visualization, but statistical inference was performed by permuting whole-mouse group labels while preserving group sizes and recomputing the mean within-group dispersion. Omnibus *P* values were estimated using label permutations; pairwise *P* values were obtained from exact two-group label permutations using the absolute difference in mean dispersion and were Bonferroni-corrected for three comparisons.

CV analysis. The coefficient of variation ($\mathrm{CV} = \sigma/\mu$) was used to quantify the spread of syllable-usage probabilities. For the intramouse analysis, a CV was computed separately for every syllable across the two recording sessions, and finite syllable-level values were averaged to yield one value per mouse. Group effects were assessed using a Kruskal–Wallis test followed, when significant, by pairwise two-sided Mann–Whitney *U*-tests with Bonferroni correction. For the intermouse analysis, the two sessions of each animal were first averaged to a single subject-level distribution, after which the CV across all animals of the same experimental group was calculated syllable-wise. Syllable-level values were retained for visualization. Statistical significance was assessed by permuting whole-mouse group labels while preserving group sizes and recalculating the group CVs on every permutation. The omnibus statistic was the sample-size-weighted between-group sum of squares of the group mean CVs. Omnibus *P* values were estimated using permutations; pairwise *P* values were obtained from exact two-group label permutations using the absolute difference in mean CV and were Bonferroni-corrected for three comparisons.

Intra-session differences in syllable usage using window-to-session KLD. The temporal stability of syllable expression within a recording was assessed in overlapping 60 s windows advanced every 30 s. For each session, the framewise syllable labels were converted to a sequence of occupancy histograms (one per window; 60 syllable bins). Windows containing no frames were skipped.

Window histograms were first converted to probability vectors pw by adding a 1×10^{-9} pseudocount to every bin (to avoid zero probabilities in the KLD calculation) and then normalizing to unit sum. The session-average distribution q was computed across all windows, and the divergence for window w was $D_{\mathrm{KL}}(p_w\|q) = \Sigma i\, p_{wi} \log(p_{wi}/q_i)$. The mean KLD over windows yielded a session-level drift index. For the base-group comparisons reported in Fig. 4f and Extended Data Fig. 8o, repeated-session indices were averaged within mouse before inference.

Distributions were visualized as grey bars indicating the mean ± s.e.m. across mouse-level means, with jittered session-level points overlaid for visualization. Group effects were tested on mouse-level means with a Kruskal–Wallis test followed, when significant, by pairwise Mann–Whitney *U*-tests with Bonferroni correction.

For temporal CV, the CV of each syllable was calculated across the same overlapping 60 s windows, and finite syllable-level CVs were averaged to yield one index per session. Repeated-session indices were then averaged within mouse. Group effects were tested with a Kruskal–Wallis test followed, when significant, by pairwise two-sided Mann–Whitney *U*-tests with Bonferroni correction.

Temporal stack plot of syllable usage. For qualitative illustration of how behavioural composition evolves over the course of a recording, we generated stack plots that lay out the relative occupancy of each syllable across consecutive, non-overlapping 60 s windows (the same window length as used for the CV and KLD analyses). For a given recording, all video frames falling inside the user-defined time slice (3 s after start to the end of the file) were binned into contiguous 60 s

segments; any partial segment at the end of the file was discarded. Within each window the number of occurrences of every syllable label (from 0 to 59) was counted, divided by the total number of frames in the window and clipped to the first 60 labels to match the palette length. The resulting 60 × *N* matrix, where *N* is the number of windows, was plotted with matplotlib.stackplot, producing a coloured band for each syllable whose vertical thickness represents its moment-to-moment fractional usage.

Directed behavioural studies

Male and female mice (aged 3–5 months) were used in this study, with organoids 4–5 months after differentiation. Control mice were littermates of apallial or XCX mice but lacked Cre recombinase expression in tail DNA (that is, *Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$ genotyped through Transnetyx). For behavioural testing, mice were group-housed in a Stanford University animal facility under a reversed light–dark cycle (lights off at 08:30 and on at 20:30) with free access to food and water. All directed behavioural studies were conducted during the mouse dark cycle, following a 1 week habituation period to the facility and 3 days of handling. All procedures adhered to National Institutes of Health guidelines and were approved by the Institutional Administrative Panel on Laboratory Animal Care (APLAC). Body weight growth was measured and only mice with body weight recorded at 1, 2 and 3 months of age were included in the analysis. Mice from all conditions were excluded from behavioural analysis if there were systemic morbidities unrelated to transplantation such as paraphimosis or malocclusion.

Open-field activity chamber. The assessment took place in an open-field activity arena (Med Associates, ENV-515) mounted with three planes of infrared detectors, within a specially designed sound-attenuating chamber (Med Associates, MED-017M-027). The arena is 43 cm (*L*) × 43 cm (*W*) × 30 cm (*H*) and the sound-attenuating chamber is 74 cm (*L*) × 60 cm (*W*) × 60 cm (*H*). The mice were placed in the corner of the testing arena and allowed to explore the arena for 10 min while being tracked by an automated tracking system. Parameters including distance moved, velocity, rearing, and times spent in the periphery and centre of the arena were analysed. The periphery was defined as the zone 5 cm away from the arena wall. The arena was cleaned with a 1% Virkon solution at the end of each trial.

Y-maze spontaneous alternation assay. This test is based on the tendency of rodents to explore new environments and alternate arm entries in a Y-shaped enclosure. The arena consists of three plastic arms in a Y shape, including one long arm (A, 20.32 cm × 12.7 cm × 7.62 cm) attached to two shorter arms (B and C, 15.24 cm × 12.7 cm × 7.62 cm), separated by 120°. Mice were placed in the centre of the maze facing the intersection between arms B and C and allowed to freely explore. Using an overhead camera, the number of arm entries and alternations were recorded for 5 min. An entry is recorded when all four limbs of the mouse have entered the arm. The first entry was excluded from data analysis due to potential influence from the experimenter. The Y maze was cleaned with a 1% Virkon solution between mice. After testing, the entries are scored for the number of alternations, defined as any combination of three unique consecutive arm entries (that is, ABC or BAC, but not CAC). The percentage alternation rate is calculated as the number of alternations/the total number of possible alternations × 100. Only mice that performed at least 10 arm entries within 5 min were included in the alternation comparison analysis.

CatWalk. Gait assessment was performed using the CatWalk XT apparatus and CatWalk XT v.10.7 software (Noldus IT). The operation principles of the CatWalk were described previously[72]. During the trial, mice were allowed to walk across the glass platform (dimensions, 40 cm (length), 10 cm (width)) three times in an unforced manner. Runs across the walkway were considered successful only when the run duration was between 0.5 s and 5 s. Only one run per trial was analysed based on duration and continuity of the run. Normal rodents demonstrate three categories of step sequences: cruciate, alternate and rotate. Each category can be broken down into two unique patterns, totalling six observable step patterns. For example, cruciate can be broken down into cruciate A (Ca) and B (Cb). Ca is characterized by the following sequence of paw placements: right front (RF), left front (LF), right hind (RH), left hind (LH). Cb: LF, RF, LH, RH. Aa: RF, RH, LF, LH. Ab: LF, RH, RF, LH. Ra: RF, LF, LH, RH. Rb: LF, RF, RH, LH[73]. Mice were tested after the AC and Y-maze SA assays. Phase dispersion data are circular, ranging from −50% to +75%. For diagonal paw pairs presented in Fig. 4m, all values fell between −13% and +17%. As the values are far from the −50%/+75% wrap boundary, the circular and linear means of the groups differed by less than 0.05 (for example, 4.323% versus 4.283% for the apallial LF–RH pair). Therefore, a linear two-way ANOVA with paw-pair as a within-subject factor was used.

Three-chamber social-interaction test. The test was conducted using a rectangular plexiglass apparatus (60 cm × 40 cm × 22 cm) divided into three interconnected chambers of equal size (each 20 cm wide). Transparent walls with 5 cm × 5 cm openings allowed the mouse to move freely between chambers. Each outer chamber contained a wire cup (10 cm in diameter, 11 cm height), used to enclose a stimulus mouse or an object. To reduce stress and novelty effects, two unfamiliar stimulus mice (stranger 1 and stranger 2) were habituated to the wire cups for 30 min before testing. The test consisted of three consecutive 10 min phases. In the habituation phase, the mouse was placed into the centre chamber and allowed to explore all three empty chambers freely. In the sociability phase, stranger 1 (a sex- and age-matched unfamiliar C57BL/6 mouse) was enclosed in a wire cup and placed in the right chamber, while a novel inanimate object was placed in a wire cup in the left chamber. The mouse was allowed to explore all chambers, and the time spent interacting with stranger 1 versus the object was scored to assess sociability. In the social novelty phase, stranger 2 was introduced into a new wire cup in the right chamber, while stranger 1 was moved to the left chamber. The mouse again explored the apparatus for 10 min, and interaction times with each unfamiliar mouse were recorded to evaluate social novelty preference. New wire cups were used in each session to eliminate odour cues, and the apparatus was thoroughly cleaned with 1% Virkon between trials. When applicable, mice were tested after the AC, Y-maze SA and CatWalk assays.

Aversive Pavlovian trace conditioning procedure. The Coulbourn Instruments Aversive Conditioning System and FreezeFrame software were used for the aversive Pavlovian trace conditioning assay, which consisted of three phases: training (day 1), contextual testing (day 2) and cued testing (day 2). The training and contextual testing chambers were identical, featuring aluminium walls, a grey metal grid floor (for shock delivery), red house lights and a mint scent to create a distinct context. Chambers were cleaned with a 10% Simple Green solution (Sunshine Makers) between mice. The cued testing chamber was distinct from the training chambers, featuring circular blue plastic walls and flooring, yellow house lights, continuous 80 dB 'waterfall' background noise and a vanilla scent. These chambers were cleaned with 70% ethanol between mice. All of the chambers were enclosed in sound-attenuating chambers equipped with mounted speakers, exhaust fans and cameras for behavioural recording. On training day (day 1), the mice were individually placed into the training chamber for 3 min. A tone (20 s, 80 dB, 2 kHz) was played, and a mild foot shock (0.5 mA, 2 s) was delivered 18 s after the tone ended (that is, an 18-s trace period). This procedure was repeated three times with 60 s intervals between shocks. After the final 60 s period, the mouse was returned to its home cage. On contextual testing day (day 2), mice were placed back into the same training chamber for 5 min without any tone or shock to assess their contextual memory. For cued testing (day 2), the

mice were placed into the cued testing chamber for 3 min of habituation, followed by three tone presentations (20 s, 80 dB, 2 kHz) at 80 s intervals, without shocks, to assess cue-dependent learning. Freezing behaviour was recorded by an overhead camera and analysed using FreezeFrame software. Freezing was defined as the complete absence of movement for at least 0.75 s, and the percentage of freezing during each phase was measured. When applicable, mice were tested after the AC, Y-maze SA, CatWalk and three-chamber social-interaction tests.

Active avoidance test. The active avoidance test was conducted using the Gemini Active/Passive Avoidance System (San Diego Instruments) to assess associative learning and avoidance behaviour in mice. The apparatus consisted of a two-compartment shuttle box equipped with an automated guillotine door, a stainless-steel grid floor capable of delivering mild foot shocks, built-in speakers for auditory stimulus presentation and overhead lights. The system was enclosed in a sound-attenuating chamber, and all trials were controlled and recorded using Gemini software.

Mice were trained in the active avoidance task over three consecutive days. Each training session consisted of 50 trials per day, totalling 150 trials across the experiment. At the beginning of each testing day, mice are acclimatized for 5 min inside the apparatus and freely allowed to transition between the two compartments. The animal location was detected by the infrared sensors in the compartments. Each trial began with the presentation of the house light for 10 s as the conditioned stimulus (CS) signalling the impending foot shock. If the mouse crossed into the opposite compartment during the CS period, the response was recorded as a successful avoidance, and no shock was delivered. If the mouse failed to cross during the CS, an unconditioned stimulus was administered in the form of a mild foot shock (0.4 mA, 2 s). The shock was automatically terminated after crossing or after 2 s had elapsed. Trials were separated by a variable intertrial interval of 30 ± 5 s.

The numbers of successful avoidance responses (crossing before shock onset), escape responses (crossing after shock onset) and no response (remaining in the same compartment and received 2 s of shock) were recorded for each mouse. Chambers were cleaned with 70% ethanol between animals to minimize olfactory cues. Active avoidance experiments were conducted after aversive Pavlovian trace conditioning.

Hot plate. A hot plate apparatus (IITC, 39) set to 55 °C was used in the study. The mice were placed onto the hot plate and covered by a transparent glass cylinder that was 25 cm high and 12 cm in diameter. A 30 s maximum trial duration was used during the test to minimize the exposure of animals to painful stimuli and prevent permanent tissue damage from the heat. A remote foot-switch pad was used to control the start/stop/reset function. The latency to hind paw licking, flicking or jumping was recorded. The hot plate test was performed after the active avoidance test.

VF filament. The VF test is a method used to assess mechanical sensitivity by applying calibrated filaments of increasing or decreasing force to the skin or paw and measuring the withdrawal response threshold. The VF setup included custom built wire grid flooring (grid size 1.27 cm; 61 cm length × 41 cm width) elevated 61 cm, a clear upside-down 1 l beaker and a set of VF filaments (North Coast, Touch-Test Sensory Evaluator, NC12775). Twelve filament sizes (0.02, 0.04, 0.07, 0.16, 0.4, 0.6, 1, 1.4, 2, 4, 6 and 8 g) were used in the assessment. Mice were habituated to the VF filament test setup for 10 min per day for 3 days before testing. During the testing day, the mouse was habituated under the beaker for 10 min. The test began when the mouse stopped exploring and acclimatized to the testing environment.

The up–down method[74] was used to determine the mechanical withdrawal threshold of both left and right hind paws. Testing started with the 0.4 g filament. If the mouse did not withdraw the paw, the next higher-force filament was applied. If the mouse did withdraw, the next lower-force filament was applied. This continued until four responses after the first change in withdrawal direction were recorded. The sequence of responses (X for withdrawal, O for no response) was recorded. To ensure accuracy, at least six responses around the estimated threshold were collected. The 50% withdrawal threshold was then calculated using the formula 50% threshold $= 10^{(X_f + kd)}$. X_f represents the base-10 logarithm of the force in grams of the final filament used, k is a tabular value based on the response pattern[74] and d is the difference between the base-10 logarithms of adjacent filament forces. This calculation allowed for a statistical determination of the mouse's mechanical sensitivity.

Each filament was applied from below the wire mesh to the central plantar surface of the left and right hind paws, avoiding the footpads. The filament was pressed until it slightly bent, maintaining contact for approximately 2 s. To ensure consistency, filaments were applied only when the mouse was stationary and standing on all four paws. A minimum interval of 20 s was allowed between each filament application to avoid sensitization and ensure reliable responses. A withdrawal response was considered valid only if the hind paw was completely lifted off the platform. If the mouse walked away immediately after filament application instead of simply withdrawing the paw, the filament was reapplied. In rare cases in which a mouse flinched but did not fully lift the paw, the response was not counted as a withdrawal. The VF test was performed after the hot plate test.

OEG analysis

Mice were anaesthetized with isoflurane (1–2% in oxygen), and the scalp was surgically removed to expose the skull. The skull surface was cleaned and dried before affixing a custom aluminium head plate using dental cement (C&B Metabond, Parkell) to enable head fixation and provide optical access to the dorsal aspect of the hCO transplant. Finally, the exposed skull was coated with a thin layer of cyanoacrylate adhesive (Krazy Glue, Toagosei America) to render it more optically homogeneous. Widefield fluorescence imaging was conducted using a custom-built tandem-lens macroscope comprising a pair of camera lenses (Nikkor 50mm f/1.2, Nikon) separated by a 499/654 nm dual-band dichroic mirror (499_654 ULTRA, Alluxa) mounted in a 60 mm cage cube (LC6W, Thorlabs). To isolate GCaMP-related signals, emitted fluorescence was band-pass filtered around 520 nm (FF01-520/35-50.8-D, Semrock) before being captured on a sCMOS camera (OrcaFlash 4.0v2, Hamamatsu). For most experiments, the imaged field of view was 13.312 × 13.312 mm^2 divided into 512 × 512 pixels with a pixel size of 26 µm × 26 µm. To distinguish Ca^{2+}-related fluorescence from non-Ca^{2+}-dependent artefacts, data were acquired with alternating excitation at 470 nm (Ca^{2+} sensitive) and 405 nm on consecutive frames. GCaMP8s is significantly less sensitive to Ca^{2+} at 405 nm and demonstrates an inverted response. In a subset of high-speed-imaging experiments, a 340 × 340 pixel region of interest covering the implant was recorded at around 142 Hz (7 ms exposure) with constant 470 nm illumination. Widefield imaging data were de-interleaved to separate the 470 nm and 405 nm colour channels. Normalized fluorescence ($\Delta F/F$) was calculated for each channel by dividing each pixel's timeseries by its mean, subtracting 1, then detrending linearly. The 405 nm channel was low-pass filtered below 14 Hz and regressed pixelwise onto the signal channel. The resulting, artefact-corrected residual signal $\Delta F/F$ after regression was used for all further analyses.

The pixel-wise average timeseries across the entire imaged cortical surface was extracted and z-scored. Calcium burst onsets were delineated by finding frames where the temporal derivative of the z-scored cortex-wide activity exceeded a threshold of 0.5 and brain-wide average raw $\Delta F/F_0$ activity exceeded 10% within 10 frames thereof. If multiple nearby frames met this criterion the earliest was used. Burst offsets

were defined as the frame at which activity returned to the baseline (<1% of burst peak) and did not spike again for at least 30 s. To examine the tendency of events to recruit the entire cortical surface, we calculated the fraction of pixels participating in each event—where participation was defined as having activity exceeding 3 s.d. above its session-wise mean $\Delta F/F_0$ during the event period—for each technical replicate. A null model of burst participation was then generated by independently shuffling each pixel's timeseries and recomputing the fraction of pixels whose activity exceeded 3 s.d. above its session-wise mean $\Delta F/F_0$ during the previously defined burst periods. We repeated this shuffling procedure 1,000 times per replicate to generate an estimate of the null probability of burst participation under the assumption that pixels were uncorrelated in their activity.

Visual inspection of the time course of calcium bursts revealed that they often comprised what appeared to be multiple events with several activity peaks occurring in succession, although at variable intervals across mice. To assess the sub-burst dynamics, we used calcium deconvolution to extract the discrete, spike-like events underlying the protracted bursts. The one-dimensional pixel-wise average $\Delta F/F_0$ timeseries was linearly detrended and its minimum subtracted to ensure non-negativity, then deconvolved using the OASIS AR-2 algorithm as implemented in CaImAn to detect events. Interevent intervals were then calculated as the time between events within a calcium burst. Note that the data are quantized to 32 ms due to the acquisition frame rate.

An infrared camera (Basler ace acA1440-220um) was used to record high speed (125 Hz) videos of orofacial motor activity during calcium imaging. The total motion energy was computed as the frame-wise sum of the absolute pixel-wise temporal derivative of the greyscale videos, then normalized between 0 and 1 for each session. Plotting the average calcium and motion energy timeseries across all bursts from all sessions, we found a strong correlation between neural activity and orofacial movement. To quantitatively assess this relationship, we computed the average normalized motion energy during Ca^{2+} bursts and compared this value to that when no burst was occurring.

Acute in vivo extracellular electrophysiology

Mice were briefly anaesthetized with isoflurane, and a craniotomy was opened directly above the cortical locus that mesoscale calcium imaging had identified as the origin of spontaneous activity bursts. The animal was transferred to an air-supported Styrofoam-ball apparatus and head-fixed. A silver-chloride ground/reference wire rested in a saline pool on the skull. A motorized micromanipulator (MPC-200, Sutter Instrument) advanced a 32-channel linear silicon probe (electrode spacing: 40 µm, A1x32-Edge-10mm-40-177-CM32, NeuroNexus) through the craniotomy and into the graft. The shank had been coated with the lipophilic tracer DiI (Thermo Fisher Scientific, D282) to enable subsequent histological verification of the recording trajectory. Neural signals were amplified 10,000-fold, band-pass filtered between 0.1 Hz and 7.5 kHz, and digitized at 30 kHz using an Open Ephys acquisition board coupled to an Intan RHD headstage; data streams were written continuously to disk for offline analysis. To quantify locomotor activity during head fixation, the 20-cm Styrofoam sphere was randomly patterned with 2–5 mm black dots spaced around 20 mm apart. A monochrome GigE camera (Mako U-130B, Allied Vision) captured a 100 px × 150 px region of interest on the sphere at 30 fps throughout each recording session. Frame-to-frame displacements were estimated with an FFT-based phase-correlation algorithm. The resulting *x*- and *y*-pixel shifts were converted to physical distances with an empirically determined pixel-to-centimetre calibration factor and divided by the 33.3 ms interframe interval to yield instantaneous speed in cm s^{-1}. All recordings were performed during the light phase.

LFP burst and peak detection in extracellular electrophysiology

Broadband traces (30 kHz) were downsampled to 1.5 kHz. Bursts were flagged by visual inspection as conspicuous epochs of which the envelope exceeded background fluctuations on at least one channel. Within each burst, individual extrema were extracted with scipy.signal.find_peaks using minimum and maximum polarity-specific amplitude thresholds that were empirically adjusted for each session. Detections separated by <5 ms were merged. All peaks were subsequently reviewed and, when necessary, manually corrected by an experimenter. These deliberately conservative criteria were optimized to capture clear, large-amplitude LFP events and, as a trade-off, may have excluded smaller deflections.

Tissue preparation and immunostaining

Mice were deeply anaesthetized and transcardially perfused with PBS followed by 4% paraformaldehyde (PFA). After cryoprotection in 30% sucrose, brains were cryosectioned at a thickness of 50 µm using the Leica CM1860 cryostat. For immunostaining, free-floating tissue sections were washed three times with PBS, then blocked and permeabilized for 1 h at room temperature in PBS containing 0.3% Triton X-100 and 10% normal donkey serum. Primary antibodies were diluted in the blocking buffer and incubated with the sections overnight at 4 °C. After three washes with PBS, the sections were incubated with the corresponding Alexa Fluor secondary antibodies for 2 h at room temperature. The samples were then mounted onto microscope slides using Fluoromount-G Mounting Medium (Southern Biotech).

The primary antibodies used were: anti-CTIP2 (rat, 1:200, ab123449, Abcam), anti-SATB2 (rabbit, 1:200, ab207040, Abcam), anti-HNA (mouse, 1:100, ab191181, Abcam), anti-GFP (rabbit, 1:500, A-21311, Life Technologies), anti-IBA1 (goat, 1:100, ab5076, Abcam), anti-GFAP (rabbit, 1:2,000, Z0334, DAKO), anti-GAD65/67 (rabbit, 1:500, AB1511, Chemicon), anti-ChAT (goat, 1:200, AB144P, Millipore), anti-NeuN (rabbit, 1:200, ab104225, Abcam), anti-NeuN (mouse, 1:500, mab377, Millipore Sigma), anti-netrin-G1 (mouse, 1:100, sc-271774, Santa Cruz), anti-STEM121 (mouse, 1:200, Y40410, Takara) and anti-VGLUT1 (guinea pig, 1:5,000, CU1328, Jessel lab). Nuclei were visualized using Fluoromount-G Mounting Medium with DAPI or Hoechst 33258 (Life Technologies). Immunostained sections were imaged on a confocal microscope (Leica TCS SP8). Whole-brain dorsal views were acquired using a widefield fluorescence microscope (Leica M165 FC).

Anterograde labelling and imaging of mouse spinal cord

Mice were perfused transcardially with PBS followed by 4% PFA. The spinal cord was then exposed through dorsal laminectomy and carefully dissected out. Tissues were post-fixed in 4% PFA for 2 h, washed in PBS and segmented into cervical, thoracic and lumbar regions. The samples were cryoprotected in 30% sucrose overnight, embedded in Optimal Cutting Temperature compound and cryosectioned. Serial 20 µm transverse sections of the cervical spinal cord were collected using the Leica CM3050S cryostat.

Slides were rehydrated in PBS for three 5 min intervals, followed by overnight incubation with primary antibodies prepared in 0.2% Triton X-100 at 4 °C. The next day, the slides were washed twice in 0.2% Triton X-100 and once in PBS, each for 5 min. Secondary antibodies were then applied in 0.2% Triton X-100 and incubated for 2 h at room temperature. The slides then underwent the same washing procedure before being coverslipped.

All spinal cord sections with anterograde labelling were imaged using a confocal microscope (Leica TCS SP8, Leica Biosystems). The LasX Navigator was used to obtain high-resolution ×10 tile scans with a 2 µm *z*-step size. Maximum-intensity projections were generated from the merged tile scans in ImageJ/FIJI (ImageJ v.1.53q). Additional ×40 insets were taken in ROIs for detailed examination of neuronal projections. Control sections were obtained from an age-matched C57BL/6 mouse and control *Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$ mouse with an hCO transplanted into the intact cortex through the same procedure and stained together with the XCX samples.

Neuronal density of the XCX graft

Fixed 50 µm tissue sections were stained with NeuN and HNA antibodies, with the latter used to delineate the graft area. To quantify the density of neurons within the XCX graft, we initially defined the sampling quadrants at ×10 magnification. We next used higher magnification (×40)[75] to quantify NeuN density within one optical section. *z*-stack images were acquired at ×40 magnification using a confocal microscope (Leica STELLARIS 5), with four distinct fields of view collected per quadrant (defined according to the cardinal planes) within each section. Cell segmentation and quantification were automated using CellPose[76]. Automated analysis was performed using the pre-trained generalist algorithm Cyto3, following optimization of the input parameters. The density of NeuN$^+$ cells was calculated from the 290 × 290 µm field, and this value was multiplied by the Abercrombie correction[77] because the optical section was thinner than the size of nuclei: $t/(t + h)$. t is the thickness of the optical section (3.708 µm), and h is the mean height of the nucleus, approximated h by measuring the 2D nuclear diameter (7.8585 µm) in the X–Y plane and assuming spherical nuclei. The reported mean ± s.e.m. was calculated across three mice.

Density of GAD$^+$ cells within graft

The number of GAD$^+$ cells within the graft was analysed for two timepoints: 3 weeks after transplantation and long-term transplantation in adult animals. Cortex from non-transplanted animals was used as a control. A confocal tile-scan of a randomly selected 1 mm^2 area within the graft was acquired with a ×20 objective, and GAD$^+$ cells were manually counted. For sections from 3-week post-transplantation XCX mice that contained an area of graft <1 mm^2, a tile-scan of the entire graft was acquired and all GAD$^+$ cells were counted. For non-transplanted animals, fields of view within the cortex were randomly acquired at ×20 and GAD$^+$ cells were similarly counted.

Soma size distribution analysis

Measurement of the maximum diameter of hSYN1-eYFP$^+$/hSYN1-GCaMP8s$^+$ cells within the XCX graft was performed using whole fixed 50 µm tissue sections. Antibody staining was performed against GFP, which can detect eYFP and circularly permuted GFP (GCaMP), to amplify the signal of the fluorophore. Similar to the neuronal density analysis, images were acquired at ×40 magnification within quadrants (defined according to the cardinal planes) within each section. Ten distinct fields of view were collected per quadrant within each coronal tissue section, with *z*-stack images acquired at ×40 magnification using the Keyence BZX-710 microscope. Manual annotation of soma size was then performed using the measurement and cell counter functions within ImageJ/FIJI (ImageJ v.1.53q). The maximum diameter of up to 10 cells was recorded for each field of view, choosing cells that were in focus and completely within the field of view. This process was performed using 6 sections across 3 transplanted animals and 2 cell lines.

Relative abundance of cells with VEN-like morphology

Manual inspection at ×40 was performed to identify VEN-like cells based on their distinctive morphology (bipolar or corkscrew) and large soma size[28,29]. As with the soma size distribution analysis, 6 sections across 3 transplanted animals were used after antibody staining against GFP. To obtain the total number of hSYN1-eYFP$^+$/hSYN1-GCaMP8s$^+$ cells within each section, manual counting was performed using the cell counter function within ImageJ/FIJI. The proportion of VEN-like cells was then calculated using the total number of VEN-like and hSYN1-eYFP$^+$/hSYN1-GCaMP8s$^+$ cells.

Induction of hypoxic injury

Animals were placed into a BioSpherix chamber and a calibrated ProOx P360 oxygen controller with an animal chamber. A secondary oxygen monitor (PureAire, TX-1100-DRA) validated that the 5% O_2 setpoint was within the stated error bounds of both sensors (±1%). Nitrogen displacement was used to achieve target O_2. We empirically determined the strongest sublethal protocol while minimizing mortality. Although all food, water and bedding were removed to avoid airway obstruction, one mouse in the entire cohort died during induction due to airway obstruction from the consumption of faeces during the session. A 1-h run-in period during which atmospheric O_2 was gradually decreased from 21% to 5% was used for all experimental trials. It should be noted that we implemented a graded descent into the final oxygen concentration, enabling survival in hypoxic conditions[78–81]. Animals were kept for 5 h at 5% O_2, with visual monitoring by the experimenter. For assessment of HIF1α expression, animals were euthanized immediately at the end of the session and brains extracted rapidly on ice. For behavioural studies, the oxygen concentration was slowly allowed to rise back to 21% to prevent reperfusion injury. Brains were extracted after behavioural testing and SWI, 20 days after hypoxic exposure.

Immunostaining assessment of HIF1α expression immediately after injury

Brains were collected immediately after termination of the hypoxic period and drop-fixed in 4% PFA. Fixed tissue was processed for immunostaining as described above and sectioned at a thickness of 50 µm. HIF1α expression was assessed using an anti-HIF1α antibody (mouse, 1:200, NB100-479SS, Novus Biologicals), with anti-HNA staining used to delineate the graft. Incubation with secondary antibodies was performed in the same manner as described above. Initial imaging consisted of whole-section overviews, followed by comprehensive manual scanning at high magnification to assess cellular localization. Sampling was focused primarily in the area adjacent to the graft–host interface, with efforts to keep the sampling area consistent across animals and conditions. Expression in the graft region was compared with the residual host palaeocortex within the same animal. Two control conditions were included: normoxic XCX controls (no hypoxic exposure) and hypoxia-exposed controls without transplantation.

Migration of hCO-derived cells into host brain

The number of hCO-derived cells that migrated outside the borders of the graft and into the adjacent host brain was assessed using 50-µm-thick immunostained sections. The graft area was defined by expression of contiguous HNA, which marked its boundaries. We analysed the number of HNA$^+$ cells beyond the graft using manual counting using the cell counter plugin in FIJI/ImageJ (ImageJ v.1.53q).

To obtain the total number of HNA$^+$ cells within the graft itself, cell segmentation and quantification were performed using CellPose[76]. Segmentation was reviewed and masks were adjusted manually as needed.

Density and ramification of GFAP$^+$ astrocytes and IBA1$^+$ microglia

To assess the baseline inflammatory state of the XCX graft, host-derived (that is, HNA$^-$) GFAP$^+$ astrocytes and IBA1$^+$ microglia were analysed in host tissue immediately adjacent to the graft. Analyses were restricted to HNA$^-$GFAP$^+$ astrocytes and IBA1$^+$ microglia within mouse tissue to assess the host response to the transplanted graft. The control group (*Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$) was sampled from an approximately matched location in the cortex. The cell density and number of primary processes were analysed for GFAP$^+$ astrocytes and IBA1$^+$ microglia. For both cell populations, confocal images were acquired from the residual host palaeocortex immediately adjacent to the graft. For cell density, the number of cells was manually counted within images acquired with a ×20 objective, with three animals used per condition (XCX and non-transplanted). For ramification, *z*-stacks were acquired at ×40 for randomly selected cells within this same area. The number of primary processes extending from the cell body was manually quantified from optical section stacks for each GFAP$^+$ and IBA1$^+$ cell.

For assessment of the inflammatory response to hypoxic injury, intragraft GFAP$^+$ (human, graft-derived) and IBA1$^+$ (host-derived, mouse)

cell density was analysed. To be certain that the inflammatory signal analysed was inside the graft, images were acquired at least 1 mm away from the host–graft interface (that is, there was a 1 mm buffer zone on the diameter of the graft in which no data were acquired). Astrocyte ramification was assessed using manual counting. The soma area of $IBA1^+$ cells was manually segmented in FIJI/ImageJ.

snRNA-seq analysis

Whole-brain snRNA-seq. Intact, whole mouse brains were extracted and rapidly placed in an ice-cold solution containing 75 mM sucrose, 87 mM NaCl, 2.5 mM KCl, 0.5 mM $CaCl_2$, 1.25 mM NaH_2PO_4, 7 mM $MgCl_2$, 25 mM $NaHCO_3$, 1 mM Na-ascorbate and 10 mM D-glucose. Single-nucleus dissociation was performed based on a previous report[82], and the buffer/solution names below refer to specific buffers enumerated in their supplemental materials. In brief, each brain was homogenized using a separate 15 ml Dounce homogenizer (Active Motif, 40415) on ice containing the low sucrose buffer. After centrifugation at 4 °C at 900*g* for 5 min, the supernatant was discarded, and the pellet was resuspended in 25% iodixanol solution (Stem Cell Technology, 07820). The solution was split into two centrifugation tubes and 29% iodixanol solution was layered underneath. After centrifugation at 4 °C at 3,200*g* for 20 min using a spinning-bucket centrifuge, the supernatant was discarded, the pellet resuspended in 1% BSA in PBS supplemented with 0.2 U ml^{-1} RNase inhibitor (Sigma-Aldrich, 3335399001) and nucleus fixation was performed (Parse Biosciences, ECF2003 Nuclei fixation v2). The samples were frozen down in a freezing container (Sigma-Aldrich, C1562) and stored at −80 °C until the barcoding and library preparation. We performed SPLiT-seq according to the manufacturer's recommendations adjusting the centrifugal speed to 450*g* (Parse Biosciences, ECW02050 Evercode WT Mega Kit v2). Two control (*Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$) and two apallial (*Emx1-cre*$^{+/-}$*Esco2*$^{fl/fl}$*Prkdc*$^{scid/scid}$) brains were analysed, and the sample conditions were balanced within a plate to avoid batch effects (one control and one apallial brain per plate, two plates). Samples were sequenced by Admera Health on a NovaSeq S4 2 × 150 (Illumina).

Gene expression levels were quantified for each putative nucleus using the Parse Biosciences analysis software suite (split-pipe command, v.1.1.1). Specifically, reads were mapped to a mouse (GRCm39, Ensembl release 110) reference genome created using the mkref mode and quantified using the default parameters. All subsequent analyses were performed on the filtered count matrices using the R (v.4.3.2) package Seurat (v.5.0.1)[83]. To ensure that only high-quality nuclei were included for downstream analyses, an iterative filtering process was implemented for each sample. First, low-quality nuclei with less than 200 unique genes detected and with mitochondrial counts accounting for greater than 5% of the total counts were identified and removed. Subsequently, raw gene count matrices were normalized using the default NormalizeData function Seurat workflow, which log transforms counts divided by the total counts of each nucleus multiplied by a scale factor. Given the large numbers of profiled nuclei, we implemented a memory efficient downsampling approach using dataset sketching as implemented in Seurat to preprocess each sample. Specifically, each sample was downsampled to 100,000 nuclei using the SketchData function followed by dimensionality reduction using PCA on the top variable genes. Clusters of nuclei were identified in PCA space by shared nearest-neighbour graph construction and modularity detection implemented by the FindNeighbors and FindClusters functions using a dataset dimension of 100. Cluster labels and dimensional reductions were then extended to the full dataset using the ProjectData function. We categorized clusters as neuronal or non-neuronal by reference mapping using Seurat's TransferData workflow on an adolescent whole CNS mouse scRNA-seq dataset[12]. We next performed iterative filtering on neuron and non-neuronal cell classes for each sample separately. After clustering (resolution = 1), putative low-quality cells or doublets were identified and removed based on outlier low gene counts (median below the 10th percentile), outlier high-fraction mitochondrial genes (median above the 95th percentile) or high co-expression of both neuronal (*Rbfox3*) or non-neuronal class markers (*Aqp4*, *Mag*, *Csf1r*, *Pdgfra*, *Ranbp3l*). Lastly, samples were integrated across the two Parse Mega kits (each containing a control and apallial mouse) using sketch-based reciprocal PCA integration (downsampled to 100,000 nuclei for each kit). For visualization purposes, integrated nuclei were embedded in two dimensions using UMAP. On the filtered dataset, cell class and subclass annotations were obtained using MapMyCells (performed on 30 April 2025, using Allen Brain Institute website portal), which uses a hierarchical mapping algorithm, on each nucleus independently, to annotate query nuclei to a whole mouse brain atlas, termed the Allen Brain Cell Atlas[9]. A subsequent round of filtering was performed to remove nuclei that poorly mapped to cell subclass annotations in the Allen Brain Cell Atlas (bootstrap probability < 0.5). Differential cell abundance analysis between control and apallial cell class annotations was performed using the EdgeR package[84], which uses negative binomial generalized linear methods to model the annotation counts across conditions. Cell annotations were normalized to the total number of cells in each sample. Differential gene expression analysis was performed using a pseudobulk approach for each cortical GABAergic subclass independently. Specifically, raw counts were summed across cells within each subclass for each sample and differential expression was performed using the edgeR quasi-likelihood framework with trimmed mean of *M*-values normalization.

snRNA-seq analysis of engrafted neurons. Single-nucleus isolation was conducted as previously outlined using sucrose-gradient buffers[3]. In brief, flash-frozen XCX samples were homogenized in a low-sucrose cell lysis buffer containing 0.1% Triton-X, using a 2 ml glass tissue grinder (Sigma-Aldrich/KIMBLE, D8938) on ice. The resulting crude nuclei were filtered through a 40 µm filter and centrifuged at 320*g* (Eppendorf, 5810R) in a 50-ml Falcon tube for 10 min at 4 °C. After the removal of the supernatant, the nucleus pellets were resuspended in 3 ml of low-sucrose buffer. A 12.5 ml high-sucrose solution was then carefully layered beneath the cell resuspension, followed by a final centrifugation at 320*g* for 20 min at 4 °C without brake. After discarding the supernatant, the samples were resuspended in 0.04% BSA/PBS containing 0.2 U $µl^{-1}$ RNase inhibitor (40 U $µl^{-1}$, Ambion, AM2682). A targeted recovery of 8,000 nuclei per sample was loaded onto the Chromium Next GEM Chip G. Dual-index snRNA-seq libraries were generated using the Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3.1 (10x Genomics). Libraries from different samples were pooled in equal molar ratios and sequenced by Admera Health on a NovaSeq S4 2×150 (Illumina), followed by trimming to 28 × 10 × 10 × 90 nucleotides.

Gene expression levels were quantified for each putative nucleus barcode using the 10x Genomics CellRanger analysis software suite (v.7.1.0). Specifically, reads were mapped to a combined human (GRCh38, Ensembl release 98) and mouse (mm10, Ensembl release 98) reference genome created using the mkref command and quantified using the count command with --include-introns=TRUE to include reads mapping to intronic regions. Human nuclei were identified based on a conservative requirement of at least 95% of total mapped reads aligning to the human genome, as we have described previously[3,13]. All subsequent analyses were performed on the filtered barcode matrices outputted from CellRanger using the R (v.4.1.2) package Seurat (v. 4.1.3)[85]. To ensure that only high-quality nuclei were included for downstream analyses, an iterative filtering process was implemented for each sample. First, low-quality nuclei with less than 1,000 unique genes detected and with mitochondrial counts accounting for greater than 5% of the total counts were identified and removed. Subsequently, raw gene count matrices were normalized by regularized negative binomial regression using the sctransform function (vst.flavor = "v2"), which also identified the top 3,000 highly variable genes using the default parameters. Dimensionality reduction using PCA on the top variable genes was performed, and clusters of nuclei were identified in

PCA space by shared nearest-neighbour graph construction and modularity detection implemented by the FindNeighbors and FindClusters functions using a dataset dimension of 30 (dims = 30, chosen based on visual inspection of elbow plot and used for all samples and integration analyses) with the default parameters. We subsequently performed iterative rounds of clustering (resolution = 2) to identify and remove clusters of putative low-quality cells or doublets based on outlier low gene counts (median below the 10th percentile), outlier high-fraction mitochondrial genes (median above the 95th percentile) or outlier doublet score (median above the 95th percentile)[86]. The samples were integrated using canonical correlation analysis as implemented with the FindIntegrationAnchors and IntegrateData functions with the above parameters. After removal of low-quality cells, the integrated dataset was clustered (FindClusters function; resolution = 0.5) and embedded for visualization purposes with UMAP. We identified and categorized clusters through a combination of marker gene expression and annotation through reference mapping using Seurat's TransferData workflow to classify organoid cells using the following annotated human datasets: adult motor cortex snRNA-seq[24], adult middle temporal gyrus snRNA-seq[87], fetal whole brain scRNA-seq[88] and fetal second-trimester cortex snRNA-seq[15]. Specifically, progenitor clusters were identified by the expression of *MKI67*, *TOP2A* and *EGFR*. Astrocyte clusters expressed high levels of *SLC1A3* and *AQP4* and mapped to adult astrocytes. OPCs expressed *PDGFRA* and *SOX10*. Glutamatergic neuron clusters were further classified into subclasses by mapping to annotations defined by the reference adult datasets except for L6b neurons, which also contained markers of subplate (SP) (*ST18*) and were therefore labelled L6b/SP. A small cluster of glutamatergic neurons did not obviously map to populations in every reference dataset and were labelled GluN_other.

To characterize the transcriptomic maturation state of XCX glutamatergic neurons, we compared our pseudobulk samples with pseudobulk glutamatergic neuron profiles spanning primary human cortical development from two published datasets[16,17]. We performed PCA on the combined sample counts-per-million-normalized gene expression matrix, subset to 3,000 highly variable genes calculated from the ref. 16 dataset using the sctransform function and shared across all datasets. To estimate the developmental age of our XCX neurons, we constructed a linear model using PC1 scores and $\log_2$-transformed gestational day information from the ref. 16 samples. Testing our model on the held-out ref. 17 human cortical development dataset showed better performance (higher r^2 and lower mean absolute error) when restricting to samples younger than 4 years of postnatal age. Predicted XCX graft transcriptomic ages were derived from this linear model.

MERFISH sample collection and analysis

MERFISH analysis was performed using the Vizgen MERSCOPE platform according to the manufacturer's instructions. We designed a custom 500-gene panel by cross-referencing previously published human and mouse brain MERFISH gene panels and published bulk RNA-seq data from developing human brain (Supplementary Table 4). Specifically, the panel incorporated: (1) the 140-gene human adult cortex panel designed by Allen Institute[89]; (2) around 100 shared genes present on both Vizgen's 500-gene PanNeuro mouse panel and the Allen Institute's whole brain 500-gene panel[9]; and (3) manually curated genes selected to maximize coverage of neural cell types and developmental signatures. To finalize gene selection, we used bulk RNA-seq reads per kilobase of transcript per million mapped reads (RPKM) expression values from published developing human frontal cortex data[90], ensuring that the summed RPKM values of all panel genes remained below 10,000 to minimize optical crowding per Vizgen recommendations. The panel was originally designed for transplanted human cortical organoids (t-hCO) into rat cortex and therefore includes 489 genes targeting human transcripts and 11 genes targeting the rat transcriptome to enable computational identification of host versus graft cells. Each gene is represented by 50 encoding probes with 30-nucleotide target regions, designed by Vizgen to maximize the sequence divergence between human and host species orthologues. To assess species specificity of the human-targeting probes for mouse, we performed pairwise local alignment (Biostrings, R/Bioconductor) of each 30-nucleotide probe against the longest annotated mouse orthologue transcript isoform obtained by Ensembl (biomaRt). Of the 489 human-targeting genes, 485 had identifiable mouse orthologues, and 322 of these (66.4%) had zero probes with a perfect full-length match (30 out of 30 nucleotides) to the mouse orthologue. An additional 161 genes had only 1–5 perfect-match probes out of 50, and the maximum observed for any gene was 12 (*LMO1*). As published MERFISH benchmarks have shown reduced detection efficiency with fewer than 16 probes per gene[91], this is consistent with human-targeting probes having minimal cross-reactivity with mouse host transcripts.

Whole brains were extracted as described above, embedded in a 1:1 solution of 30% sucrose in PBS:OCT (Tissue-Tek) and flash-frozen using either dry ice or within an isopentane bath pre-chilled on dry ice. Using a cryostat (Leica, CM1860), 10 µm sections were collected and adhered to a MERSCOPE Slide (Vizgen, 20400001) then placed into a 6 cm Petri dish. The slide was left in the cryostat at −20 °C for 25–30 min to allow the section to fully adhere to the slide. The samples were fixed directly on the slide by incubating with 4% PFA for 15 min at room temperature. The slide was then washed three times with 5 ml of 1× PBS for 5 min each and allowed to dry at room temperature for 1 h. The slide was stored in 5 ml of 70% ethanol at 4 °C for a minimum of 24 h and a maximum of 1 month before proceeding with sample processing.

The samples were processed according to the MERSCOPE User Guide for Fresh and Fixed Frozen Tissue Sample Preparation (Vizgen, 91600002 Rev F). In brief, the samples were incubated for 36–48 h with 50 µl of MERSCOPE Gene Panel Mix (Vizgen, 10400003) at 37 °C in a humidified incubator. After probe hybridization, the samples were embedded in a 4% polyacrylamide gel (Vizgen, 20300004) and incubated in Clearing Solution (Vizgen, 20300003) at 37 °C for 24 h. After clearing, the samples were then incubated with 3 ml DAPI and PolyT Staining Reagent (Vizgen, 20300021) for 15 min at room temperature on a rocker while protected from light. The samples were then incubated with 5 ml formamide wash buffer (Vizgen, 20300002) for 10 min at room temperature and subsequently moved to 5 ml sample prep wash buffer (Vizgen, 20300001). The samples were imaged according to the MERSCOPE Instrument User Guide (Vizgen, 91600001 Rev G) using a MERSCOPE 500 Gene Imaging Cartridge (Vizgen, 20300019).

Raw imaging data were processed using the Vizgen analysis pipeline based on the MERlin python package[92], which included cell segmentation with Cellpose[76] based on DAPI and PolyT staining. The resulting cell-by-gene count matrix was filtered to include human cells (>95% of total counts from human genes) with at least 50 mRNA molecules to remove putative low-quality cells. Data were preprocessed using the Seurat (v.5.0.1) R package workflow. Specifically, we performed sctransform-based normalization with modified clipping parameters (−10, 10) to account for the effect of outliers, as recommended by the Seurat authors for single-molecule FISH experiments. Dimensionality reduction using PCA on human genes was performed, and cells were annotated using Seurat's TransferData workflow with our XCX graft snRNA-seq dataset as reference.

Statistics and reproducibility

All numerical data are expressed as means ± s.e.m. unless otherwise specified, with biological replicate (that is, mouse) serving as the unit of replication (n) when possible. Experiments and analyses were conducted in a blinded manner whenever feasible. Sample sizes were estimated empirically, based on previous studies[3]. Randomization was used in the stereological analysis of soma size. Images were acquired from an equally distributed grid throughout the tissue for downstream analysis. Randomization of subjects was performed to ensure that both cell lines were represented.

When s.d. values were significantly different between groups, Welch's ANOVA was used. All *t*-tests were two-tailed. Post hoc testing was only performed for main effects or interactions shown to be significant in the overall ANOVA, with Holm–Sidak correction for multiple comparisons unless otherwise specified. Geisser–Greenhouse's correction was applied whenever applicable. Data were log-transformed if a *Q*–*Q* plot, residuals plot and homoscedasticity plot demonstrated that the residuals of fitting a parametric model were skewed. If log-transformation was required and the dataset contained zeros, we added a constant of 1 to all values before log transformation. There was one exception: while the omnibus *F*-test underwent Geisser–Greenhouse correction, the post hoc test for Extended Data Fig. 7t did not due to technical limitations related to fractional degrees of freedom. Statistical analyses of the MRI data, immunostaining, body weights, breeding strategy and targeted behavioural approaches were carried out using Prism 10 (GraphPad).

Group × sex interactions were tested for all the directed behavioural tasks in Fig. 4 and Extended Data Fig. 9. We did not find any significant interactions, so data from male and female mice were pooled. These studies were not powered to detect modest effects of sex, but 3 out of 48 directed behavioural results (CatWalk swing duration, Y-maze distance, AC vertical time) revealed a main effect of sex and details are presented in Supplementary Table 5. Sex × genotype effects were not estimable in the hypoxic injury experiments because the apallial group contained only female mice. Sex was therefore assessed in the two groups where it could be: a genotype × sex ANOVA restricted to control and XCX mice revealed no main effect of sex and no genotype × sex interaction for any of the CatWalk measures in Fig. 5k–n and Extended Data Fig. 10f–i. Sexes were pooled for the analyses shown.

Three datasets are compositional, with components summing to 100% within each animal: paws on ground simultaneously (Fig. 4l), step-sequence usage (Fig. 4n) and fixel direction fractions (Extended Data Fig. 3j). As each animal shares the same mean across components, between-subject main effects are uninformative and were not reported. Compositional distributions are shown descriptively, and statistical testing was applied to a contrast that is less affected by the compositional nature of the data. For step sequences, the contrast analysed was the alternate:cruciate log-ratio (Fig. 4o) because rotate sequences were negligible. For fixel direction fractions, control mice exhibited an equal proportion of all three principal directions, so the axis × group interaction was used, and each axis was also compared against the null hypothesis of no directional preference (33.3%). The proportion of time using different paw support strategies (Fig. 4l) is presented descriptively.

Micrographs depict representative results and were repeated with the following *n* values: Fig. 2n,o (3 mice); Fig. 2p (3 mice); Fig. 3b (3 mice); Fig. 5b (3 mice); Extended Data Fig. 1e (1 mouse per condition); Extended Data Fig. 3b (3 mice); Extended Data Fig. 3c (1 mouse); Extended Data Fig. 3f (3 mice); Extended Data Fig. 7d (3 sagittal sections/1 mouse, presence of subcortical fibre tracts replicated in coronal sections from 3 mice); Extended Data Fig. 7e,f (3 mice); Extended Data Fig. 7j (5 mice); Extended Data Fig. 7m,n (3 mice); Extended Data Fig. 7o (1 mouse); Extended Data Fig. 7p,q (3 mice); Extended Data Fig. 10a (4 mice); Extended Data Fig. 10b (2 mice); Extended Data Fig. 10c (4 mice); and Extended Data Fig. 10d (2 mice).

Ethics statement

All experiments involving human cells were predicated on ethical guidelines and safeguards and were adherent to all relevant guidelines and regulations, including the International Society for Stem Cell Research (ISSCR). Human donors in this study consented to the use of their cells to generate hiPS cells and derived cells, including for in vivo transplantation. The source of the cells and their institutional approvals are listed in the reporting summary. This study benefited from prospective and ongoing consultation with bioethicists at Stanford University, review by the Stanford Neuroscience Institute Executive Committee, and evaluation and guidance provided by an external independent ad hoc ethics committee. Experiments were shared transparently and discussed, including the scientific rationale for the work and sequence of studies, laboratory practices and safeguards, ethical dimensions of study design, methods and data interpretation, the possibility of altered or improved capacities in the experimental mice, and potential social, medical and ethical implications of the work. Approval for transplantation of human cortical organoids into mice was obtained from the Stanford Stem Cell Research Oversight (SCRO) committee and Institutional Animal Care and Use Committee (IACUC). To assess altered or improved capacities in the experimental mice, systematic and comprehensive locomotor and cognitive testing was performed and their wellbeing was monitored throughout. In addition to fully adhering to all relevant guidelines and regulations, we recommend that all future experiments with this model should be performed in close consultation with ethicists in designing, conducting, interpreting and communicating this work.

Reporting summary

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

Data availability

Single-nucleus gene expression data are available at Gene Expression Omnibus (GSE301194). Due to participant consent and privacy, we are not able to make raw sequencing data public for the XCX samples but they are available from the authors on reasonable request. Spatial transcriptomics data have been deposited at Zenodo[93] (https://doi.org/10.5281/zenodo.21138437) and are available as of the date of the publication. The following public datasets were used for snRNA-seq analysis: human genome sequence information from Ensembl (http://ftp.ensembl.org/pub/release-98/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz) and human gene annotation from GENCODE (http://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_32/gencode.v32.primary_assembly.annotation.gtf.gz); information on the mouse genome sequence is from Ensembl (http://ftp.ensembl.org/pub/release-98/fasta/mus_musculus/dna/Mus_musculus.GRCm38.dna.primary_assembly.fa.gz); the Allen Brain Institute generated human adult snRNA-seq data from the primary motor cortex (https://portal.brain-map.org/atlases-and-data/rnaseq; accessed May 2022); the HNOCA (https://datasets.cellxgene.cziscience.com/83b34db4-0319-44ee-b81d-815e3005be30.h5ad; accessed March 2026); transplanted cortical organoid into adult mouse snRNA[23] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE185472; accessed February 2026). The following public datasets were used for mouse whole-brain snRNA-seq analysis: mouse genome sequence information from Ensembl (https://ftp.ensembl.org/pub/release-110/fasta/mus_musculus/dna/Mus_musculus.GRCm39.dna.primary_assembly.fa.gz); mouse brain scRNA-seq atlas from ref. 12 (https://storage.googleapis.com/linnarsson-lab-loom/l5_all.loom; accessed March 2022). DWI data are available at the OSF[94] (https://osf.io/5sdx6/overview). All other data are available as source datasheets.

Code availability

Custom codes to analyse DWI data are available at GitHub (https://github.com/garikoitz/apallial_dmri). Otherwise, we did not generate novel analysis pipelines and refer to previously published code.

53. Paşca, A. M. et al. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat. Methods* **12**, 671–678 (2015).
54. Yoon, S.-J. et al. Reliability of human cortical organoid generation. *Nat. Methods* **16**, 75–78 (2019).
55. Zhang, Y. et al. Fast and sensitive GCaMP calcium indicators for imaging neural populations. *Nature* **615**, 884–891 (2023).

56. Fenno, L. E. et al. Comprehensive dual-and triple-feature intersectional single-vector delivery of diverse functional payloads to cells of behaving mammals. *Neuron* **107**, 836–853 (2020).
57. Gordon, J. A. & Stryker, M. P. Experience-dependent plasticity of binocular responses in the primary visual cortex of the mouse. *J. Neurosci.* **16**, 3274–3286 (1996).
58. Li, X., Morgan, P. S., Ashburner, J., Smith, J. & Rorden, C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J. Neurosci. Methods* **264**, 47–56 (2016).
59. Tournier, J.-D. et al. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* **202**, 116137 (2019).
60. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
61. Tustison, N. J. et al. The ANTsX ecosystem for quantitative biological and medical imaging. *Sci. Rep.* **11**, 9069 (2021).
62. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
63. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
64. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
65. Brett, M. et al. nipy/nibabel: 5.4.1 (version 5.4.1). *Zenodo* https://doi.org/10.5281/zenodo.18944560 (2026).
66. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
67. Veraart, J. et al. Denoising of diffusion MRI using random matrix theory. *Neuroimage* **142**, 394–406 (2016).
68. Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A. & Sijbers, J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *Neuroimage* **103**, 411–426 (2014).
69. Raffelt, D. A. et al. Investigating white matter fibre density and morphology using fixel-based analysis. *Neuroimage* **144**, 58–73 (2017).
70. Tournier, J. D., Calamante, F. & Connelly, A. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. *Proc. Intl. Soc. Mag. Reson. Med.* **18**, 1670 (2010).
71. Bogovic, J. A., Hanslovsky, P., Wong, A. & Saalfeld, S. Robust registration of calcium images by learned contrast synthesis. In *Proc. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* 1123–1126 (IEEE, 2016).
72. Wang, X. H. et al. Quantitative assessment of gait and neurochemical correlation in a classical murine model of Parkinson's disease. *BMC Neurosci.* **13**, 142 (2012).
73. Hamers, F. P. T., Koopmans, G. C. & Joosten, E. A. J. CatWalk-assisted gait analysis in the assessment of spinal cord injury. *J. Neurotrauma* **23**, 537–548 (2006).
74. Chaplan, S. R., Bach, F. W., Pogrel, J. W., Chung, J. M. & Yaksh, T. L. Quantitative assessment of tactile allodynia in the rat paw. *J. Neurosci. Methods* **53**, 55–63 (1994).
75. Mouton, P. R. *Unbiased Stereology: A Concise Guide* (JHU Press, 2011).
76. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
77. Abercrombie, M. Estimation of nuclear population from microtome sections. *Anat. Record* **94**, 239–247 https://doi.org/10.1002/ar.1090940210 (1946).
78. Cristancho, A. G. et al. Deficits in seizure threshold and other behaviors in adult mice without gross neuroanatomic injury after late gestation transient prenatal hypoxia. *Dev. Neurosci.* **44**, 246–265 (2022).
79. Ilacqua, A. N., Kirby, A. M. & Pamenter, M. E. Behavioural responses of naked mole rats to acute hypoxia and anoxia. *Biol. Lett.* **13**, 20170545 (2017).
80. Nakada, Y. et al. Hypoxia induces heart regeneration in adult mice. *Nature* **541**, 222–227 (2017).
81. Magalhães, J. et al. Acute and chronic exposition of mice to severe hypoxia: the role of acclimatization against skeletal muscle oxidative stress. *Int. J. Sports Med.* **26**, 102–109 (2005).
82. Gautier, O. et al. Challenges of profiling motor neuron transcriptomes from human spinal cord. *Neuron* **111**, 3739–3741 (2023).
83. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
84. Chen, Y., Chen, L., Lun, A. T., Baldoni, P. L. & Smyth, G. K. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res.* **53**, gkaf018 (2025).
85. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
86. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).
87. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
88. Braun, E. et al. Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, eadf1226 (2023).
89. Gabitto, M. I. et al. Integrated multimodal cell atlas of Alzheimer's disease. *Nat. Neurosci.* **27**, 2366–2383 (2024).
90. Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
91. Xia, C., Babcock, H. P., Moffitt, J. R. & Zhuang, X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci. Rep.* **9**, 7721 (2019).
92. Emanuel, G., Eichhorn, S. & Zhuang, X. MERlin-Scalable and extensible MERFISH analysis software, v0.1.6. *Zenodo* https://doi.org/10.5281/zenodo.3758539 (2020).
93. Kaganovsky, K. et al. MERFISH data for 'Developmental xenocortication using human-derived organoids in mice'. *Zenodo* https://doi.org/10.5281/zenodo.21138437 (2026).
94. Kaganovsky, K. et al. DWI data for 'Developmental xenocortication using human-derived organoids in mice'. *OSF* https://osf.io/5sdx6/overview (2026).
95. Chen, J. G. et al. Giotto Suite: a multiscale and technology-agnostic spatial multiomics analysis ecosystem. *Nat. Methods* **22**, 2052–2064 (2025).

Acknowledgements We thank M. Shamloo, N. Saw and the members of the Stanford Behavioral and Functional Neuroscience Laboratory for behavioural testing; M. Abe and the staff at the Neurosciences Preclinical Imaging Lab for MRI support; M. Onesto for immunostaining and imaging support; and N. Sakai for the establishment and management of the breeding colony.

Author contributions S.-J.Y. maintained and differentiated hiPS cells. K.K. and O.R. performed the transplantation experiments. X.C., A.V. and K.K. prepared the single-nucleus profiling experiments. K.W.K. performed the scRNA-seq analysis. K.K., K.W.K. and N.R. performed and analysed the MERFISH experiments. P.M.H., K.K., A.A., J.L.S. and J.A.K. performed, analysed and supervised immunocytochemistry and imaging. K.K., A.A. and S.O. performed MRI. G.L.-U. and B.W. analysed the DWI data. T.G., F.G. and I.S. conducted, analysed and supervised in vivo LFP recordings. J.K., A.D.W. and K.D. performed, analysed and supervised the whole-cortex GCaMP imaging. A.A., S.O. and K.K. monitored animal wellbeing. K.K., P.M.H. and S.O. performed rabies tracing. T.G. and I.S. performed, analysed and supervised MoSeq behavioural characterization. P.M.H. and H.V. characterized the cellular response to hypoxic injury. K.K., K.W.K., T.G., P.M.H. and S.P.P. conceived the project and designed the experiments. K.K., K.W.K., T.G., P.M.H. and S.P.P. wrote the manuscript with input from all of the authors. S.P.P. supervised all aspects of the work.

Funding This work was supported by the Stanford Big Idea Project on Brain Organogenesis (Wu Tsai Neuroscience Institute) (to S.P.P. and K.D.), the Kwan Funds (to S.P.P.), the Senkut Funds (to S.P.P.) and Brain & Behavior Research Foundation and Evelyn Toll Family Foundation (to K.W.K.). S.P.P. is a CZI Ben Barres Investigator and a CZ BioHub Investigator.

Competing interests Stanford University holds patents for the generation of cortical organoids (listing S.P.P. as an inventor) and a provisional patent application for transplantation of organoids (listing S.P.P., O.R., F.G., K.D. and K.W.K. as inventors).

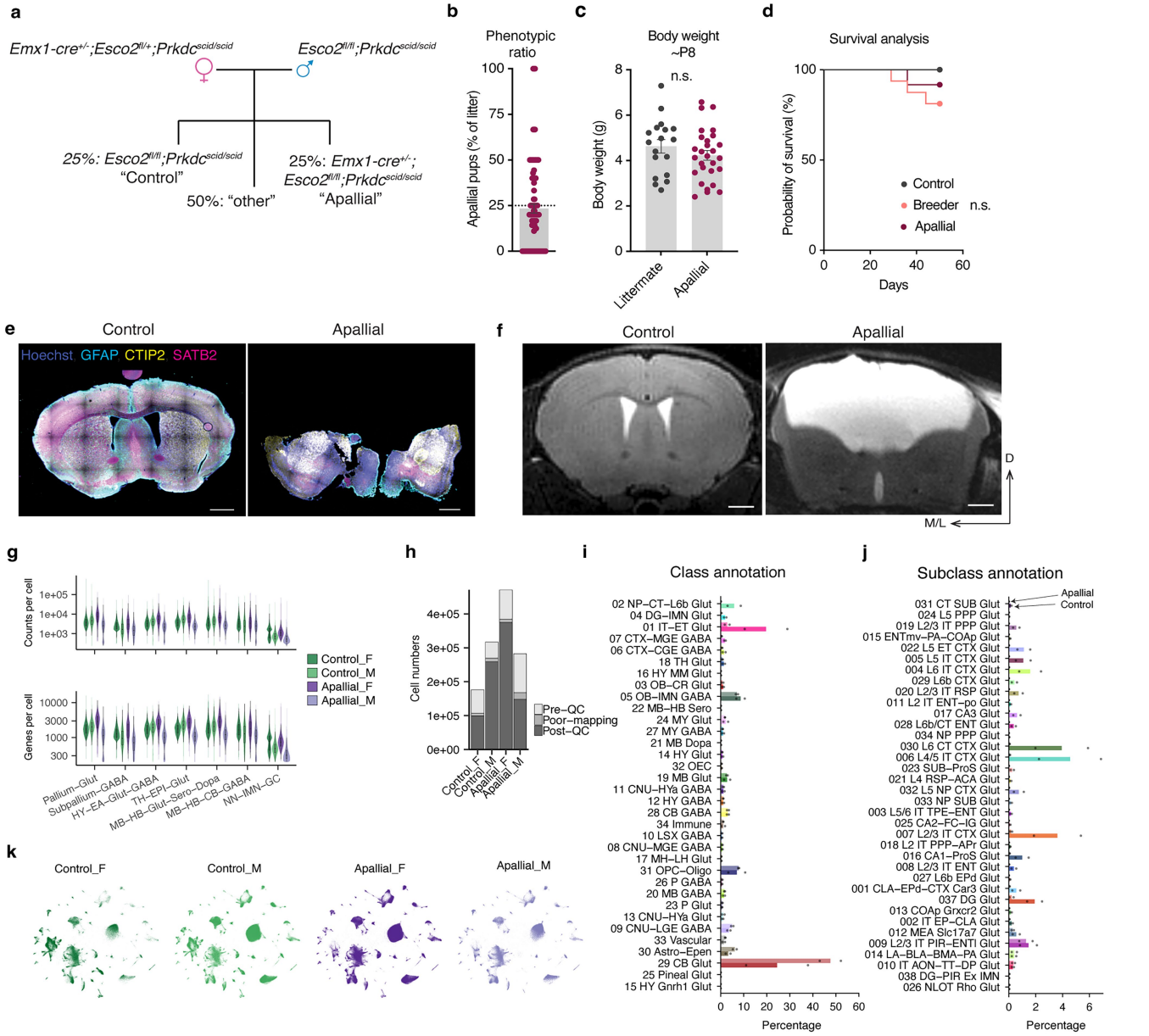
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-11032-2.

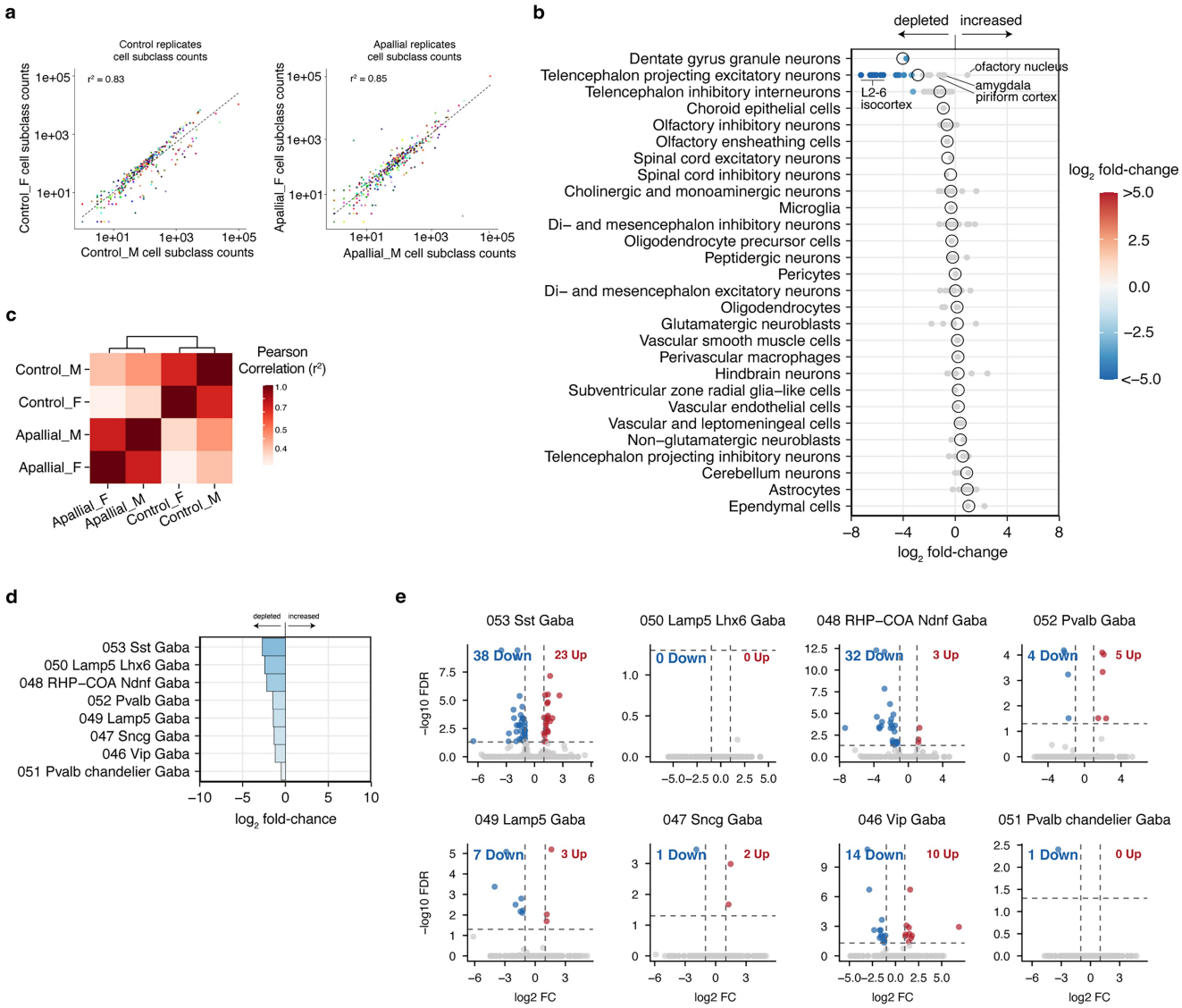
Correspondence and requests for materials should be addressed to Sergiu P. Paşca.

Peer review information *Nature* thanks Laurent Nguyen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

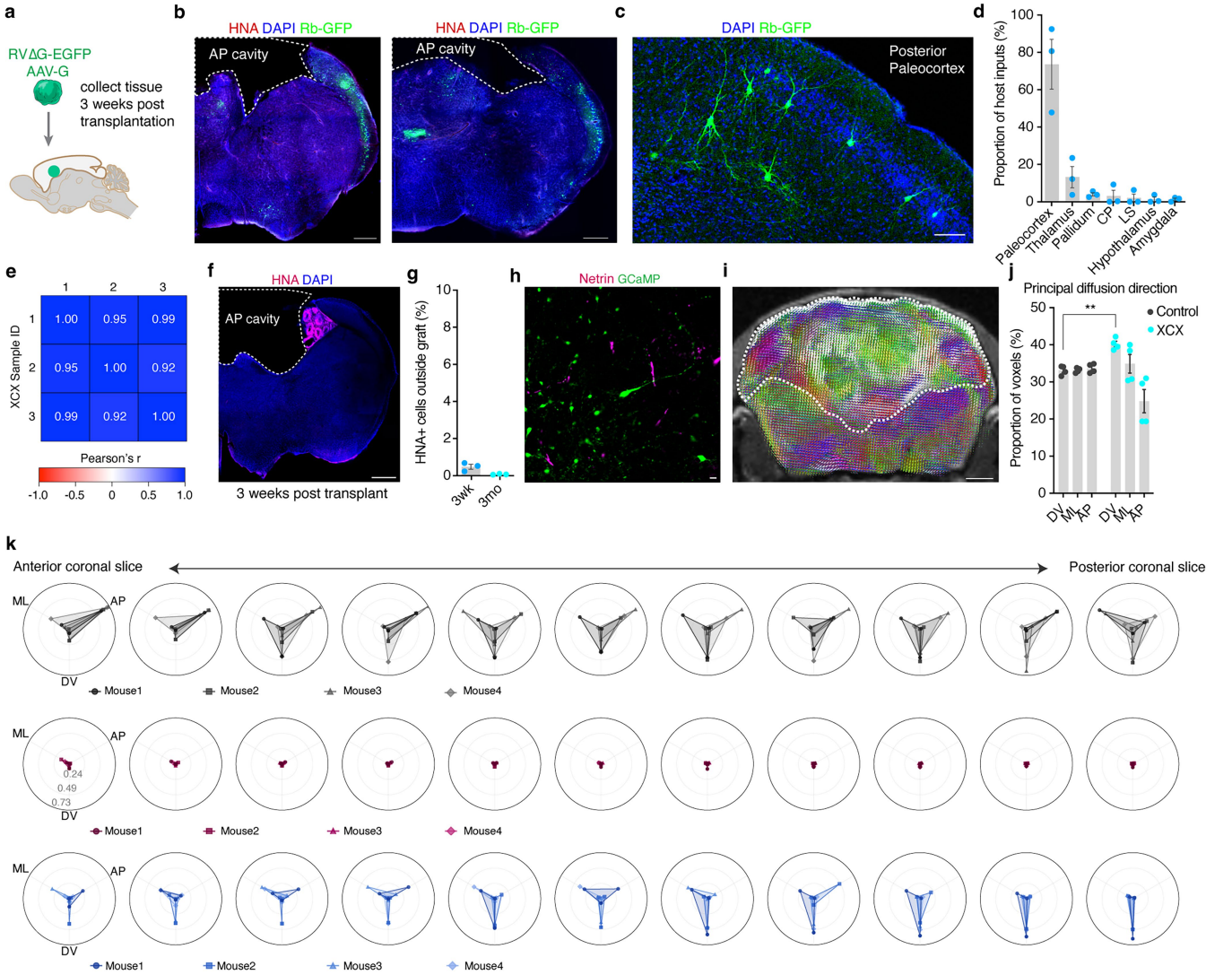
Reprints and permissions information is available at http://www.nature.com/reprints.



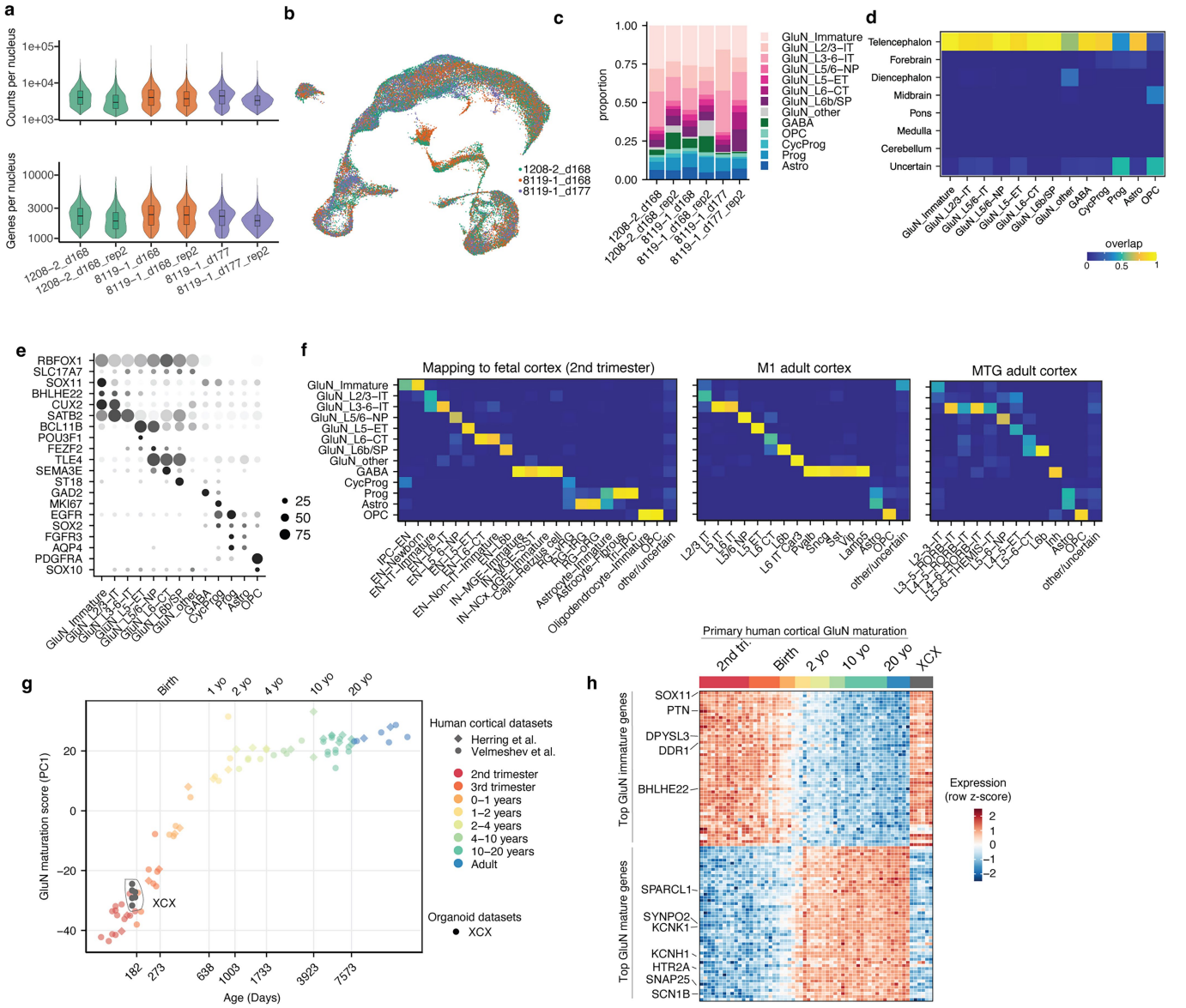
Extended Data Fig. 1 | Generation of the apallial mouse. a, Schematic illustrating the breeding strategy used to generate mice with selective deletion of *Esco2* within cells expressing the pallium marker *Emx1*. The *Prkdc*$^{scid/scid}$ background enabled transplantation without the need for exogenous immunosuppression. **b**, Quantification of the proportion of litter expressing an apallial phenotype, closely reflecting expected Mendelian phenotypic ratio (25% dashed line, $n = 91$ litters, 46 breeding cages). **c**, Body weight from P5-11 ($n = 18$ control, 28 apallial, unpaired t-test: t = 1.128, df=44, P = 0.2654). The average age is P8 for both groups. **d**, Kaplan-Meier survival curve of survival probability ($n = 11$ control/*Esco2*$^{fl/fl}$;*Prkdc*$^{scid/scid}$, 16 breeder/*Emx1-cre*$^{+/-}$;*Esco2*$^{fl/+}$;*Prkdc*$^{scid/scid}$, 12 apallial/*Emx1-cre*$^{+/-}$;*Esco2*$^{fl/fl}$;*Prkdc*$^{scid/scid}$ mice, Log-rank test, P = 0.2861). **e**, Representative immunostaining of a coronal section. Scale bar, 1 mm. **f**, Representative T2-weighted coronal MRI of 2-month-old control and apallial mice. Scale bar, 1 mm. **g**, snRNA-seq quality metrics after filtering showing the distribution of the number of counts and number of genes per nucleus across annotated populations. Each colour corresponds to a single mouse ($n = 4$ mice/880,149 single-nucleus profiles after QC). Embedded box plots: horizontal line denotes median; lower and upper hinges correspond to the first and third quartiles; whiskers extend 1.5 times the interquartile range. **h**, Number of cells per sample meeting quality control (QC) thresholds and filtered for poor mapping to Allen Brain Cell Atlas (post-QC). **i**,**j**, Bar plots showing annotated cell class (**i**) or subclass (**j**) proportions (%) for each sample (dots); mean ± SEM for ($n = 2$ control and $n = 2$ apallial mice). **k**, UMAP visualization of each sample. Bars plot and error bars represent mean ± SEM.



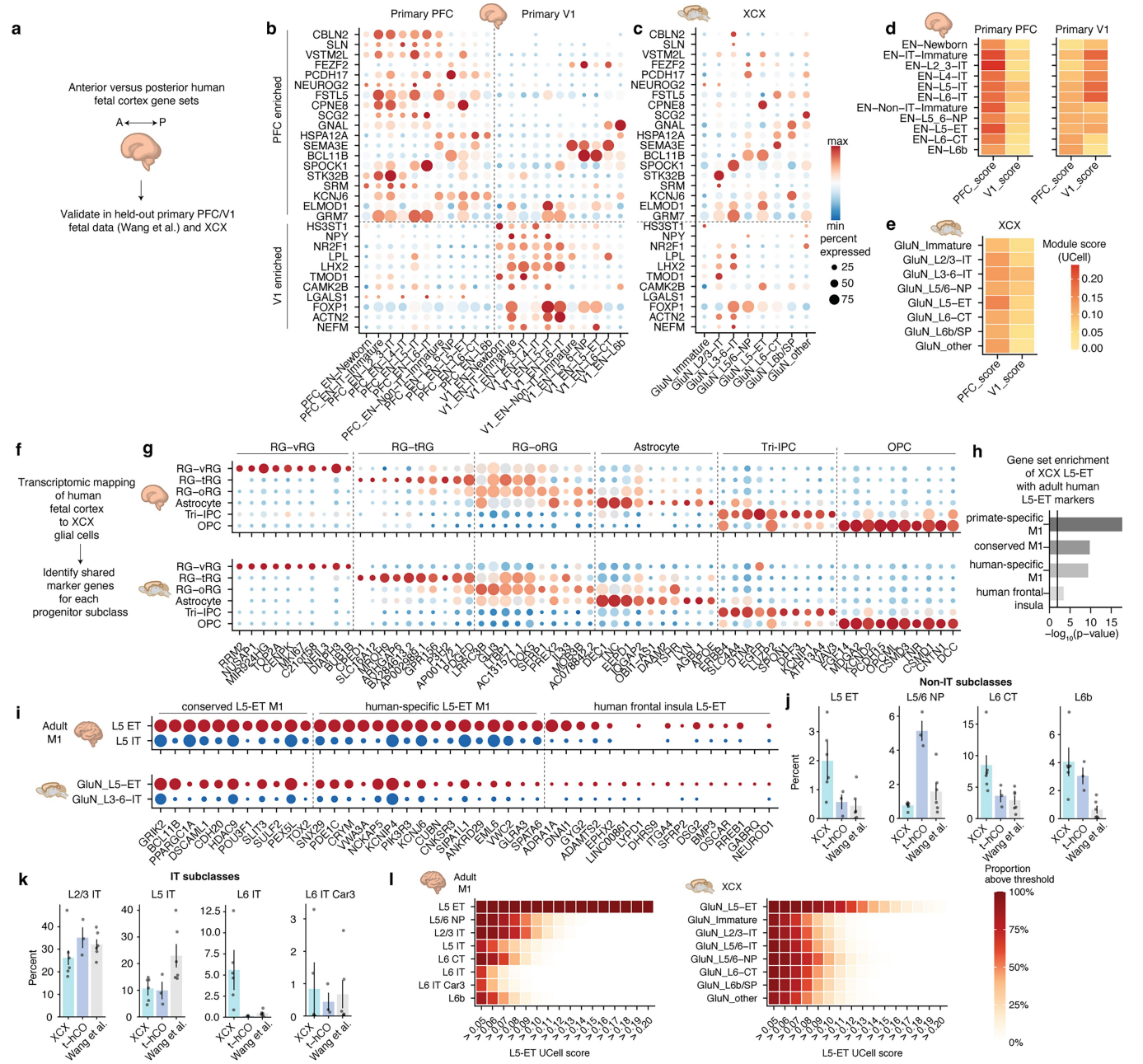
Extended Data Fig. 2 | snRNA-seq characterization of apallial mice. a, Scatter plots showing correlation of cell subclass annotations across conditions (Pearson correlation of log-transformed data shown). Dashed line denotes linear fit. **b**, Dot plot of log-transformed proportion changes between control and apallial annotated nuclei by label transform to Zeisel et al.[12] Black circles show annotations to cell classes (Taxonomy rank 4) and grey dots show annotated subclasses (cluster labels). Statistically significant (FDR < 0.1) subclass differences are coloured. **c**, Heatmap of Pearson correlation matrix calculated from cell subclass annotation log-transformed counts of each whole brain snRNA-seq sample. Dendrograms denote hierarchical clustering. **d**, Bar plot of log-transformed proportion changes between control and apallial CGE and MGE derived GABAergic neuron subclasses. **e**, Volcano plots showing differentially expressed genes (|log2 fold-change| > 1, FDR < 0.05; edgeR) for each cortical GABAergic subclass (pseudobulk) between control and apallial mice. Dashed lines denote significance thresholds. Non-significant genes (grey) were randomly downsampled for visualization.



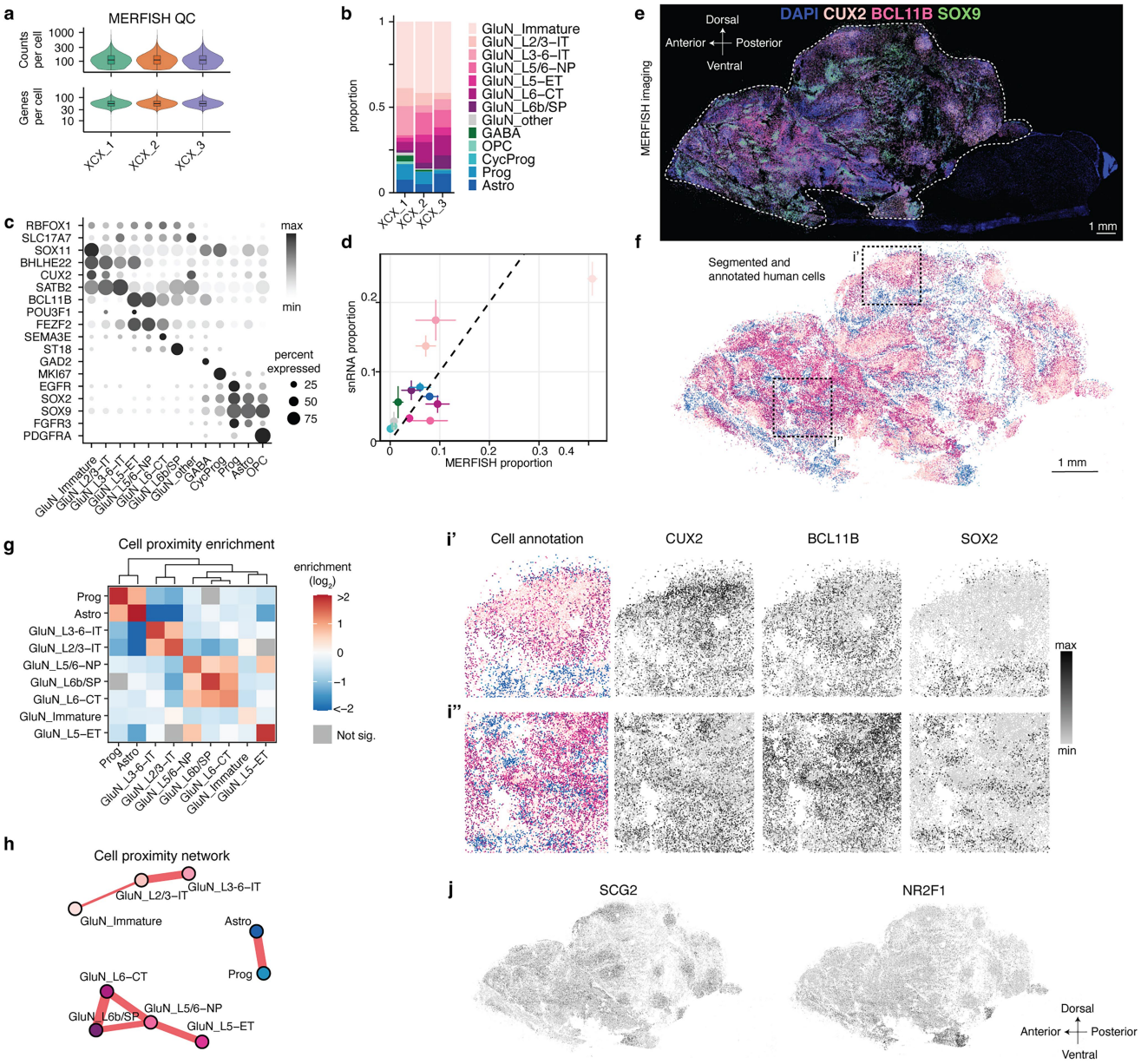
Extended Data Fig. 3 | Supplemental anatomical characterization using rabies tracing and diffusion-weighted imaging. a, Schematic depicting the rabies tracing approach. **b**, Representative coronal hemi-sections demonstrating Rb-GFP labelling of input cells in host paleocortex, thalamus (i.e. anterior thalamus on the left section and ventromedial thalamus on the right section), and pallidum (i.e. magnocellular and diagonal band nuclei in the section on the left). The series of sections used for analysis were also immunostained with HNA and only HNA-/GFP+ cells were included for analysis. Sections consecutive to the analysed series were immunostained for TH and PV in order to reveal caudate putamen and reticular thalamus (respectively) to aid identification of regions. Scale bar, 500 µm. **c**, High magnification image of Rb-GFP-labelled cells in the host paleocortex 1.6 mm from the posterior edge of the graft. Scale bar, 100 µm. **d**, Distribution of Rb-GFP-labelled cells across regions of the host brain ($n = 3$ mice). CP- caudate putamen, LS- lateral septum. **e**, Heat map of Pearson's correlation matrix comparing the regional distributions of host-derived labelled cells in three XCX mice. Correlations were calculated across the brain regions shown in (**d**). XCX sample 1 vs XCX sample 2 (P = 0.001), XCX sample 1 vs XCX sample 3 ($P = 5.797 \times 10^{-6}$), XCX sample 2 vs XCX sample 3 (P = 0.004). **f**, Coronal section 3 weeks post-transplant. Graft borders are defined by HNA immunostaining, and the dashed line indicates the apallial (AP) cavity. Scale bar, 500 µm. **g**, Quantification of HNA^+ cells outside the graft border at 3 weeks (0.486 % ± 0.132, $n = 3$ mice) and 3 months (0.080 % ± 0.014, $n = 3$ mice) post-transplantation; mean ± SEM. **h**, Netrin-G1+ projections within graft. Scale bar, 20 µm. **i**, Coronal T2w MRI in XCX -4.7 months post-differentiation with fibre orientation distributions (FODs) estimated from diffusion-weighted imaging (DWI) superimposed. Green FODs indicate a dorso-ventral (DV) preference for water diffusion, red: anterior-posterior (AP), and blue: medial-lateral (ML). Graft margin is indicated by white dashed line. Scale bar, 1 mm. **j**, The distribution of the dominant directions of water diffusion within control isocortex ($n = 4$ mice) and XCX ($n = 4$ mice). Because each animal's DV, ML, and AP proportions sum to 100%, these data are compositional and dependent. The direction x group interaction is therefore the contrast of interest: Two-way ANOVA with direction as a within-subject factor: interaction F (1.082, 6.494) = 7.399 (P = 0.0306); Holm-Sidak corrected posthoc tests for Control vs XCX: DV (P = 0.0014), ML (P = 0.5367), AP (P = 0.1262). As a complementary analysis, we tested whether each axis differs from the null expectation of no directional preference (i.e. 33.3% for each axis). Holm-Sidak corrected p-values: XCX DV: (P = 0.0173), remaining comparisons: P > 0.05. Together, these analyses indicate that the XCX group demonstrates a principal direction of water diffusion along the dorsal-ventral axis. **k**, Radar plots depicting the principal directions of the fixels across 11 consecutive slices. Rows represent the experimental groups: control (top), apallial (middle), and XCX (bottom). Columns represent the fixed, indexed slice position (Slices 1–11) across all subjects. Note that the comparison in Fig. 2h displays a specific, per-subject anatomical landmark (the anterior margin of the thalamus), which maps variably to Slices 2–6 depending on the individual. This supplementary figure is not aligned to anatomical landmarks, but rather the relative position in the scan sequence. To quantify alignment, the fixels were clustered into two main populations, aligned to a principal reference direction, weighted by peak amplitude, and normalized by total voxels. Each plot depicts $n = 4$ mice per condition, differentiated by the shape of the points, and shade of the respective colour. All experiments were performed using 8119-1 and 0524-1 cell lines. Bars plot and error bars represent mean ± SEM. **P < 0.01.



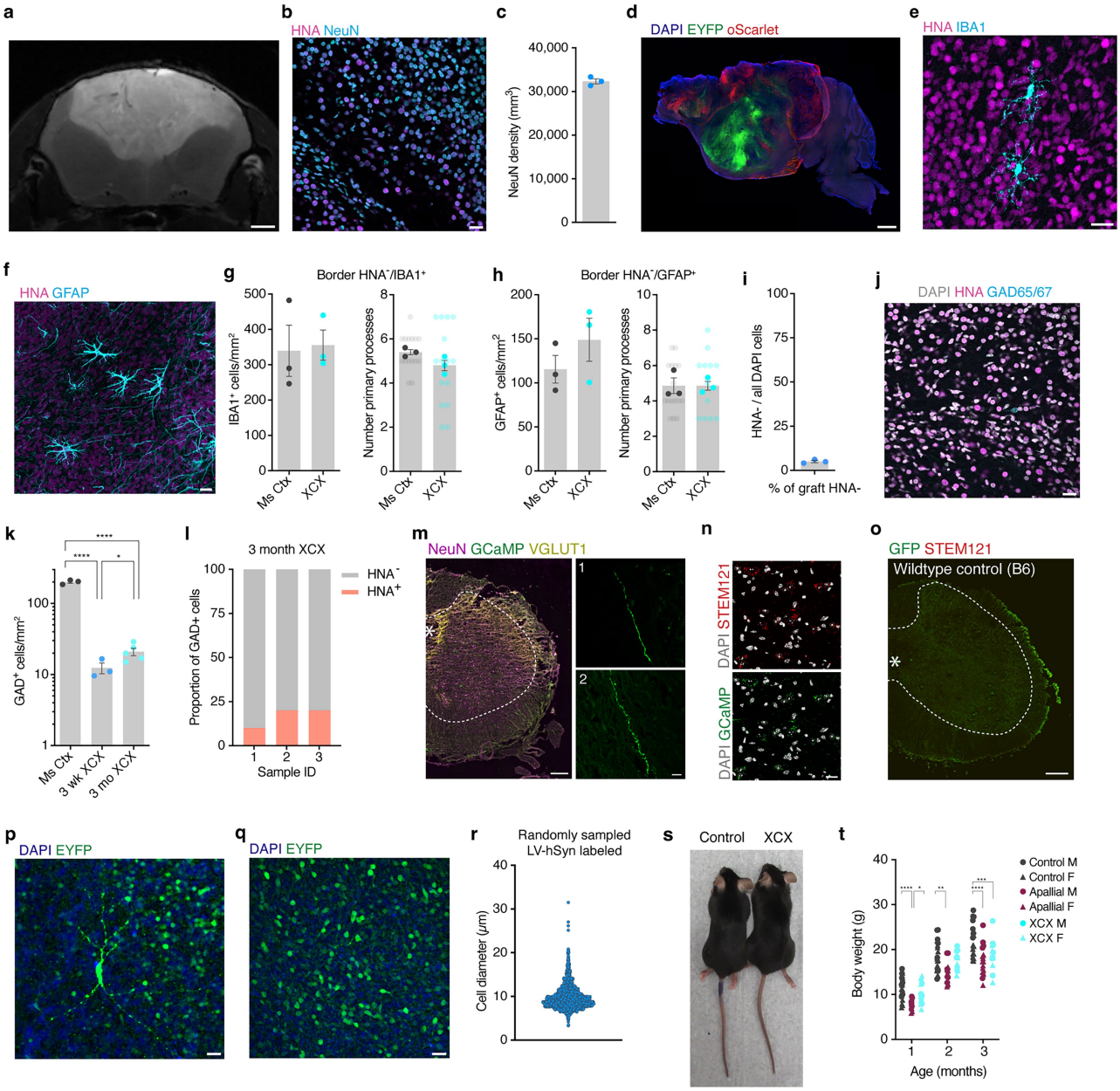
Extended Data Fig. 4 | XCX graft snRNA-seq data quality and characterization. a, snRNA-seq quality metrics after filtering showing the distribution of the number of human counts and number of human genes per nucleus in each sample. Each colour corresponds to a single mouse with technical replicates presented side-by-side ($n = 3$ mice/6 technical replicates/52,080 single-nucleus profiles). Box plots: horizontal line denotes median; lower and upper hinges correspond to the first and third quartiles; whiskers extend 1.5 times the interquartile range. **b**, Same integrated UMAP as shown in Fig. 2i, coloured by XCX animals. **c**, Cell type proportions across XCX graft 10x samples coloured by cluster annotation. **d**, Quantification of proportion of indicated XCX cell cluster mapping to primary human fetal brain regions by transfer labelling to reference human whole brain reference atlas[88]. **e**, Dot plot of marker gene expression in select clusters. **f**, Quantification of proportion of indicated primary human annotated cell cluster mapping to indicated XCX annotated cell cluster by transfer labelling to reference developing human cortex (*left*)[15], human adult M1 cortex (*middle*)[24] or human adult MTG cortex (*right*)[87] atlases. **g**, The first principal component (PC1) calculated on pseudobulk glutamatergic (GluN) neuron gene expression from human cortical development samples[16,17] and XCX samples using highly variable genes derived from Velmeshev et al. and plotted versus sample age (see Methods, the average age of XCX samples was 171 days post-differentiation, primary human samples are plotted with respect to gestational age). **h**, GluN neuron gene expression (pseudobulk and scaled for each gene) across snRNA-seq samples of the top 25 genes positively or negatively correlated with primary human sample age. All experiments were performed using 1208-2 and 8119-1 cell lines.



Extended Data Fig. 5 | XCX graft snRNA-seq data characterization. **a**, Cortical arealization analysis schematic. **b**,**c**, Dot plot of anterior and posterior cortical gene markers in glutamatergic neuron subclasses from (b) second trimester primary prefrontal cortex (PFC, *left*) and V1 (*right*)[15] and (c) XCX. **d**,**e**, Gene set module scores (calculated from UCell)[22] using genes from panels b-c applied to glutamatergic neuron subclasses from (d) primary PFC and V1 and (e) XCX. **f**, Progenitor cell analysis schematic. **g**, Dot plot of shared marker genes of indicated progenitor subclasses from (top) primary second trimester cortical data[15] and (*bottom*) XCX. Progenitor subclass annotations in the XCX graft obtained from label transfer to the primary fetal cortical atlas. **h**, Gene set enrichment analysis (one-sided Fisher's exact test with Bonferroni correction) of XCX L5-ET marker genes (primate-specific M1: $P = 2.34 \times 10^{-18}$, conserved M1: $P = 1.70 \times 10^{-10}$, human-specific M1: $P = 4.04 \times 10^{-10}$, human frontal insula: $P = 3.92 \times 10^{-4}$, fold change > 1) with adult human L5-ET gene sets obtained from M1[24] and frontoinsular cortex[20]. Vertical line indicates Bonferroni corrected p-value threshold (0.05/4). **i**, Dot plot of shared L5-ET genes in (*top*) adult human M1 and (*bottom*) XCX. **j**,**k**, Proportion of (j) non-intratelencephalic (IT) and (k) IT projecting glutamatergic neuron subclasses in each sample obtained from transfer labelling to adult M1 reference atlas for the following transplanted cortical organoid datasets: XCX (this study; $n = 3$ mice/6 technical replicates), t-hCO ($n = 3$ rats)[3], and cortical organoids transplanted into adult mice ($n = 6$)[23]. **l**, Heatmaps showing the proportion of cells within annotated glutamatergic neuron subclasses exceeding L5-ET UCell score thresholds in (left) adult human M1 and (right) XCX. All experiments were performed using 1208-2 and 8119-1 cell lines. Bars plot and error bars represent mean ± SEM.

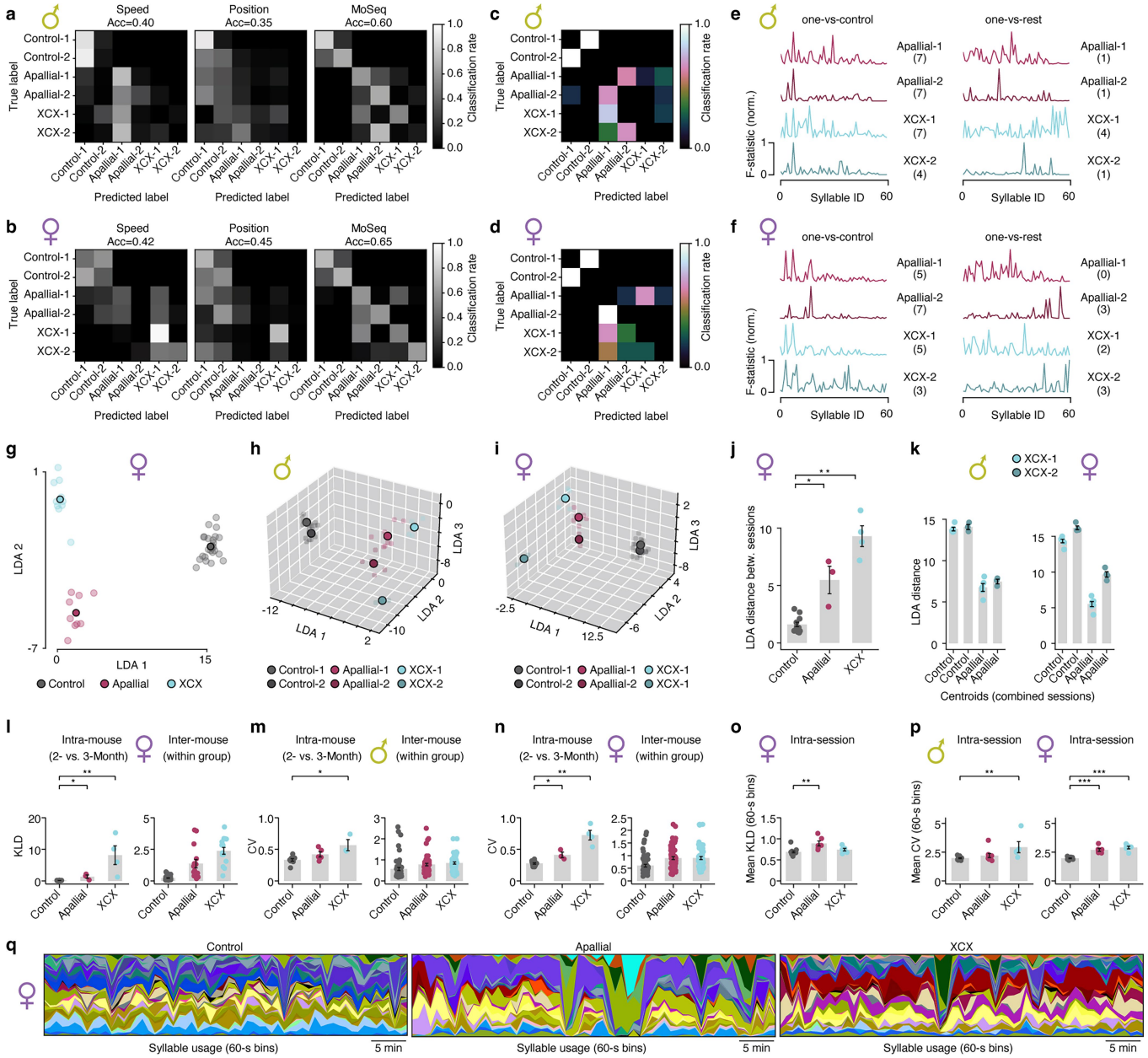


Extended Data Fig. 6 | XCX graft spatial transcriptomics characterization. **a**, MERFISH quality metrics after filtering showing the distribution of the number of human counts and number of human genes per cell in each sample ($n = 3$ mice, 296,101 cells after QC). Embedded box plots: horizontal line denotes median; lower and upper hinges correspond to the first and third quartiles; whiskers extend 1.5 times the interquartile range. **b**, Cell type proportions across XCX samples, coloured by cell type annotation derived from label transfer to XCX snRNA-seq. **c**, Dot plot of marker gene expression in select clusters. **d**, Scatter plot of MERFISH ($n = 3$ mice) versus snRNA-seq ($n = 3$ mice) cell type proportions. Dashed line denotes y = x; error bars denote SEM. **e**,**f**, Representative MERFISH of XCX mice at 4-months post-differentiation highlighting (e) transcript counts of indicated genes and (f) spatial organization of annotated cells, coloured as in b. **g**, Heatmap showing log2 enrichment of spatial proximity between cell type pairs, averaged across three independent samples, based on Delaunay triangulation networks and 1,000 label permutations (Giotto)[95]. Values shown only for interactions significant in ≥ 2 out of 3 samples (FDR < 0.05). **h**, Spatial interaction network where nodes represent cell types and edges indicate significantly enriched spatial proximity (FDR < 0.05 in ≥ 2 samples, log2 enrichment > 0). Layout: Fruchterman-Reingold algorithm. **i'-i''**, Insets from f showing spatial organization of annotated cells and normalized gene expression of indicated genes within segmented cells. **j**, Normalized expression of arealization genes across graft shown in e. All experiments were performed using 1208-2 and 8119-1 cell lines.



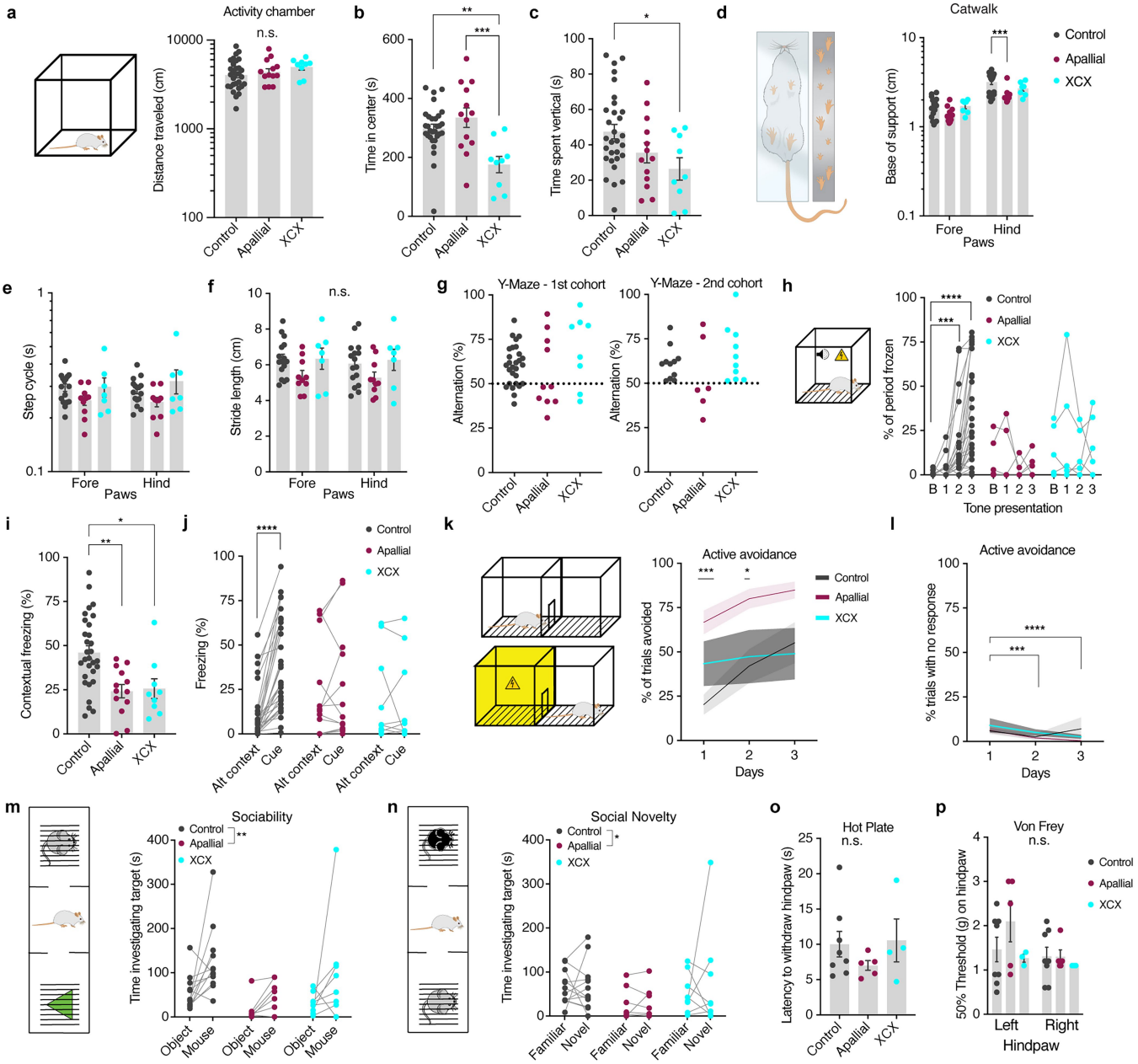
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Supplemental graft characterization.
a, Representative T2-weighted coronal MRI of XCX mice 3 months post-transplantation. Scalebar, 1 mm. **b**, Representative immunostaining of NeuN within graft, scale bar, 20 µm. **c**, Quantification of neuronal density within the graft (mean = 32,340, SEM = 575.1, *n* = 3 mice/48 fields of view). **d**, Sagittal brain section. Two organoids were transplanted per hemisphere, with one organoid expressing EYFP and the other expressing oScarlet. Scale bar, 1000 µm. **e**, Representative immunostaining of host-derived IBA1 + /HNA- microglia within the XCX graft. Scale bar, 20 µm. **f**, Representative immunostaining of $GFAP^+/HNA^+$ astrocytes located centrally in the graft, tiling the graft region. Scale bar, 20 µm (*right*). **g**,**h**, Comparison of the density and morphology of microglia (**g**) and astrocytes (**h**) at the graft-host border compared with control mouse cortex to assess for baseline inflammation. Quantification of the density (n = 3 animals per condition) and number of primary processes of host-derived $IBA1^+$ microglia and $GFAP^+$ astrocytes in control and XCX mice. **i**, The proportion of total host-derived cells within the organoid graft, as determined by HNA^- $DAPI^+$ nuclei divided by total $DAPI^+$ nuclei. **j**, Representative immunostaining of host-derived $HNA^-/GAD65/67^+$ neurons within graft ~3 months post-transplant. Scale bar, 20 µm. **k**, Quantification of GAD^+ cell density within the graft at 3 weeks (*n* = 3 XCX mice) and 3 months post-transplantation (*n* = 5 XCX mice), with control *Esco2*$^{fl/fl}$;*Prkdc*$^{scid/scid}$ mouse cortex as a control (*n* = 3 mice). ANOVA after log-transform F (2, 8) = 123.3 ($P = 9.74 \times 10^{-7}$); Holm-Sidak corrected posthoc tests: Mouse cortex vs XCX 3 weeks ($P = 1.53 \times 10^{-6}$), Mouse cortex vs XCX 3 months ($P = 2.12 \times 10^{-6}$), XCX 3 weeks vs XCX 3 months (P = 0.0165). **l**, Proportion of intragraft GAD^+ cells which are host- versus graft-derived at 3-months post-transplantation. **m**, Anterograde tracing from graft labels mouse cervical spinal cord with $GCaMP^+$ projections, as shown with immunostaining. The asterisk indicates the location of the central canal. Scale bar, 200 µm *(hemisection)*, 20 µm *(insets)*. **n**, $GCaMP^+/STEM121^+$ double-positive puncta are visible surrounding spinal cord nuclei. Scale bar, 10 µm. **o**, Immunostaining against STEM121 and GFP in a 3-month-old B6 wild-type mouse reveals only background signal. This mouse was within the age range of the XCX mice used for spinal cord imaging (90–110 days). This section was immunostained and imaged alongside the XCX mouse spinal cords, with identical acquisition parameters and digital brightness/contrast adjustment. Scale bar, 200 µm. **p**,**q**, Example images of graft-derived cell with large soma size (**p**), and graft-derived $EYFP^+$ cells without bipolar morphology (**q**). Scale bar, 20 µm. **r**, Distribution of soma diameter of cells labelled with hSYN1_EYFP or hSYN1_GCaMP8s (25th percentile: 8.058, median: 9.251, 75th percentile: 11.05, *n* = 1098 cells, 3 mice). **s**, Dorsal view of control and XCX adult mice. **t**, Body weight for control (*n* = 8 male, *n* = 14 female mice), apallial (*n* = 5 male, *n* = 12 female mice), and XCX (*n* = 7 male, *n* = 5 female mice) at 1, 2, and 3 months of age. Three-way ANOVA with age as a within-subject factor: group main effect F (2, 45) = 17.4953 ($P = 2.39 \times 10^{-6}$), sex main effect F (1,45) = 12.7332 (P = 0.0009), age main effect F (1.831, 82.41) = 310.1914 ($P = 7.47 \times 10^{-38}$), group x sex interaction F (2,45) = 1.1699 (P = 0.3197), group x time interaction F (3.662, 82.41) = 2.7583 (P = 0.0373), sex x time interaction F (1.8313, 82.41) = 11.8451 ($P = 5.1582 \times 10^{-5}$), group x sex x time interaction F (3.6627, 82.41) = 1.2498 (P = 0.2971). Sidak corrected posthoc tests - Month 1: Control vs Apallial ($P = 7.982 \times 10^{-6}$), Control vs XCX (P = 0.1014), Apallial vs XCX (P = 0.0302). Month 2: Control vs Apallial (P = 0.0019), Control vs XCX (P = 0.655), Apallial vs XCX (P = 0.0861). Month 3: Control vs Apallial ($P = 4.335 \times 10^{-5}$), Control vs XCX (P = 0.0005), Apallial vs XCX (P = 0.9655). All experiments were performed using 8119-1 and 0524-1 cell lines. Bars plot and error bars represent mean ± SEM. *P < 0.05; ****P < 0.0001.



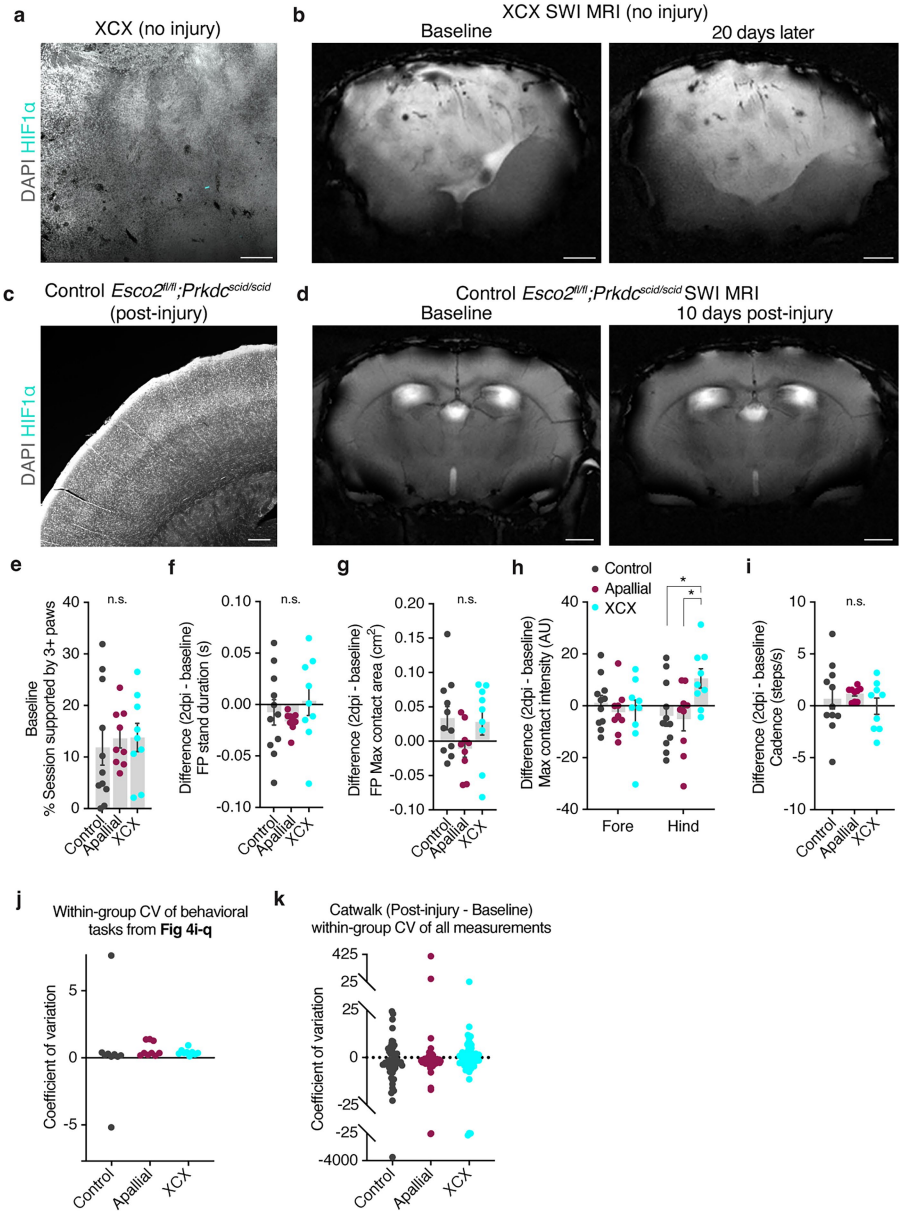
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Behavioural pattern of control, apallial and transplanted mice during spontaneous exploration. a, Normalized classification matrices (across rows and columns) showing the performance of a linear classifier in discriminating between group sessions for the male dataset. Classifiers were trained using one of the following behavioural measures: speed, position, or MoSeq syllables. An ideal classifier performance corresponds to a diagonal white with otherwise black fields (classification rate of 1). The overall accuracy of the classifier is indicated on the top. **b**, Same as in Extended Data Fig. 8a for the female dataset. **c**, Held-out confusion matrices (across rows and columns) summarizing the classification of a given group-session pair when excluded during training of a classifier for the male dataset. The displayed matrix combines the result of six separate classifiers. Since the classifier must allocate a recording from the held-out group-session pair to another label, the approach reveals (by design) the group-session pair with which the held-out is most confused with and thus, the relationship among group-session pairs. The diagonal of the matrix remains dark (unlike in a), since the correct group-session classification is impossible (see Methods for details of held-out classification). **d**, Same as in Extended Data Fig. 8c for the female dataset. **e**, Normalized F statistic ("behavioural fingerprints"; see Methods for details) of male mice highlighting the relevance of each indicated syllable for distinguishing a given group at a given age either from the corresponding age-matched control (one-vs-control) or all other conditions (one-vs-rest). The number of significant syllables is indicated in parentheses (Holm-Bonferroni-corrected $P < 0.01$ from the F-test). **f**, Same as in Extended Data Fig. 8e for the female dataset. **g**, LDA projection as in Fig. 4c. The spatial arrangement mirrors that in males. **h**, Three-dimensional LDA plot of group-session pairs in male mice. **i**, Same as in Extended Data Fig. 8h for the female dataset. **j**, Similarity of mean behavioural summaries of female mice within group-session pairs (similar to Fig. 4d). Euclidean distance in LDA space between the first and second session for each group. Dots represent mice with both usable sessions (Control n = 12, Apallial n = 3, XCX n = 4). The LDA session distance of both Apallial and XCX mice diverge significantly from Control (Kruskal–Wallis $H_2 = 13.26$, $P = 0.0013$). Bonferroni-corrected two-sided Mann-Whitney U tests showed Apallial > Control ($P = 0.013$) and XCX > Control ($P = 0.0033$), whereas XCX and Apallial mice do not differ ($P = 0.17$). **k**, LDA distance of each XCX mouse session to the pooled Control centroid and the pooled Apallial centroid. Reference centroids were calculated after first averaging repeated sessions within each reference mouse (male: Control n = 11, Apallial n = 9; female: Control n = 13, Apallial n = 6). Dots show paired distances for individual XCX mice (first session: males n = 5, females n = 6; second session: males n = 3, females n = 4). Across sex and session comparisons, XCX mouse recordings were descriptively closer to the Apallial centroid, but no comparison was significant after Bonferroni correction across embedding metrics and sessions (two-sided Wilcoxon signed-rank; first session: males $P = 0.25$, females $P = 0.125$; second session: males $P = 1.0$, females $P = 0.50$). **l**, KLD of syllable-usage distributions in female mice (similar to Fig. 4e). *Left*, intra-mouse: KLD between first (2-months of age) and second session (3-months of age) of the same animal. *Right*, inter-mouse: KLD for each unique mouse–mouse pair within a group (sessions first averaged per mouse). Dots represent single mice (intra; Control n = 12, Apallial n = 3, XCX n = 4) or mouse–pair values for visualization (inter; Control n = 13, Apallial n = 6, XCX n = 6 mice); inter-mouse inference was performed by permuting whole-mouse group labels. Intra-mouse Kruskal–Wallis tests indicated significant group effects (intra, $H_2 = 12.9$, $P = 1.58 \times 10^{-3}$). Bonferroni-corrected two-sided Mann-Whitney U tests showed Apallial > Control ($P = 0.013$, Bonferroni) and XCX > Control ($P = 0.0033$); Apallial and XCX mice were not different ($P = 0.69$, Bonferroni). The inter-mouse permutation test indicated a group effect ($P = 0.010$), but no Bonferroni-corrected pairwise comparison was significant (Control versus Apallial $P = 0.413$; Control versus XCX $P = 0.173$; Apallial versus XCX $P = 0.714$). **m**, Coefficient of variation (CV) of syllable frequencies in male mice. Left, intra-mouse: CV of each syllable between first (2-months of age) and second session (3-months of age) of the same animal; syllable-level CVs were averaged within mouse before inference. Dots represent mouse–syllable values for visualization; n denotes mice (Control n = 9, Apallial n = 5, XCX n = 3). Bars in the left panel show the mean ± SEM across mouse-level means. A Kruskal-Wallis test indicated a significant group effect ($H_2 = 7.74$, $P = 0.0209$). Bonferroni-corrected two-sided Mann-Whitney U tests showed XCX > Control ($P = 0.027$); Apallial versus Control ($P = 0.249$) and Apallial versus XCX ($P = 1.0$) were not significant. Right, inter-mouse: CV across mice within a group. Dots represent syllables; significance was assessed by permuting whole-mouse group labels and recomputing the CV. The permutation test did not indicate a significant group effect ($P = 0.545$; Control n = 11, Apallial n = 9, XCX n = 5 mice). **n**, CV of syllable frequencies in female mice. Panels mirror Extended Data Fig. 8m. Left, intra-mouse: dots represent mouse–syllable values for visualization; n denotes mice (Control n = 12, Apallial n = 3, XCX n = 4). A Kruskal–Wallis test showed a significant group effect ($H_2 = 13.26$, $P = 0.0013$). Bonferroni-corrected two-sided Mann–Whitney U tests showed Apallial > Control ($P = 0.013$) and XCX > Control ($P = 0.0033$); Apallial and XCX were not different ($P = 0.17$). Right, inter-mouse: dots represent syllables; significance was assessed by permuting whole-mouse group labels and recomputing the CV. The permutation test did not indicate a significant group effect ($P = 0.108$; Control n = 13, Apallial n = 6, XCX n = 6 mice). **o**, KLD between syllable usages of the entire session and 60-s windows with 30-s overlap in female mice (averaged per session); as in Fig. 4f. The window-averaged index was first calculated per session and then averaged across repeated sessions within mouse for inference. Dots represent sessions for visualization; n denotes mice (Control n = 13, Apallial n = 6, XCX n = 6). Grey bars show the mean ± SEM of the mouse-level means. Kruskal–Wallis tests revealed significant group effects intra-session ($H_2 = 10.85$, $P = 0.0044$). Post-hoc two-sided Mann–Whitney U tests were Bonferroni-corrected: Apallial > Control ($P = 0.0027$); Control versus XCX ($P = 0.318$) and Apallial versus XCX ($P = 0.279$) were not significant. **p**, CV of syllable usage across 60-s windows with 30-s overlap in mice (averaged per session). The syllable-averaged CV was first calculated per session and then averaged across repeated sessions within mouse for inference. Dots represent sessions for visualization; n denotes mice, and bars show the mean ± SEM of the mouse-level means. *Left*, male mice (Control n = 11, Apallial n = 9, XCX n = 5): Kruskal–Wallis tests showed a significant intra-session group effect ($H_2 = 8.75$, $P = 0.0126$). Bonferroni-corrected pair-wise Mann–Whitney U tests indicated XCX differs from Control ($P = 0.0055$); no other comparisons reached significance (Apallial versus Control, $P = 0.590$; Apallial versus XCX, $P = 0.249$). *Right*, female mice (Control n = 13, Apallial n = 6, XCX n = 6): Kruskal–Wallis tests revealed a robust intra-session group effect ($H_2 = 18.30$, $P = 1.06 \times 10^{-4}$). Bonferroni-corrected post-hoc Mann–Whitney U tests showed Apallial > Control ($P = 2.21 \times 10^{-4}$) and XCX > Control ($P = 2.21 \times 10^{-4}$); Apallial and XCX did not differ ($P = 0.93$). **q**, Stack plots as in Fig. 4g. For j–p, grey bars and error bars show mean ± SEM. Means are calculated across mice in j and k; across mice (left) or displayed mouse-pair values (right) in l; across mouse-level means (left) or syllable-level CV values (right) in m and n; and across mouse-level means in o and p. For j and l-p, asterisks above brackets denote Bonferroni-corrected two-sided pairwise P values: *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$; ns, $P \geq 0.05$. Brackets identify pairwise comparisons and do not denote omnibus tests. All experiments were performed using 8119-1 and 0524-1 cell lines.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Directed behavioural investigation of transplantation in apallial mice. a-c, Activity chamber for Control ($n = 30$), Apallial ($n = 13$), and XCX ($n = 9$) mice. **a**, *Left-* Schematic of the activity chamber. *Right-* distance travelled in the activity chamber. ANOVA after log transform $F(2, 49) = 1.330$ ($P = 0.2739$). **b**, Time spent in the centre of the activity chamber. ANOVA $F(2, 49) = 8.634$ ($P = 0.0006$); Holm-Sidak corrected posthoc tests: Control vs Apallial ($P = 0.2335$), Control vs XCX ($P = 0.0019$), XCX vs Apallial ($P = 0.0006$). **c**, Time spent vertically (i.e. rearing) in the activity chamber. ANOVA $F(2, 49) = 3.905$ ($P = 0.0267$); Holm-Sidak corrected posthoc tests: Control vs Apallial ($P = 0.1858$), Control vs XCX ($P = 0.0370$), XCX vs Apallial ($P = 0.3318$). **d**–**f**, CatWalk for Control ($n = 16$), Apallial ($n = 10$), and XCX ($n = 7$) mice. Also see table S5. **d**, *Left-* Schematic of gait analysis in the CatWalk task, *Right-* The average distance between forepaws (i.e. the distance between front right and front left paws) and hind paws. Two-way ANOVA with paw as a within-subject factor: group main effect $F(2, 30) = 4.278$ ($P = 0.0232$), paw main effect $F(1, 30) = 221.4$ ($P < 0.0001$), interaction $F(2, 30) = 8.864$ ($P = 0.0009$); Holm-Sidak corrected posthoc tests: Hind paw Control vs Apallial ($P = 0.0006$), all others $P > 0.05$. **e**, The average duration of the step cycle (stand time + swing time). Two-way ANOVA with paw as a within-subject factor after log-transform: group main effect $F(2, 30) = 2.064$ ($P = 0.1445$), paw main effect $F(1, 30) = 0.03704$ ($P = 0.8487$), interaction $F(2, 30) = 4.765$ ($P = 0.0160$); Holm-Sidak corrected posthoc tests: all $P > 0.05$. **f**, The average stride length. Two-way ANOVA with paw as a within-subject factor: group main effect $F(2, 30) = 2.071$ ($P = 0.1437$), paw main effect $F(1, 30) = 4.071$ ($P = 0.0526$), interaction $F(2, 30) = 0.8018$ ($P = 0.4579$). **g**, 1st (*left*) and 2nd (*right*) cohort of Y-Maze behaviour demonstrated consistent results. Internal replication of Y-maze behavioural results conducted months apart using new hCO differentiations and new mouse breeders supports the reliability of behavioural readouts in XCX mice. *First cohort-* control ($n = 28$), apallial ($n = 10$), and XCX ($n = 8$) mice. *Second cohort-* control ($n = 11$), apallial ($n = 6$), and XCX ($n = 10$) mice. Together, the data from these two cohorts comprise Fig. 4q. **h**, *Left-* Schematic of the aversive Pavlovian trace conditioning assay (Control $n = 30$, Apallial $n = 13$, XCX $n = 9$). *Right-* Freezing in response to tones during training. X-axis titles: "B"=freezing during baseline; (1,2,3)=freezing during tone in order of tone presentation. Two-way ANOVA with toneNumber as a within-subject factor after log transform: group main effect $F(2, 49) = 3.084$ ($P = 0.0548$), toneNumber main effect $F(2.333, 114.3) = 10.19$ ($P < 0.0001$), interaction $F(4.666, 114.3) = 14.84$ ($P < 0.0001$); Holm-Sidak corrected posthoc tests comparing freezing during tone presentation within each group, only comparisons to baseline shown: Control group: Baseline vs Tone2: ($P = 0.0003$), Baseline vs Tone3: ($P < 0.0001$), Tone1 vs Tone2: ($P = 0.0006$), Tone1 vs Tone3: ($P < 0.0001$), Tone2 vs Tone3: ($P < 0.0001$). All other comparisons $P > 0.05$. **i**, Freezing in the previously conditioned context without foot shock 1 day after conditioning. ANOVA $F(2, 49) = 8.526$ ($P = 0.0007$); Holm-Sidak corrected posthoc tests Control vs Apallial ($P = 0.0023$), Control vs XCX ($P = 0.0104$), XCX vs Apallial ($P = 0.8581$) **j**, The effect of presentation of the cue previously associated with foot shock on freezing in an altered context without foot shock. Two-way ANOVA with tone as a within-subject factor: group main effect $F(2, 49) = 0.1425$ ($P = 0.8676$), tone main effect $F(1, 49) = 11.21$ ($P = 0.0016$), interaction $F(2, 49) = 13.20$ ($P < 0.0001$); Holm-Sidak corrected posthoc tests comparing freezing within-subject: Control ($P < 0.0001$), all others $P > 0.05$. **k**, *Left-* Schematic of the active avoidance procedure (Control $n = 13$, Apallial $n = 9$, XCX $n = 6$). *Right-* Learning curve for active avoidance over days. Two-way ANOVA with day as a within-subject factor: group main effect $F(2, 25) = 5.510$ ($P = 0.0104$), day main effect $F(1.465, 36.63) = 11.79$ ($P = 0.0004$), interaction $F(2.930, 36.63) = 2.283$ ($P = 0.0965$); Holm-Sidak corrected posthoc tests indicated with asterisks on the figure: Day 1: Control vs Apallial ($P = 0.0005$), Day 2: Control vs Apallial ($P = 0.0235$), all others $P > 0.05$. **l**, Percent of trials with no response over days in active avoidance. Two-way ANOVA with day as a within-subject factor after log transform: group main effect $F(2, 25) = 0.9370$ ($P = 0.4051$), day main effect $F(1.896, 47.39) = 15.59$ ($P < 0.0001$), interaction $F(3.791, 47.39) = 0.8710$ ($P = 0.4836$); same group sizes indicated in (k). Holm-Sidak corrected posthoc tests Day 1 vs Day 2: ($P = 0.0003$), Day 1 vs Day 3: ($P < 0.0001$), Day 2 vs Day 3: ($P = 0.1298$). **m**, *Left-* Schematic of the three-chamber sociability assay (Control $n = 12$, Apallial $n = 7$, XCX $n = 8$). *Right-* Preference for a novel mouse over a novel object. Two-way ANOVA after log-transform with target as a within-subject factor: group main effect $F(2, 24) = 8.112$ ($P = 0.0020$), target main effect $F(1, 24) = 8.817$ ($P = 0.0067$), interaction $F(2, 24) = 0.2044$ ($P = 0.8165$); Holm-Sidak corrected posthoc tests: overall investigation time Control vs Apallial: ($P = 0.0017$), all others $P > 0.05$. Log transform applied due to heteroscedasticity demonstrated in the residual plot. **n**, Preference for a novel over a familiar mouse. Two-way ANOVA after log-transform with target as a within-subject factor: group main effect $F(2, 24) = 3.585$ ($P = 0.0434$), target main effect $F(1, 24) = 0.005222$ ($P = 0.9430$), interaction $F(2, 24) = 0.2917$ ($P = 0.7496$); Holm-Sidak corrected posthoc tests: overall investigation time Control vs Apallial: ($P = 0.0390$), all others $P > 0.05$; same group sizes indicated in (m). Log transform applied due to heteroscedasticity demonstrated in the residual plot. **o**, Latency to withdraw hind paw in the hotplate thermal sensation assay (Control $n = 8$, Apallial $n = 5$, XCX $n = 4$). ANOVA $F(2, 14) = 0.8389$ ($P = 0.4528$). **p**, 50% threshold for paw withdrawal in the von Frey mechanical sensation assay (Control $n = 8$, Apallial $n = 5$, XCX $n = 3$). Two-way ANOVA with paw as a within-subject factor after log transform: group main effect $F(2, 13) = 0.7641$ ($P = 0.4856$), paw main effect $F(1, 13) = 1.283$ ($P = 0.2778$), interaction $F(2, 13) = 0.4368$ ($P = 0.6552$). All experiments were performed using 8119-1 and 0524-1 cell lines. Bars plot and error bars represent mean ± SEM. *$P < 0.05$;**$P < 0.01$; ***$P < 0.001$;****$P < 0.0001$.



Extended Data Fig. 10 | Extended characterization of control conditions for low oxygen exposure and post-injury behavioural changes. a, Representative immunostaining for HIF1α in a control XCX graft that did not undergo hypoxic injury. The field of view is completely within the graft interior, as indicated by the organoid-like cytoarchitecture. Scale bar, 250 µm. **b**, Susceptibility-Weighted Imaging sequence of the same graft 20 days apart, suggestive of stable hypointense vasculature in the absence of injury. Scale bar, 1 mm. **c**, Representative immunostaining for HIF1α in a control mouse (*Esco2*$^{fl/fl}$;*Prkdc*$^{scid/scid}$ mouse without a transplant) collected immediately following the hypoxic injury protocol. Scale bar, 250 µm. **d**, Susceptibility-Weighted Imaging sequence of the same control mouse before injury (*left*) and 10 days post-injury (10dpi, *right*). Scale bar, 1 mm. **e**–**i**, Additional data to support detection of within-animal behavioural readouts of hypoxic injury using the CatWalk task across control ($n = 11$), apallial ($n = 9$), and XCX ($n = 9$) mice. **e**, Percent of session supported by 3+ paws during the baseline test prior to injury. ANOVA F (2, 26) = 0.1454 (P = 0.8654). **f**, Within-animal difference between performance on 2 days post-injury (2dpi) and baseline for forepaw (FP) stand duration. Welch's ANOVA W(2.000, 13.03) = 1.061 (P = 0.3741). **g**, Within-animal difference between performance on 2 days post-injury (2dpi) and baseline for forepaw (FP) max contact area. ANOVA F (2, 26) = 2.085 (P = 0.1446). **h**, Within-animal difference between performance on 2 days post-injury (2dpi) and baseline for the maximum intensity of the paw prints during maximum contact. Two-way ANOVA with paw as a within-subject factor: group main effect F (2, 26) = 1.874 (P = 0.1737), paw main effect F (1, 26) = 0.3405 (P = 0.5646), interaction F (2, 26) = 3.839 (P = 0.0346); Holm-Sidak corrected posthoc tests for Hindpaw: Control vs XCX (P = 0.0433), XCX vs Apallial (P = 0.0373), all others P > 0.05. **i**, Within-animal difference between performance on 2 days post-injury (2dpi) and baseline for step cadence. Welch's ANOVA W(2.000, 13.27) = 1.131 (P = 0.3519). **j**, Coefficient of variation (CV) across specific behavioural tasks in Fig. 4i–q. **k**, Coefficient of variation (CV) across all within-animal differences post injury. All experiments were performed using 8119-1 and 0524-1 cell lines. Bars plot and error bars represent mean ± SEM. All statistical tests were two-tailed. *P < 0.05.

nature portfolio

Corresponding author(s): Sergiu Pașca

Last updated by author(s): Jul 29, 2026

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	BZ-X analyzer (Keyence) and Las-X 3.5.7 (Leica) software were used for acquiring immunohistochemistry images. PyDAQmx v1.4.5 and pypylon 1.7.2 were used for OEG acquisition. The open ephys GUI (v0.4.4.1) and Motion Sequencing acquisition software (kinect2-nidaq v0.2.3) were used to collect electrophysiological and behavioral data, respectively. All MRI was performed in vivo on a 7 T Bruker scanner (Stanford University) running ParaVision 360.3.5.
Data analysis	The following custom- and publicly-available software packages were used for data analysis; implementation details are provided in the Methods section: ImageJ 1.53q, QuPath v0.5.1, CellPose 3.0.8.dev19+g0ce3653, Python 3.7 and 3.10, Motion Sequencing (MoSeq2-app 1.3.1, moseq2-extract 1.2.0, moseq2-model 1.2.0, moseq2-pca 1.2.0, moseq2-stat 0.6.0, moseq2-viz 1.3.0), NumPy 1.20 / 1.26.4, SciPy 1.7.0 / 1.15.1, scikit-learn 1.0.2 / 1.6.1, StatsModels 0.13.5, Pandas 1.3.5 / 2.2.3, Seaborn 0.12 / 0.13.2, Matplotlib 3.1.2 / 3.5.3 / 3.10.0, scikit-image 0.16.2, H5py 2.10 / 3.12.1, Dask 2.30, Bokeh 2.4.2 / 3.6.3, Plotly 4.14.3, Neo 0.14.0, and Numba 0.61.2. OEG analysis also used: Tifffile 2024.9.20, CaImAn v1.12.1, CMasher 1.9.2., MRtrix3 V.3.0.3, ANTs v.2.3.5

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Single-nucleus gene expression data are available under the Gene Expression Omnibus accession number GSE301194. Spatial transcriptomics data have been deposited at Zenodo (https://doi.org/10.5281/zenodo.21138437) and are available as of the date of the publication. The following public datasets were used for XCX graft snRNA-seq analysis: human genome sequence information from Ensembl (http://ftp.ensembl.org/pub/release-98/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz) and human gene annotation from GENCODE (http://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_32/gencode.v32.primary_assembly.annotation.gtf.gz); information on the mouse genome sequence is from Ensembl (http://ftp.ensembl.org/pub/release-98/fasta/mus_musculus/dna/Mus_musculus.GRCm38.dna.primary_assembly.fa.gz); the Allen Brain Institute generated human adult snRNA-seq data from the primary motor cortex (https://portal.brain-map.org/atlases-and-data/rnaseq; accessed May 2022); the Human Neural Organoid Cell Atlas (https://datasets.cellxgene.cziscience.com/83b34db4-0319-44ee-b81d-815e3005be30.h5ad; accessed March 2026); Wang et al. transplanted organoid snRNA (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE185472; accessed February 2026). The following public datasets were used for mouse whole-brain snRNA-seq analysis: mouse genome sequence information from Ensembl (https://ftp.ensembl.org/pub/release-110/fasta/mus_musculus/dna/Mus_musculus.GRCm39.dna.primary_assembly.fa.gz); mouse brain scRNA-seq atlas from Zeisel et al. 2018 (https://storage.googleapis.com/linnarsson-lab-loom/l5_all.loom; accessed March 2022). DWI data are available https://osf.io/5sdx6/overview, and related code https://github.com/garikoitz/apallial_dmri. All other data are available as source datasheets or on zenodo.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Sample sizes were estimated empirically, based on previous studies (Birey et. al., Nature 2017; Marton et., Nature Neuroscience 2019; Pasca et al., Nature Medicine 2019, Khan et al., Nature Medicine 2020, Myura et al., Nature Biotechnology 2020, Revah et al., Nature 2022, Gschwind et al., Neuron 2023).

Data exclusions	Mice were excluded if they presented morbidity and did not complete the core battery of behavioral tasks.
Replication	Micrographs depict representative results and were repeated with the following n: Fig. 2n, o (3 mice); Fig. 2p (3 mice); Fig. 3b (3 mice); Fig. 5b (3 mice); Extended Data Fig. 1e (1 mouse/condition); Extended Data Fig. 3b (3 mice); Extended Data Fig. 3c (1 mouse); Extended Data Fig. 3f (3 mice); Extended Data Fig. 7d (3 sagittal sections/1 mouse, presence of subcortical fiber tracts replicated in coronal sections from 3 mice); Extended Data Fig. 7e,f (3 mice); Extended Data Fig. 7j (5 mice); Extended Data Fig. 7m,n (3 mice); Extended Data Fig. 7o (1 mouse); Extended Data Fig. 7p,q (3 mice); Extended Data Fig. 10a (4 mice); Extended Data Fig. 10b (2 mice); Extended Data Fig. 10c (4 mice); Extended Data Fig. 10d (2 mice).
Randomization	Randomization was used in the stereological analysis of soma size. Images were acquired from an equally distributed grid throughout the tissue for downstream analysis. Randomization of subjects was performed to ensure both cell-lines were represented. Most analyses used the entire population, so randomization was not required.
Blinding	Data were collected and analyzed by investigators blinded to the group identity or automatically analyzed using Motion Sequencing.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used

Primary antibodies:
1. ChAT (AB144P, Millipore)
2. CTIP2 (ab123449, abcam)
3. GAD65/67 (AB1511, Chemicon)
4. GFAP (Z0334, DAKO)
5. GFP (A21311, Invitrogen)
6. HIF-1α (NB100-479SS, Novus Biologicals)
7. HNA (ab191181, abcam)
8. HNA (235-1R, NeoBiotechnologies)
9. IBA1 (ab5076, Abcam)
10. NeuN (ab104225, Abcam)
11. NeuN (mab377, Millipore Sigma)
12. Netrin-G1 (sc-271774, Santa Cruz)
13. SATB2 (ab207040, abcam)
14. STEM121 (Y40410, Takara)
15. vGlut1 (guinea pig; from Jessel lab CU1328)

Secondary antibodies:
1. Donkey anti-mouse 405 (715-475-150; Jackson ImmunoResearch)
2. Donkey anti-rabbit 488 (A-21206; Invitrogen)
3. Donkey anti-rabbit 568 (A10042; Invitrogen)
4. Donkey anti-mouse 647 (A32787; Invitrogen)
5. Donkey anti-goat Plus 647 (A32849; Invitrogen)
6. Donkey anti-rabbit Plus 647 (A32795; Invitrogen)
7. Donkey anti-goat Cy3 (705-165-003 or 705-165-147; Jackson ImmunoResearch)
8. Donkey anti-guinea pig 405 (706-475-148; Jackson ImmunoResearch)
9. Donkey anti-rabbit 405 (711-475-152; Jackson ImmunoResearch)
10. Donkey anti-chicken 488 (703-545-155; Jackson ImmunoResearch)

Validation

All antibodies have been validated and previously used for immunofluorescence in mouse or human. The strategies for validation and the number of studies using this antibody are from the company's website. Specific citations refer to relevant papers using the antibody.

Primary Antibodies
1. ChAT (AB144P; Millipore); https://www.sigmaaldrich.com/US/en/product/mm/ab144p

2. CTIP2 (ab123449; Abcam); https://www.abcam.com/en-us/products/primary-antibodies/ctip2-antibody-25b6-ab123449
3. GAD65/67 (AB1511; Chemicon); https://www.sigmaaldrich.com/US/en/product/mm/ab1511
4. GFAP (Z0334; DAKO); https://www.agilent.com/store/en_US/Prod-Z033429-2/Z0334
5. GFP (A21311; Invitrogen); https://www.thermofisher.com/antibody/product/GFP-Antibody-Polyclonal/A-21311
6. HIF-1α (NB100-479SS; Novus Biologicals); https://www.novusbio.com/products/hif-1-alpha-antibody_nb100-479ss
7. HNA (ab191181; Abcam); https://www.abcam.com/en-us/products/primary-antibodies/human-nuclear-antigen-antibody-235-1-ab191181
8. HNA (235-1R; NeoBiotechnologies); https://neobiotechnologies.com/product/human-nuclei-235-1r/
9. IBA1 (ab5076; Abcam); https://www.abcam.com/en-us/products/primary-antibodies/iba1-antibody-ab5076
10. NeuN (ab104225; Abcam); https://www.abcam.com/en-us/products/primary-antibodies/neun-antibody-1b7-ab104225
11. NeuN (mab377; Millipore Sigma); https://www.sigmaaldrich.com/US/en/product/mm/mab377
12. Netrin-G1 (sc-271774; Santa Cruz); https://www.scbt.com/p/netrin-g1-antibody-a-8
13. SATB2 (ab207040; Abcam); https://www.abcam.com/en-us/products/primary-antibodies/satb2-antibody-epr20038-ab207040
14. STEM121 (Y40410; Takara); https://www.takarabio.com/products/antibodies-and-elisa/primary-antibodies-and-elisas-by-research-area/stem-cell-research-antibodies/stem-antibodies?srsltid=AfmBOopq0EG5ZBSLBKPh2kt_voLpWIWK4qvx09PHCywhmVRlmu9RMBbN
15. vGlut1 (guinea pig; from Jessel lab CU1328)

Secondary Antibodies
1. Donkey anti-mouse 405 (715-475-150; Jackson ImmunoResearch); https://www.jacksonimmuno.com/catalog/products/715-475-150
2. Donkey anti-rabbit 488 (A-21206; Invitrogen); https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21206
3. Donkey anti-rabbit 568 (A10042; Invitrogen); https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-10042
4. Donkey anti-mouse 647 (A32787; Invitrogen); https://www.thermofisher.com/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A32787
5. Donkey anti-goat Plus 647 (A32849; Invitrogen); https://www.thermofisher.com/antibody/product/Donkey-anti-Goat-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A32849
6. Donkey anti-rabbit Plus 647 (A32795; Invitrogen); https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A32795
7. Donkey anti-goat Cy3 (705-165-003 or 705-165-147; Jackson ImmunoResearch); https://www.jacksonimmuno.com/catalog/products/705-165-003
8. Donkey anti-guinea pig 405 (706-475-148; Jackson ImmunoResearch); https://www.jacksonimmuno.com/catalog/products/706-475-148
9. Donkey anti-rabbit 405 (711-475-152; Jackson ImmunoResearch); https://www.jacksonimmuno.com/catalog/products/711-475-152
10. Donkey anti-chicken 488 (703-545-155; Jackson ImmunoResearch); https://www.jacksonimmuno.com/catalog/products/703-545-155

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	1208-2 (UCLA IRB#13-000296); 8119-1 (UCLA IRB#13-000296); 0524-1 (Stanford IRB-27795).
Authentication	Genomic integrity was assessed by SNP microarray "GSAMD-24v2--0" and pluripotency was assessed by immunocytochemistry (Khan et al., Nature Medicine 2020)
Mycoplasma contamination	All cell lines tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	This study utilized both male and female Esco2fl/fl (JAX, B6N.129S-Esco2tm1.1Ge/J), Emx1-cre (JAX, B6.129S2-Emx1tm1(cre)Krj/J), C57BL SCID (JAX, B6.Cg-Prkdcscid/SzJ), and Swiss Webster (Charles River, Crl:CFW(SW), active dams custom ordered) mice. Mice were group housed with littermates and kept on a 12:12 light:dark cycle and at a temperature of 20-23C and 40-60% humidity.
Wild animals	This study did not involve wild animals.
Reporting on sex	Data were collected from both sexes. For unsupervised segmentation of spontaneous behavior with Motion Sequencing, males (n=43) and females (n=44) were modeled separately. Group x sex interactions were tested for all the directed behavioral tasks in Fig. 4 and Extended Data Fig. 9. We did not find any significant interactions, so data from male and female mice were pooled. These studies were not powered to detect modest effects of sex, but 3 out of 48 directed behavioral results (Catwalk Swing Duration, Y-maze distance, AC vertical time) revealed a main effect of sex and details are presented in Table S5. Sex × genotype effects were not estimable in the hypoxic injury experiments because the apallial group contained only female mice. Sex was therefore assessed in the two groups where it could be: a genotype × sex ANOVA restricted to control and XCX mice revealed no main effect of sex and no

genotype × sex interaction for any of the CatWalk measures in Fig. 5k–n and Extended Data Fig. 10f–i. Sexes were pooled for the analyses shown.

Field-collected samples: This study did not involve samples collected from the field.

Ethics oversight: For datasets generated by the authors: Prior to starting experiments, we sought approval from the Stanford Administrative Panel on Laboratory Animal Care (APLAC). Beyond this conventional review procedure, we requested higher-level consultation with bioethicists at Stanford who provided an extensive overview of implications. Soon after, we consulted with the Stanford Neuroscience Institute's Executive Committee that included leading neuroscientists. As we initiated these experiments and established the apallial model in the laboratory, we also confidentially sought feedback from several neuroscientists around the world. In parallel, Stanford University School of Medicine convened an external committee composed of neuroscientists, legal scholars, ethicists, and patient advocates to develop a framework for oversight of these studies. We presented this work to the committee on multiple occasions, and their feedback substantially informed our approach.

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

Plants

Seed stocks: *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes: *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication: *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

Magnetic resonance imaging

Experimental design

Design type: Diffusion-weighted imaging (DWI), and susceptibility weighted imaging (SWI).

Design specifications: *Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*

Behavioral performance measures: *State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s): Structural, 2D SWI

Field strength: 7T

Sequence & imaging parameters: 2D T2w Turbo RARE sequence was used to obtain transverse structural MRI for volume measurements (TR/TE = 4500/41 ms, FOV = 18 x 18 mm2, matrix = 258 x 258 yielding 70 x 70 □m2 in-plane resolution, slice thickness = 0.5 mm, RARE Factor = 8, NEX = 2, scan time = 4 min 3 sec). For some animals, 3D RAREvfl sequence was used (TR/TE = 1800/90.26 ms, FOV = 15 x 15 x 8 mm3, matrix = 150 x 150 x 80 yielding 0.100 mm isotropic resolution, RARE Factor = 43, NEX = 1.6, scan time = 10 min 33 sec with Compressed Sensing) and data are reconstructed with Smart Noise Reduction option. For high resolution anatomical T2W images for DWI co-registration, two RARE (Rapid Acquisition with Relaxation Enhancement) sequences were acquired per animal: (i) a low-resolution scan (200 × 200 × 500 μm^3) for co-registration reference, and (ii) a high-resolution scan (100 × 100 × 300 μm^3; RARE factor = 4, TR/TE = 4000/30 ms, 2 averages, flip angle = 180°) used as the primary anatomical reference for visualization and registration. 2D SWI data were acquired (TR/TE = 858.8/18 ms, FOV = 15 x 15 mm2, matrix = 280 x 280 yielding 54 x 54 □m2 in-plane resolution, slice thickness = 0.5 mm, NEX = 6, scan time 21 min 54 sec).

Area of acquisition: Whole brain

Diffusion MRI: ☒ Used ☐ Not used

Parameters: Diffusion-weighted imaging (DWI) data were acquired with a 4-segment diffusion-weighted EPI sequence (Bruker:DtiEpi) using the following parameters: b = 1000 s/mm^2, 40 non-collinear diffusion-encoding directions plus 5 b = 0 images (45 volumes total), diffusion timing δ/Δ = 4.5/10.3 ms, TR/TE = 3500/21 ms, in-plane resolution 150 × 150 μm^2, slice thickness 500 µm, FOV = 16 × 15 mm^2, matrix = 107 × 100, bandwidth = 250 kHz, second-order B_0 map shimming, three FOV saturation bands. The B0 map was obtained prior to the sequence and first- and second-order shims were applied using mapshim. NiFti formatted data were converted

from Bruker ParaVision format using dcm2niix.

Preprocessing

Preprocessing software: MRTrix3

Normalization: *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template: Allen Atlas for the control group

Noise and artifact removal: *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring: *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings: *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested: *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

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

Anatomical location(s): *Describe how anatomical locations were determined (e.g. specify whether automated labeling algorithms or probabilistic atlases were used).*

Statistic type for inference (See Eklund et al. 2016): *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction: *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

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