

Integrated analysis of molecular atlases unveils modules driving developmental cell subtype specification in the human cortex

Received: 22 September 2023

Accepted: 27 February 2025

Published online: 21 April 2025

 Check for updates

Patricia R. Nano¹, Elisa Fazzari¹, Daria Azizad¹, Antoni Martija¹,
Claudia V. Nguyen¹, Sean Wang¹, Vanna Giang¹, Ryan L. Kan¹, Juyoun Yoo^{2,3},
Brittney Wick⁴, Maximilian Haeussler⁴ & Aparna Bhaduri¹✉

Human brain development requires generating diverse cell types, a process explored by single-cell transcriptomics. Through parallel meta-analyses of the human cortex in development (seven datasets) and adulthood (16 datasets), we generated over 500 gene co-expression networks that can describe mechanisms of cortical development, centering on peak stages of neurogenesis. These meta-modules show dynamic cell subtype specificities throughout cortical development, with several developmental meta-modules displaying spatiotemporal expression patterns that allude to potential roles in cell fate specification. We validated the expression of these modules in primary human cortical tissues. These include meta-module 20, a module elevated in FEZF2⁺ deep layer neurons that includes TSHZ3, a transcription factor associated with neurodevelopmental disorders. Human cortical chimera experiments validated that both FEZF2 and TSHZ3 are required to drive module 20 activity and deep layer neuron specification but through distinct modalities. These studies demonstrate how meta-atlases can engender further mechanistic analyses of cortical fate specification.

The human cerebral cortex, expanded compared to rodents and other mammals¹, enables diverse biological processes that distinguish humans from other species, including judgment, perception and language. Many of these differences begin during development², and the signaling pathways and cell types that promote expanded cortical size and function in humans also create vulnerabilities toward a variety of neurodevelopmental and neuropsychiatric disorders³. As such, the study of human brain development is crucial to understanding adult human cell types, function and disease.

Existing efforts to parcellate cells in the developing brain into cell types have relied heavily on marker genes^{4–6}, which has successfully delineated cell types and subtypes but may not fully encompass the span of gene programs represented during a complex process such as development. Mechanistic investigations of these marker genes have

provided foundational principles of cortical cell fate specification; however, these genes alone are insufficient to define developmental cell types existing on a continuum in which several marker genes can be co-expressed in the same cell. A nuanced, complete and unbiased picture of the gene networks driving cell type specification can reveal the emergence of biological function in developmental datasets. Thus, combining the power of multiple datasets and leveraging additional methods of interrogating their identity are essential to describing the cellular diversity of human cortical development.

The value of inventorying the cell types and states that exist in human brain development has been well appreciated by work as part of the BRAIN Initiative Cell Census Network⁷, the Human Cell Atlas⁸ and individual laboratories^{4–6,9–19}, resulting in numerous single-cell RNA sequencing datasets in the last few years that focus on cataloging cell

¹Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA. ²Department of Neuroscience, Columbia University, New York, NY, USA. ³Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA. ⁴Genomics Institute, University of California, Santa Cruz, Santa Cruz, CA, USA. ✉e-mail: ABhaduri@mednet.ucla.edu

types. Although these resources are important, each study is limited by the realities of rare samples: no one study has yet to comprehensively profile the entire span of cortical regions or stages with enough sample numbers to instill confidence that the atlasing effort is complete.

Gene co-expression efforts have historically been used in bulk RNA sequencing studies to provide unique insight into gene modules that explain key disease mechanisms^{20–23}, with more recent applications in single-cell data²⁴. In the current study, we applied a meta-module gene co-expression strategy based on iterative hierarchical clustering to integrate recently published single-cell transcriptomic profiles of the developing and adult human cortex and extract gene networks that describe not only cell type but also key developmental processes and signatures of maturation. To focus our efforts on the gene networks that establish cortical cell types, our collection of developmental datasets focused on stages of peak neurogenesis as well as the transition from neurogenesis to gliogenesis. These timepoints enable our analyses to capture when radial glia, the neural stem cells of the cortex, specify into multiple subtypes and give rise to the vast majority of cell type populations in the cortex.

Using this strategy, we found networks with relevance to the neurogenesis-to-gliogenesis transition and the establishment of specific neuronal subtypes found in the adult human brain. We used immunostaining to validate the cell type and temporal activity patterns of key meta-modules in primary developing human cerebral cortex. Furthermore, we functionally interrogated a deep layer-associated meta-module using human cortical chimeroids—a recently described model of the developing human cortex in which cortical organoids are generated by combining multiple pluripotent stem cell lines²⁵. Across three donor lines, our experiments demonstrated the ability of subtle differences in meta-module activity to cascade into dramatic changes in cell type composition. These gene networks represent a diverse array of developmental processes, comprising a resource applicable to a broad range of questions concerning the development of the human cortex. Our findings also suggest that the meta-atlas strategy can be leveraged to derive biological insight across atlasing efforts that continue to be a centerpiece in the field.

Results

Iterative clustering identifies gene networks in meta-atlas

Our meta-atlas of the developing human cortex consists of seven recently published single-cell transcriptomic datasets^{3,6,9,12,13,16,18} containing 599,221 cells from 96 individuals, spanning gestational weeks (GWs) 6–40 and 8 months postnatal (Supplementary Tables 1 and 2). We first performed rigorous quality control and co-clustered our dataset using the established reciprocal principal component analysis (PCA) pipeline²⁶ (Methods) to establish that expected cell types and published marker genes could be clearly identified, enabling visualization and verification of our input datasets (Supplementary Table 3). Our meta-atlas contains the expected distribution of cell types and subtypes (Fig. 1a, Extended Data Fig. 1 and Supplementary Fig. 1), expressing appropriate marker genes (Fig. 1b and Supplementary Table 4).

Our integration analysis, as well as parallel alternative approaches^{27–29} (Supplementary Fig. 1c), instills confidence in the integrity of our large meta-atlas dataset, and pseudotime analyses^{30,31} validate our ability to demonstrate that our integration shows the continuum of cell states (Supplementary Fig. 1d). However, technical variation between studies may obscure important biological processes, emphasizing the need for module-based analysis to further explore dynamic developmental processes.

To use these data to extract gene networks that effectively describe human cortical development, we performed a meta-module analysis. We first identified the key sources of transcriptomic variation within each individual in our dataset, clustering the cells within each individual and identifying cluster marker genes (Fig. 1c and Methods). We then scored the ability of each marker gene to describe a cluster using

a metric based on the specificity and enrichment of a marker gene's expression to its assigned cluster (Methods).

Next, we identified the genes that are linked not only within cells of the same individual but also across datasets and the entire meta-atlas (Extended Data Fig. 2). We aggregated the cluster markers from all individuals in our meta-atlas and isolated marker genes that surpassed the 90th percentile of gene scores across the entire meta-atlas. This generated a filtered list of cluster markers that are most representative of the transcriptional signatures detected within each individual, which were then hierarchically clustered into meta-modules (Fig. 1d). In this way, we generated 225 meta-modules that consist of genes that co-express across different individuals, datasets and ages—genes whose association with each other withstands technical noise (Supplementary Table 5). We observed significantly higher gene expression correlations within our modules, validating our technical approach (Supplementary Fig. 2a,b). Upon comparing our meta-modules to widely used gene co-expression and regulon analyses^{21,32}, we found that our approach binned genes into sets in almost entirely unique ways (Supplementary Fig. 2c and Supplementary Tables 6 and 7). We, therefore, hypothesize that our meta-modules provide an orthogonal co-expression strategy to discovering biological processes important to human cortical development identified in a wholly unbiased manner.

Meta-modules capture broad and cell-type-specific processes

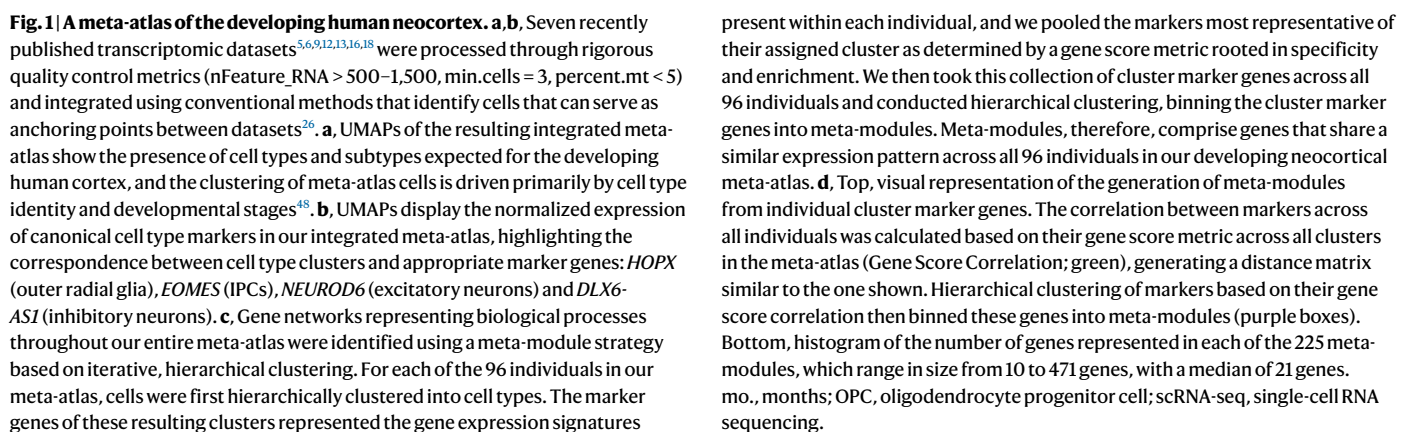
We examined the biological roles of these networks through detailed annotation of each meta-module, ranging from 10 to 471 genes in size. By surveying the signaling pathways, subcellular localizations, transcription factors (TFs)³³ and cell types represented in each meta-module as well as conducting literature review into meta-module genes, we were able to assign biological processes to the majority of meta-modules (Fig. 2a, Supplementary Tables 8 and 9 and Supplementary Fig. 3). These roles spanned a wide range, including synapse function, immune function, cell fate and cell division (Fig. 2a).

To test the accuracy of our annotations, we first examined whether meta-modules annotated with cell-type-specific processes were more active in expected cell types (Supplementary Tables 10 and 11). These analyses scored cells using a meta-module activity score based on average meta-module gene expression (Methods), similar to other published methods (Extended Data Fig. 3). We found that the activity of meta-modules with functions specific to vascular or immune activity were enriched in endothelial and microglial clusters, respectively (Fig. 2b). Similarly, some meta-modules related to oligodendrocyte and astroglial function were specific to those cell types.

To more systematically explore the cell-type-specific patterns of our meta-modules, we developed a module specificity score that measures how enriched a module is in a particular cell type or subtype (Methods). Interestingly, we observed that developmental cell types are characterized by groups of modules instead of singular, highly specific modules (Fig. 2d, Supplementary Fig. 3e and Supplementary Table 12), suggesting that, in development, individual modules can be shared by multiple cell types, and each cell type is distinguished, instead, by a unique combination of modules.

Adult meta-atlas highlights cell fate signature dynamics

To evaluate how well our meta-modules derived from the developing cortex could explain adult cell types, we sought to link modules between the developing and adult human cortices. To achieve this, we generated a parallel meta-atlas from 16 recently published^{14,15,17,20,34–45} transcriptional datasets from the adult human brain that consists of 2.6 million cells across 274 individuals aged 25 years or older (Extended Data Fig. 4 and Supplementary Table 13). We generated a uniform manifold approximation and projection (UMAP) for visualization and annotated the adult cell types by mapping the broad and granular annotations from one of the datasets comprising this meta-atlas³⁶ (Fig. 3a and Supplementary Tables 14 and 15). This resulted in the generation of



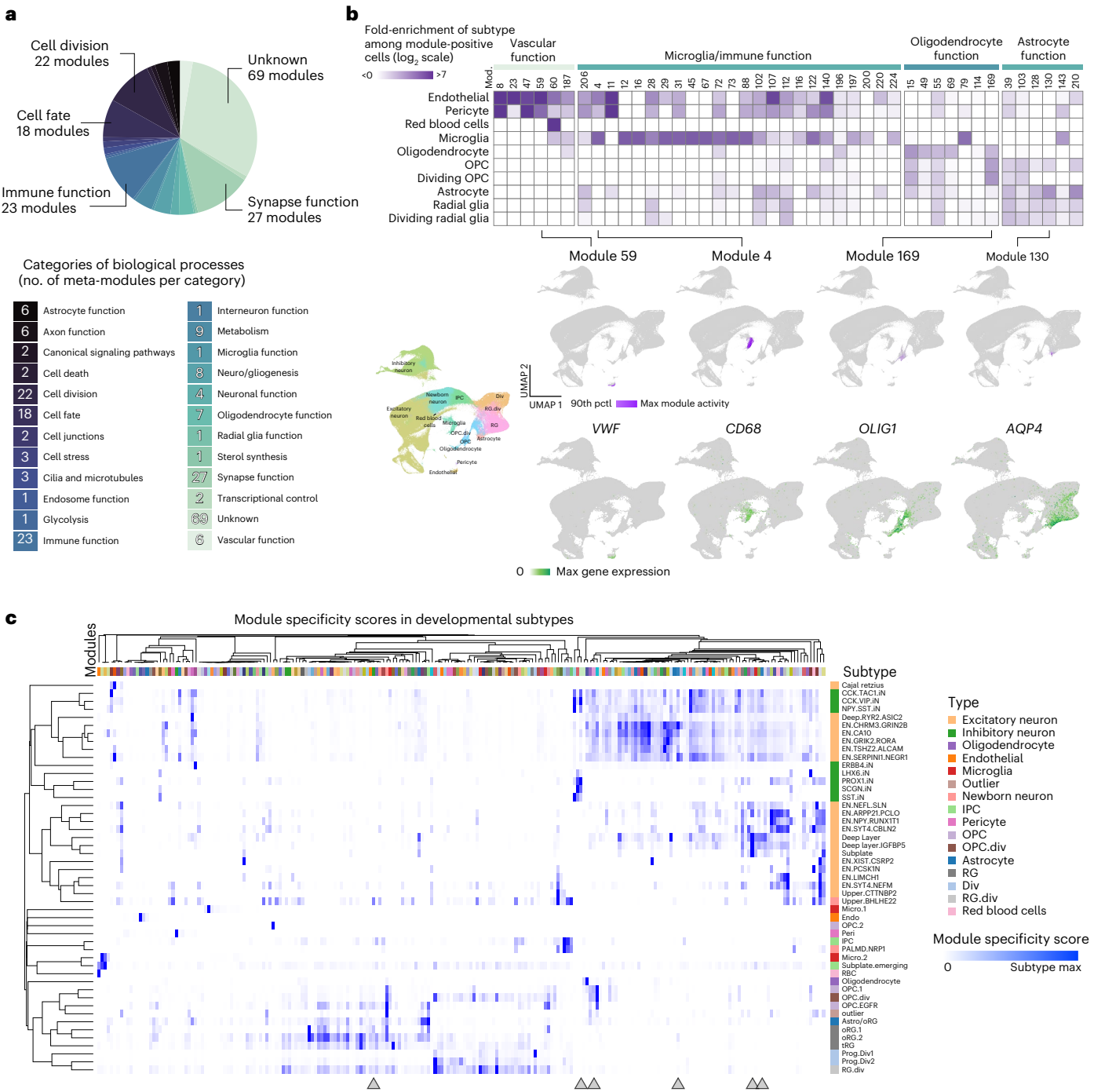
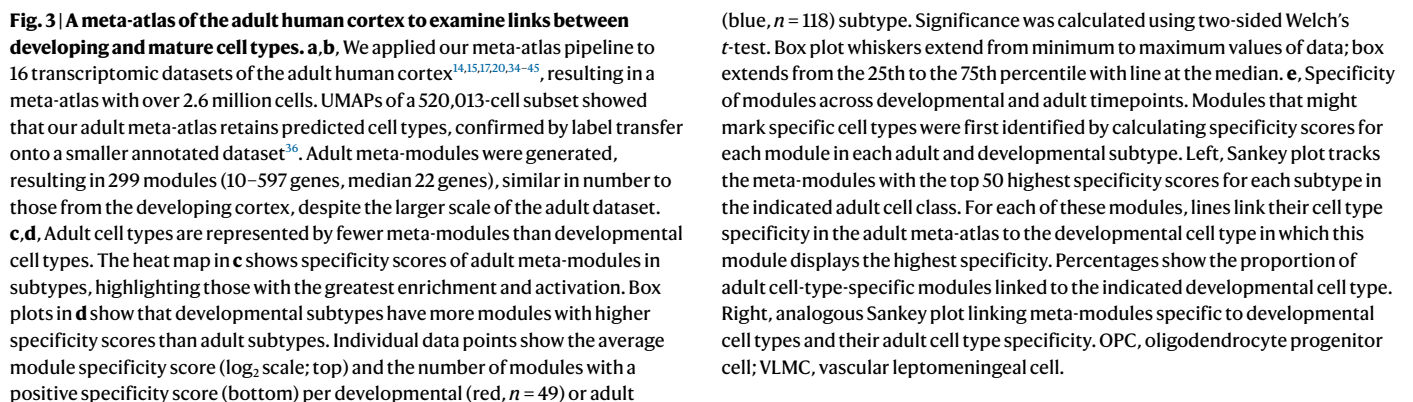


Fig. 2 | Meta-modules reveal developmental, cortical gene networks. a, Of the 225 modules, 156 were confidently assigned biological annotations, with most modules associated with synaptic function, immune function, cell division and cell fate. Meta-modules were annotated based on rigorous literature review of meta-module genes and an enrichment analysis of terms from signaling pathway databases (WikiPathway 2021 Human, KEGG 2021 Human and Elsevier Pathway Collection), transcriptional regulatory collections (ChEA 2016, ENCODE and ChEA Consensus TFs from ChIP, TF Perturbations Followed by Expression and TRRUST Transcription Factors 2019) and Gene Ontology sets (GO Biological Process 2021, GO Molecular Function 2021 and GO Cellular Component 2021). **b**, Meta-modules associated with cell-type-specific functions (that is, that of vascular cells, microglia/immune cells, oligodendrocytes and astrocytes) were active predominantly in those corresponding cell types. Enrichment analysis was calculated by first isolating the cells in the meta-atlas displaying the greatest

activity for the indicated module (module-positive cells) and then evaluating proportional enrichment for a given subtype in each module (purple). UMAPs are shown highlighting the distribution of module activity (top UMAP) and cell-type-specific marker gene expression (bottom UMAP) for select meta-modules: module 59 (vascular cells), 4 (microglia), 169 (oligodendrocyte lineage) and 130 (astrocytes). **c**, Developmental subtypes are represented by multiple modules that show similar levels of subtype specificity and enrichment, as measured by our module specificity score. This metric scores the relative activity of the module in a given cell type as well as the enrichment of a given cell type among module-positive cells. Heat map displaying the module specificity scores for all modules in all developmental subtypes demonstrates that subtypes are overall characterized by several high-scoring modules (blue). Gray arrowheads highlight example vignette modules (from left to right: modules 156, 144, 94, 134, 20 and 189). OPC, oligodendrocyte progenitor cell; pctl, percentile.



a similar number of meta-modules as generated by our developmental meta-atlas (299 meta-modules; Fig. 3b and Supplementary Table 16), instilling confidence in the fact that the principles of our developmental meta-atlas are cross-applicable to new, substantially larger datasets.

As in development, we leveraged our module specificity score to analyze how these adult modules are distributed among adult cell types. In contrast to development, we observed that adult subtypes were largely characterized by one or two strikingly specific adult modules (Fig. 3c and Supplementary Tables 17 and 18). Given this discrepancy, we sought to explore whether these trends are due to differences between the modules or between the cell types in development versus adulthood. Thus, we performed reciprocal analyses where adult modules were analyzed in developmental cell types and developmental modules were scored on adult cell types (Extended Data Fig. 5a and Supplementary Tables 19 and 20). This analysis convincingly showed that, regardless of which set of modules is being scored, developmental cell types express a broader range of modules with greater specificities, whereas adult cell types are more restricted in their module expression (Fig. 3d).

This naturally led to the question of how gene programs are changed or refined from development to adulthood. We directly compared the composition of modules between adult and development, noting that, although many adult modules shared a significant overlap with developmental modules (Extended Data Fig. 5b), the two module collections overall represent unique groupings of genes. Among neuronal populations in our developmental meta-atlas, we observed that later-born subtypes, such as upper layer neurons, were characterized by a greater ratio of adult versus developmental modules (Extended Data Fig. 5c). We sought to more broadly examine this trend, exploring how modules that are specific to cell subtypes in the adult could be tied to developmental cell subtypes. We also examined the converse, analyzing how modules specific to developmental subtypes can characterize subtypes in the adult (Fig. 3e and Extended Data Fig. 5d). Overall, most modules specific to excitatory neuron subtypes in the adult mark excitatory neurons during development (Fig. 3e). However, when mapping these subtype-specific modules to individual excitatory neuron subtypes, we found that the modules that mark excitatory neuron subtypes in the adult trace back to a variety of subtypes in development (Extended Data Fig. 5d), alluding to the presence of transcriptomic refinement at the cell subtype level. By conducting this reciprocal analysis, we also confirmed that the modules specific to certain developmental subtypes are representative of a broad array of adult cell subtypes, most notably among developmental progenitors (Fig. 3e and Extended Data Fig. 5d). Intrigued by these trends, we more closely examined the modules specific to progenitors during development (Extended Data Fig. 5e). As expected, most of these modules (40–60%) are specific to astrocytes in the adult. However, we found that developmental and adult modules that are specific to progenitors during development can mark distinct sets of neuronal cell types in the adult. In one interesting example, 6% of modules specific to radial glia during development were most specific to inhibitory neurons in the adult, but developmental modules marked different subtypes of inhibitory neurons than did adult modules. This is consistent with recent work⁴⁶ that observed the local dorsal generation of inhibitory neurons between GW 15 and GW 18. Further examination of our dataset in this manner can reveal other gene programs that might allude to the refinement of progenitors toward adult subtypes.

Previous examinations of human cortical development noted large differences in cell types and gene expression programs between human developmental and adult timepoints⁵⁹, with substantial changes in gene expression occurring between peak neurogenesis and early childhood periods^{22,47,48}. Our comparative meta-atlas approaches, therefore, confirmed long-standing observations: developmental subtypes are distinguished by subtle transcriptomic differences that sharpen in the adult human cortex, suggesting that, to decipher the

programs that drive cell fate specification, it is essential to view these processes from a developmental lens.

Meta-modules reveal gene programs initiating cell fate

The idea that module activity dynamics in progenitor cells during development can inform analyses of cortical cell type specification is particularly exemplified by meta-module 156. These efforts are facilitated by our meta-module activity metric, which enables the comparison of meta-modules from our developing meta-atlas with our adult meta-atlas. In addition, our module activity metric enables the evaluation of modules in external datasets spanning other timepoints. These datasets provide further validation of the module activity patterns observed in our developmental meta-atlas as well as provide a dataset in which to determine how our modules change over timepoints not included in our atlases.

In our developmental meta-atlas, module 156 activity is particularly elevated within progenitor cells and becomes more active in radial glia throughout development. These patterns were recapitulated in the second-trimester timepoints of other developmental cortical profiles as well as in the radial glia progenitors from the third trimester through adolescence (Fig. 4a and Supplementary Fig. 4a)¹⁷. Meta-module 156 is also specific to radial glia, as radial glial subtypes display the highest specificity scores for this module. These module expression patterns culminate in the high restriction of module 156 to glial subtypes in our adult meta-atlas, as can be seen via both module activity and specificity measurements (Fig. 4a).

Consistent with a role for meta-module 156 in gliogenesis, annotation of module 156 genes indicated a role in radial glia biology, neuronal activity and response to stimuli (Supplementary Tables 8 and 9). One of the meta-module 156 member genes, *QKI*, has reported roles as a pan-glial progenitor marker⁴⁹. We validated this model further by determining the co-localization of meta-module 156 genes in progenitors of the ventricular zone (VZ) in primary human cortex from GW 16 and GW 20 (Fig. 4b and Supplementary Figs. 4b,c and 5).

We focused first on *QKI* and the meta-module 156 gene *PDLIM5*, a post-synaptic scaffolding protein with roles in dendritic spine development^{50,51} and neuropsychiatric disorders^{52–54} but no reported role in glial fate specification. Both *QKI* and *PDLIM5* expression was observed as early as GW 16 most prominently in the intermediate zone (IZ) and VZ, with *QKI* levels increasing in intensity at GW 20 (Fig. 4b and Supplementary Figs. 4b,c and 5a). These expression patterns exhibited partial co-localization within the VZ, in which punctate *PDLIM5* staining can be seen surrounding *QKI*⁺ nuclei. To validate the module further, we also stained for co-expression of other module members, including *SALL1*, a TF⁵⁵, which we observed to be co-localized with *QKI* as predicted by the module (Fig. 4b and Supplementary Fig. 5b). We further observed the co-localization of two additional module members, *ZFP36L1/2*, which was described as mediating the transition between neural progenitors and glia in both development and brain cancer⁵⁶ and *FOS* (Supplementary Fig. 5c). *FOS* is an early response gene⁵⁷ typically associated with neurons, and its role in progenitors is not well described, but our meta-module construction and immunostaining validate that it is expressed in the progenitor zone and that it co-localizes with other module 156 genes. These findings lend further support to the model that meta-module 156 may represent a gene program that signals a shift within progenitors from neurogenesis to gliogenesis—a process that has been well connected to timing and cell subtype but has not been well defined molecularly⁵⁸. Interestingly, although module 156 contains members with described roles in gliogenesis (notably, *QKI* and *ZFP36L1/2*), these regulators were not grouped together in established analyses, such as STRING interaction network analyses (Supplementary Fig. 5b). Module 156, therefore, demonstrates how our meta-module generation approach can reveal relationships between gene sets that can provide an orthogonal description of cell fate specification. Nonetheless, our current results

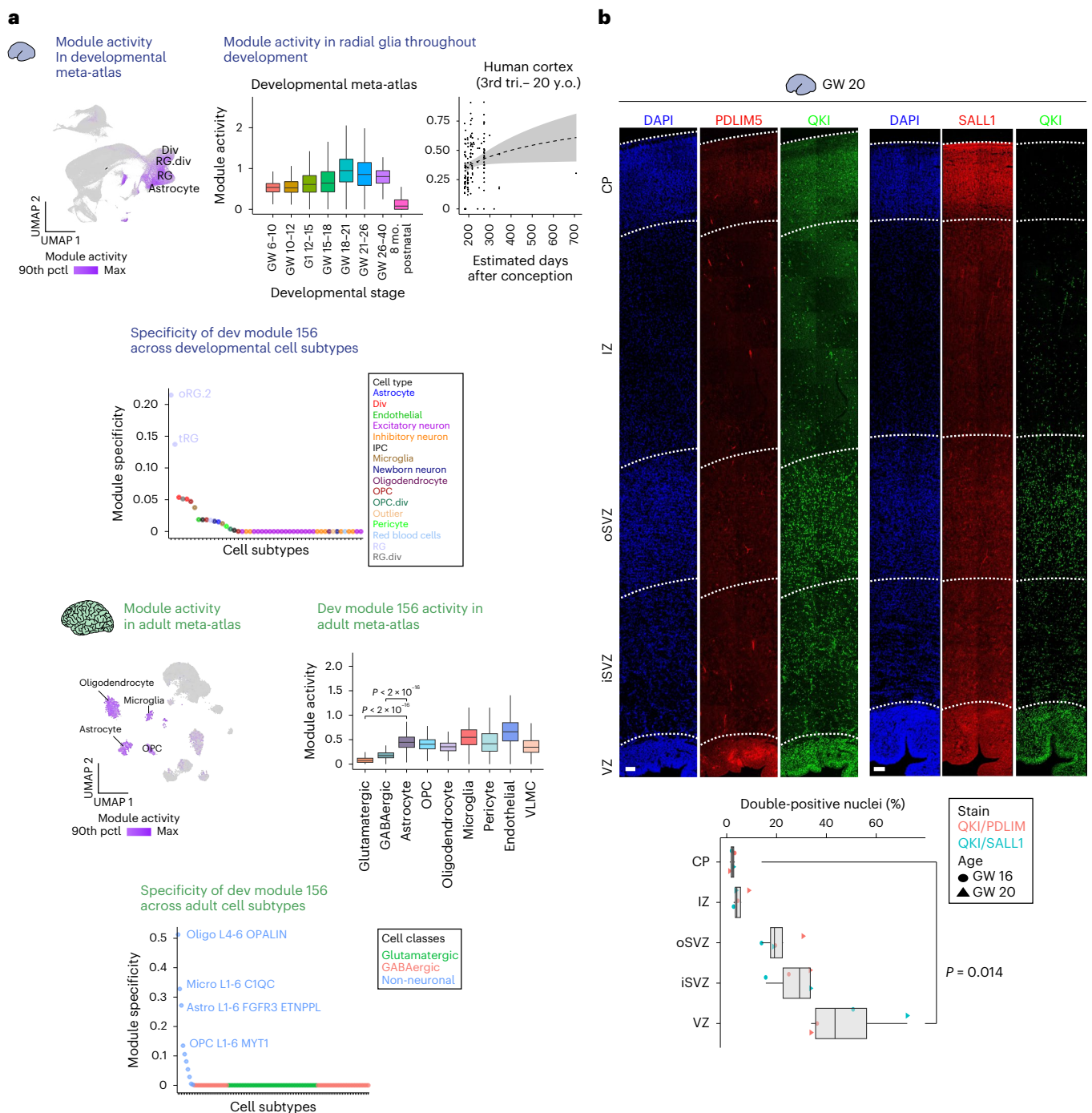


Fig. 4 | Developmental modules reflect initiation of cell fates in neural progenitors. a, Module 156 identified by comparing cell type distribution of module activity across developing (top) and adult (bottom) meta-atlas. Top-left UMAP displays binarized module 156 activity during development, enriched among progenitors. Within glial progenitors, module 156 activity increases throughout most developmental stages in our meta-atlas (middle box plots) and between the third-trimester and young adulthood timepoints in the Velmeshev et al.¹⁷ datasets (middle scatter plot with linear regression model; error bands indicate 95% confidence interval). Module 156 also displays the highest specificity scores in oRG.2 and tRG radial glia subtypes in the developing meta-atlas (below). These patterns culminate in module 156 activity and specificity patterns highly restricted to glial cells in our adult meta-atlas (bottom). Significance among the module activities of glutamatergic neurons ($n = 258,734$), GABAergic neurons ($n = 109,360$) and astrocytes ($n = 29,058$) in the adult meta-atlas was calculated with two-sided Welch's t -tests. **b**, The expression of module 156 genes in developing human cortical VZ was validated with immunofluorescent staining for QKI (green) and either PDLIM5 or SALL1 (red)

in primary human cortical tissues from GW 16 and GW 20. Top, representative immunofluorescent staining of QKI (green) and PDLIM5 (red) or QKI (green) and SALL1 (red) in the GW 20 cortex. Lines demarcate cortical layers identified based on well-established cell density patterns⁸⁵, visualized by DAPI. VZ, inner and outer subventricular zones (iSVZ and oSVZ, respectively), IZ and CP are shown. Scale bar, 100 μ m. Bottom, module 156-expressing cells are significantly enriched in VZ, as evaluated by quantification of double-positive (QKI⁺/PDLIM5⁺ or QKI⁺/SALL1⁺) nuclei across cortical layers in stained GW 16 and GW 20 slices ($n = 4$; dots represent measurements from each image). Staining in each image was quantified as the number of QKI⁺ nuclei overlapping with PDLIM5⁺ or SALL1⁺ puncta, normalized by the total number of DAPI⁺ nuclei. Significance values were calculated with two-sided Welch's t -tests. All box plot whiskers extend from minimum to maximum values of data; box extends from the 25th to the 75th percentile with line at the median. mo., months; OPC, oligodendrocyte progenitor cell; pct., percentile; tri., trimester; VLMC, vascular leptomeningeal cell; y.o., years old.

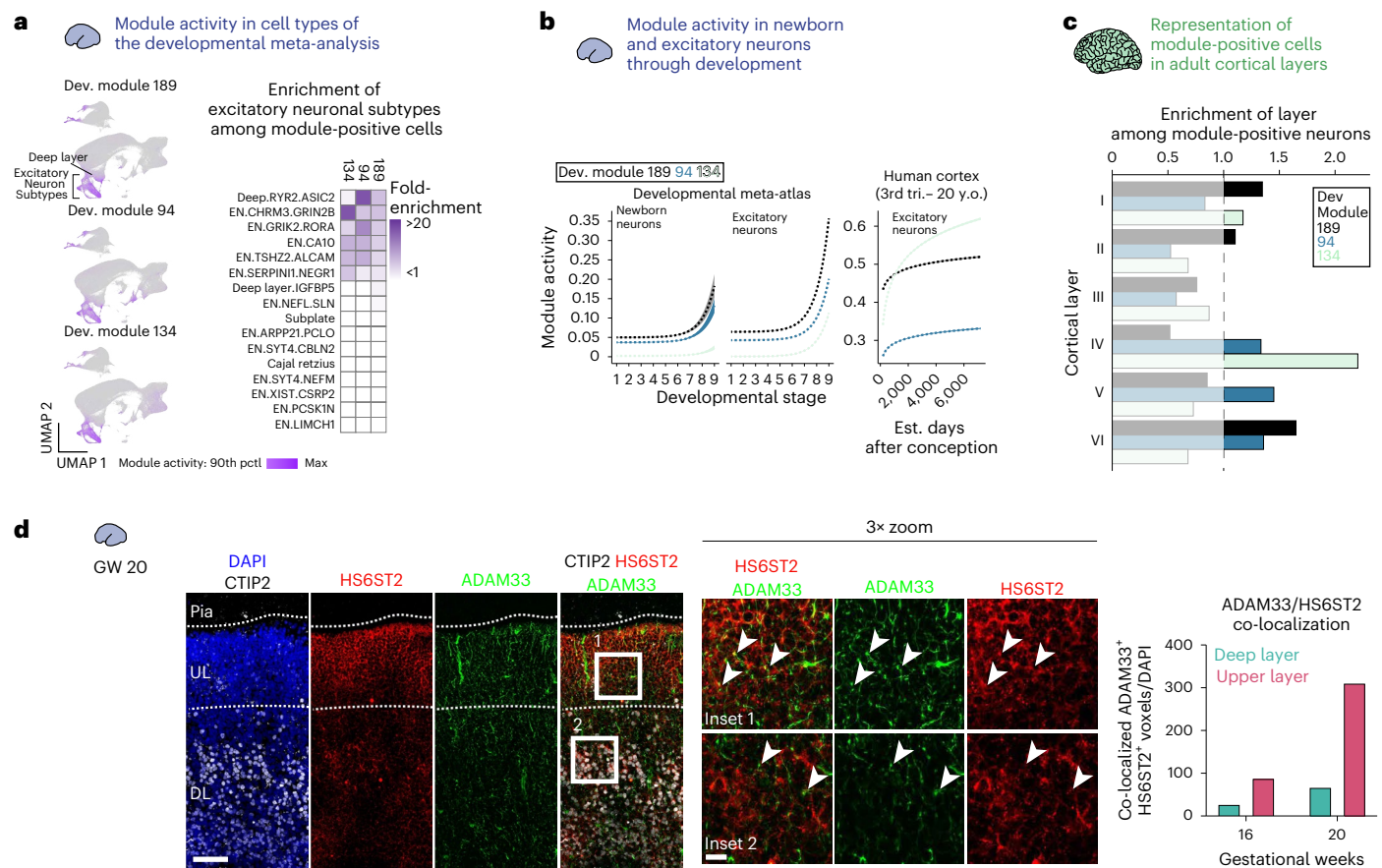


Fig. 5 | Developmental modules can reflect the maturation of adult cortical excitatory neurons. **a**, Modules 189, 94 and 134 were identified through an analysis identifying modules specific to neuronal subtypes during development but broad in the adult. The UMAPs (left) and heat maps (right) highlight this subtype enrichment, plotting the maximum module activity score (left) for each module as well as the fold enrichment of neuronal subtypes among cells positive for these modules (heat map, right). **b**, Module activity is represented as exponential and logarithmic regressions, showing that all modules are increasing in activity over time both in the developmental meta-atlas and during the third-trimester through young adulthood timepoints of the Velmeshev et al.¹⁷ dataset. Deep layer modules are more activated at earlier timepoints, consistent with the inside-out manner of neurogenesis. **c**, Analysis of module activity in the Jorstad et al.³⁶ adult cortex dataset highlighted cell type specificity for module 134 (layer IV) and modules 189 and 94 (deep layers), consistent with their order

of activation. Enrichment scores were calculated by first isolating neurons from the adult dataset and then evaluating proportional enrichment for a given layer in each module; the dashed line at 1 indicates above which enrichment is greater than the expected distribution. **d**, Enrichment of module 134 member-expressing cells in the upper versus deep layers of the cortex was validated using immunofluorescence for HS6ST2 (red) and ADAM33 (green) in primary human cortical tissues from GW 16 and GW 20. Left, representative images from GW 20 cortical tissues, which were co-stained with CTIP2 (white), showing that the co-localization of module 134 members is greater in neurons residing above the CTIP2 layer (3×-zoom insets and white arrows). These observations were further confirmed by quantifying the number of co-localized ADAM33⁺ and HS6ST2⁺ voxels, normalized by total number of DAPI⁺ nuclei, in upper versus deep layers (right bar graphs). Scale bar, 20 μm. DL, deep layer; pctl, percentile; tri., trimester; UL, upper layer; y.o., years old.

are hypothesis generating based upon our observed co-expression and predicted protein–protein interaction analysis, although future work could experimentally test this observation.

Meta-modules reveal gene programs that refine cell type

We next examined the ability of our meta-modules to represent the adoption of specific neuronal subtype identities. Our analyses revealed three meta-modules that were active in specific excitatory neuronal subtypes during development (Fig. 5a and Supplementary Table 21). These include meta-modules 189 (characterized by a mix of genes associated with the induction of neurogenesis) and 94 (associated with ion channels and synaptic signaling) active in deep layer neuronal subsets. We also assessed meta-module 134 activity primarily in *GRIN2B* excitatory neurons and comprising genes characterized to have roles in synapses. For this analysis, we leveraged data from one of the studies in our adult meta-atlas that included layer-dissected and sequenced cell types³⁶ (Extended Data Fig. 6) to identify how these modules are expressed in adult cortical layers. All three meta-modules

increased in activity over developmental time, with meta-module 134 emerging later in newborn neurons than the other two (Fig. 5b and Supplementary Fig. 6a). To link these meta-modules to adult neuronal subtypes, we calculated the meta-module activity scores in cells from the human adult cortex. In the adult, the meta-modules were active in a broader range of cell types (Extended Data Fig. 6b) but did display some layer-specific activity (Fig. 5c and Supplementary Table 22). Although meta-modules 94 and 189 retain their specificity to deep layer neuronal subtypes in the adult cortex, meta-module 134 becomes specifically enriched in layer IV.

Given that meta-module 134 activity is enriched in layer IV of the adult human cortex, we explored whether this module might signal the specification of layer IV cells in the developing human cortex (Fig. 5d and Extended Data Fig. 6b). In the cortical plate (CP) of the developing human cortex (GW 16 and GW 20), we detected the expression of two meta-module 134 members, both of which have undercharacterized roles in brain development: *ADAM33*, a metalloprotease with elevated expression in the brain according to mouse studies⁵⁹, and *HS6ST2*, a

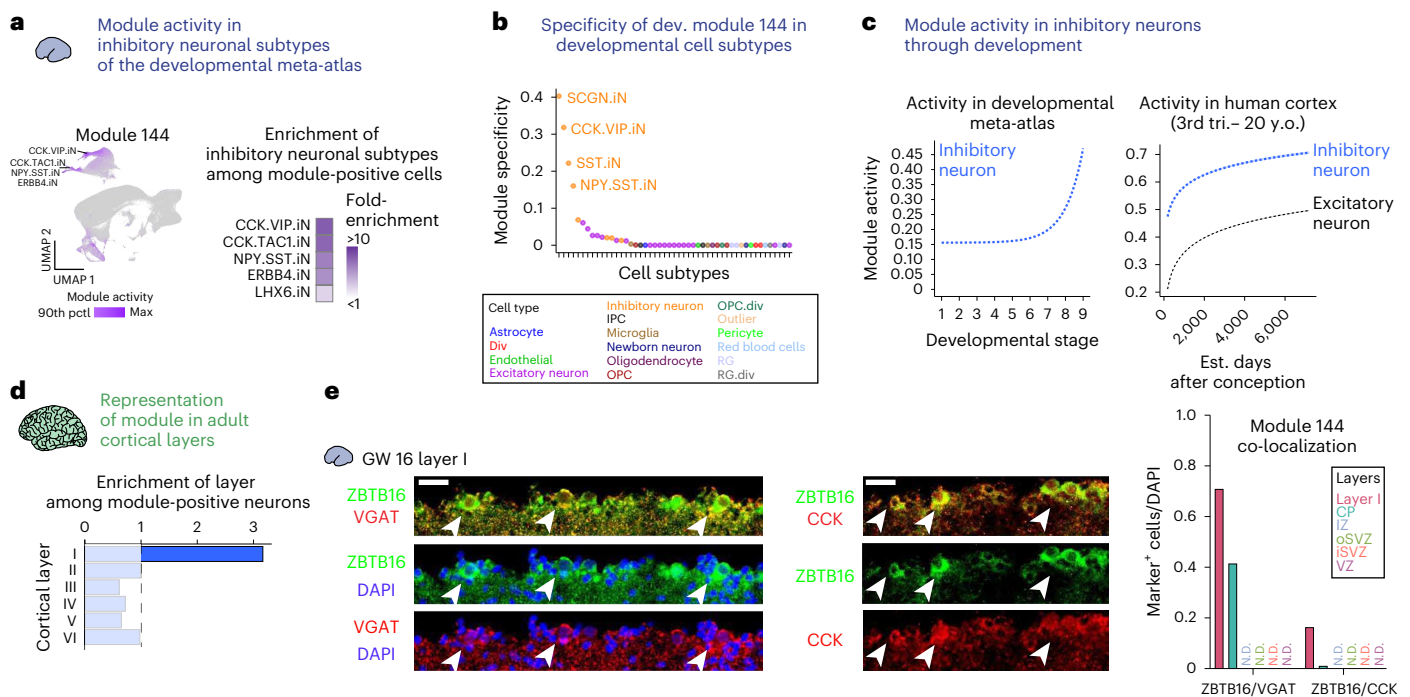


Fig. 6 | Developmental modules can reflect the maturation of adult cortical inhibitory neurons. a, b, Module 144 also displays a pattern of neuronal-enriched activity in development that broadens in adult, but it remains restricted within GABAergic neuronal subtypes. **a,** The UMAP and heat map highlight the restriction of module 144 activity in inhibitory neuronal subtypes in our developmental meta-atlas, with CCK-expressing interneuron subtypes displaying the greatest enrichment among module 144-positive cells. **b,** Consistent with these module activity patterns, module 144 specificity scores are highest among inhibitory neuronal subtypes. **c,** Exponential regression model of module 144 activity shows an increase in activity within inhibitory neurons throughout developmental stages of both our developmental meta-atlas and the third-trimester through young adulthood timepoints of Velmeshev et al.¹⁷ **d,** Calculation of module 144 activity in the adult cortex³⁶ showed an enrichment of layer I cells among module 144-positive neurons. Enrichment scores were calculated by first isolating neurons from the adult dataset and

then evaluating proportional enrichment for a given layer in each module; the dashed line at 1 indicates above which enrichment is greater than the expected distribution. The proportion of layer I cells was around three-fold greater among module 144-positive neurons than among all cortical neurons. **e,** Immunofluorescent staining for module 144 members VGAT, ZBTB16 and CCK in GW 16 primary human cortical tissues validated the enrichment of this module in layer I. Left, co-stains of ZBTB16 (green) and VGAT (red) display the enrichment of these module genes in layer I, where cells can be found to co-express these genes as highlighted by the white arrows. Middle, co-stains of ZBTB16 (green) with CCK (red) in layer I of GW 16 primary human cortex also indicate the co-expression of these module genes (white arrows). Right, these observations are supported by the quantification of the fraction of DAPI⁺ nuclei co-expressing both ZBTB16 and either VGAT or CCK across cortical layers, where co-expressing cells were not detected (N.D.) outside of the CP. Scale bar, 20 μ m. OPC, oligodendrocyte progenitor cell; pct, percentile; tri., trimester; y.o., years old.

heparan sulfate sulfotransferase linked to glioma⁶⁰ and X-linked intellectual disability⁶¹. ADAM33 was detected in the cytoplasm and plasma membrane of cells throughout the CP, whereas HS6ST2 was concentrated in upper layers (Fig. 5d and Supplementary Fig. 6b), particularly in cells lying above those expressing deep layer marker CTIP2 (Fig. 5d). At GW 20, upper layer cells also displayed partial co-localization of ADAM33 and HS6ST2 (Fig. 5d), as predicted by the enrichment of meta-module 134 activity in layer IV of the adult human cortex.

Another meta-module, 144, was detected to be specifically active in our developmental meta-atlas, predominantly in interneuron subtypes expressing cholecystikinin (CCK), a gene encoding a precursor for peptide hormones (Fig. 6a,b and Supplementary Table 21). Within these cells, meta-module activity also increases throughout development (Fig. 6c and Supplementary Fig. 6c). When exploring meta-module 144 activity in the adult human brain, we noted that it was enriched in layer I, which contains specific subtypes of inhibitory neuron populations^{62–64} (Fig. 6d, Extended Data Fig. 6b and Supplementary Table 22). We, therefore, tested the prediction that meta-module 144 may indicate the specification of interneuron subtypes. Meta-module 144 includes *SLC32A1*, which encodes the VGAT vesicular GABA transporters expressed almost exclusively in CCK.TAC1.iN and CCK.VIP.iN cells (Supplementary Fig. 6d). This module also contains the TF *ZBTB16* (also known as *PLZF*), which, intriguingly, has roles in neocortical thickness, deep layer specification and neurodevelopmental phenotypes in

the mouse cortex^{65,66}. Consistent with the enrichment of meta-module 144 activity in layer I of the adult human cortex, VGAT, ZBTB16 and CCK were detected in layer I of the developing human cortex (GW 16; Fig. 6e and Supplementary Fig. 6e), and cells displayed co-expression of cytosolic ZBTB16 with either VGAT⁺ or CCK⁺ puncta (Fig. 6e). These findings lend support to the model that meta-module 144 may indicate the initiation of layer I, CCK⁺ interneuron fates.

Across multiple meta-modules, we found that we can validate temporal and spatial expression patterns predicted from our data by grouping previously unconnected genes into meta-modules. The temporal dynamics of our meta-modules, therefore, mirror key changes in cell fate identity throughout development, suggesting that our meta-module analyses can be an unbiased way of revealing the previously inaccessible transcriptomic shifts that drive human cortical cell fate specification.

Meta-module 20 correlates with FEZF2⁺ deep layer neurons

To further examine how our meta-module analysis can inform mechanistic interrogation of cortical cell fate specification, we focused on meta-module 20. Meta-module 20 steadily increases in activity throughout the entirety of developmental stages assessed in our atlas (Fig. 7a and Extended Data Fig. 7), a pattern that is most prominent in deep layer neuronal subtypes. Interestingly, module 20 activity consistently remains low in upper layer neurons and inhibitory

neurons (Fig. 7a and Extended Data Fig. 7a). These patterns suggest that module 20 might reflect subtype-specific or maturation-specific signatures among excitatory neurons, but distinguishing between these two possibilities using our developmental meta-atlas proved difficult given the relative immaturity of upper layer neurons (Extended Data Fig. 7). We, therefore, examined the activity of module 20 in adult cortical subtypes, again leveraging the layer-specific adult dataset³⁶. In the adult human cortex, the cells with the greatest meta-module 20 expression are almost exclusively within deep layer neurons (Fig. 7b and Supplementary Table 23). Closer examination of these meta-module 20^{high} cells revealed an enrichment specifically in subtypes associated with the well-established deep layer terminal specification factor *FEZF2* (refs. 67–69) (Fig. 7b), even though *FEZF2* is not a gene within module 20. Although *FEZF2* is generally sparse in the adult, we did observe that these *FEZF2*-annotated deep layer cells display significantly high levels of both *FEZF2* and module 20 (Fig. 7b and Extended Data Fig. 7d).

We, thus, hypothesized that meta-module 20 may represent a gene network by which adult *FEZF2*⁺ subtype identities are established in early development. During gestation, *FEZF2* expression preceded meta-module 20 activity, peaking transiently in GW 10–15 and largely diminishing by the time meta-module 20 activity peaks (Fig. 7c). Although *FEZF2* expression in deep layer neuronal subtypes of the human cortex is well established^{48,70–72}, the vast majority of mechanistic analyses of *Fezf2* function have been conducted in the mouse^{67–69}. We, therefore, calculated the activity of meta-module 20 in recently published single-cell transcriptomic atlases of the developing and adult mouse cortex^{73,74}, finding that the activity of this module in the mouse cortex is less restricted than in the human (Supplementary Figs. 7 and 8). In the developing mouse cortex, meta-module 20 displays a sustained increase in activity in not only deep layer neuronal subtypes but in neuronal subtypes in the upper layers as well (Supplementary Fig. 7b). Meta-module 20 activity is retained in the adult mouse cortex, with activity highest among deep layer neuronal subtypes. However, unlike in the human cortex, meta-module 20 and *Fezf2* expression in these subtypes are no longer tightly correlated (Supplementary Fig. 8c). Meta-module 20 may, therefore, harbor putative roles in the execution of the specification of *FEZF2*⁺ deep layer neurons in the human cortex, distinct from its expression patterns in the mouse.

Fig. 7 | Module 20 activity may pattern *FEZF2*-expressing deep layer subtypes during development. a, Module 20 was identified in an analysis for modules with dynamic, cell-type-specific activities. Module 20 is enriched in neuronal subtypes (UMAP), increasing in activity in deep layer neurons (box plot). Box plot data represent cells in each developmental stage: Deep Layer, Stage (1) = 6; (2) = 114; (3) = 1,012; (4) = 443; (5) = 13,590; (6) = 2,450; (7) = 3; (9) = 1,835; EN.SERPINI1.

NEGRI, Stage (1) = 135; (2) = 20; (3) = 10; (4) = 52; (5) = 1,717; (6) = 1,215; (7) = 12; (9) = 6,485; Subplate, Stage (1) = 10; (2) = 178; (3) = 3,611; (4) = 663; (5) = 1,836; (6) = 1,968; (7) = 2; (9) = 461; EN.PCSK1N, Stage (1) = 33; (2) = 7; (3) = 72; (4) = 110; (5) = 11,745; (6) = 11,829; (7) = 32; (9) = 83; Upper.CTTNBP2, Stage (1) = 2; (2) = 33; (3) = 114; (4) = 272; (5) = 7,670; (6) = 16,279; (7) = 38; (9) = 2. b, In the Jorstad et al.³⁶ adult cortex dataset, module 20 is most active among deep layer subtypes (UMAP, cells with the top 10th percentile of module 20 activity in blue). Module 20-positive adult cells are enriched for deep layer and white matter cells (bar plot; dashed line, expected distribution). The average module 20 activity of layer V/VI subtypes annotated as *FEZF2* expressing by Jorstad et al.³⁶ is significantly higher than all other subtypes in that layer (box plots represent average module 20 activity of layer V/VI subtypes; dots show average activity of each individual subtype). Layer V: *n* = 19 *FEZF2*⁺ subtypes, 90 *FEZF2*[−] subtypes; Layer VI: *n* = 18 *FEZF2*⁺ subtypes, 85 *FEZF2*[−] subtypes. c, Sparse *FEZF2* expression in our developing meta-atlas (left UMAP), peaks in deep layer neurons at developmental stages preceding peak module 20 activity. *FEZF2* expression (log₁₀-normalized CPM; green) and module 20 activity (purple) in deep layer neurons are shown overlaid in the right box plot for each developmental stage. Number of deep layer neurons per stage described in a.

FEZF2 targets meta-module 20 genes, including *TSHZ3*

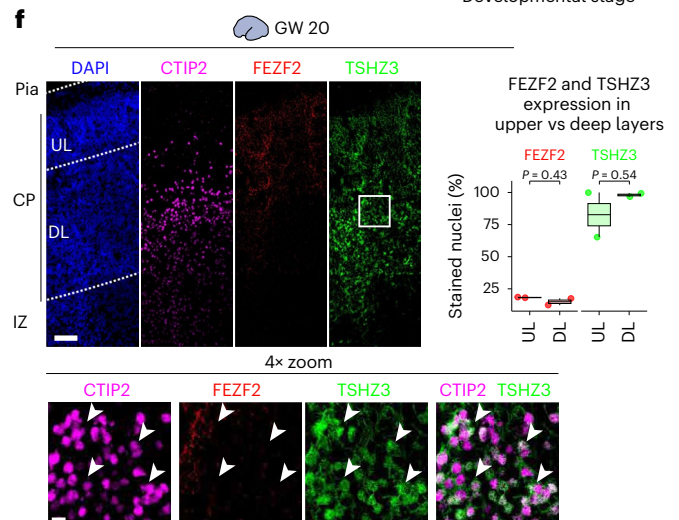
We hypothesized that, despite not being a member of meta-module 20, *FEZF2* induces the activation of meta-module 20 within maturing deep layer neurons. To investigate this model, we compared the overlap between meta-module 20 genes and putative *Fezf2* targets⁷⁵. Strikingly, almost half of meta-module 20 genes are putative *Fezf2* targets (35 of 74 genes) (Fig. 7d). To test whether meta-module 20 represents a bridge between *FEZF2* and full specification of these deep layer neurons, we searched for TFs among the putative *Fezf2* targets within meta-module 20. There is only one TF in module 20, *TSHZ3*, which was recently implicated in autism spectrum disorder⁷⁶. Previous transcriptomic efforts identified both *TSHZ3* and *FEZF2* as members of a neocortical gene network related to neuronal differentiation, axonal generation and neural projection that peaks in activity during development⁴⁸. In the mouse cortex, *Tshz3* depletion resulted in the altered expression of deep layer neuronal markers, including the upregulation of *Fezf2* (ref. 76). However, how these factors interact with each other and with other meta-module 20 genes to specify cell fates during human brain development is unclear. To investigate the role of *TSHZ3* in bridging the activity of *FEZF2* and meta-module 20, we compared the temporal dynamics of *TSHZ3* and module 20 expression during development. Interestingly, *TSHZ3* displays a transient activation in GW 15–21 (Fig. 7d), peaking in between the peaks of *FEZF2* and meta-module 20 activation. Notably, in the adult cortex, *TSHZ3* is broadly expressed (Supplementary Fig. 13d), potentially indicating a temporal restriction for its role in specifying *FEZF2*⁺ deep layer neurons. Taken together, these results suggest a model in which *FEZF2* acts within deep layer neurons early during development to activate *TSHZ3*, thus inducing the activation of meta-module 20 to promote the specification of *FEZF2*⁺ deep layer subtypes found in the adult.

FEZF2/*TSHZ3* co-localization in human deep layer neurons

To validate the cell-type-specific relationship between *FEZF2* and *TSHZ3* in the developing human cortex, we explored *FEZF2* and *TSHZ3* expression in sections of cortical tissue from GW 16 and GW 20 samples (Fig. 7e). In the cortex, we found that *FEZF2* expression is highest throughout the CP (Fig. 7e and Supplementary Fig. 13e), as expected from its well-established roles in the specification of deep layer neuronal identities. *TSHZ3* was also co-expressed with *FEZF2* throughout the CP in this stage (Fig. 7e and Supplementary Fig. 13e).

d, Venn diagram shows that 35 of 74 module 20 genes are putative *Fezf2* targets in mouse cortical progenitors⁷⁵, including the TF *TSHZ3* (UMAP, expression in developing meta-atlas). Box plot shows *TSHZ3* expression (log₁₀-normalized CPM; green) spikes within human deep layer neurons transiently in the middle of peak *FEZF2* expression and module 20 activity (purple). Number of deep layer neurons per stage described in a. e, Validation of *FEZF2* (red) and *TSHZ3* (green) expression throughout the CP via immunofluorescent staining of GW 16 primary human cortical tissues (3×-zoom insets highlight *FEZF2*⁺/*TSHZ3*⁺ cells). Lines demarcate cortical layers based on cell density patterns⁸⁵. Box plots show the significantly enriched co-expression of *FEZF2* and *TSHZ3* in the CP (including upper layers (UL) and deep layers (DL)) compared with ventricular zones (VZ), including inner subventricular zone (iSVZ), outer subventricular zones (oSVZ) and intermediate zone (IZ), IZ and CP). Stains from two GW 16 slices (dots) were quantified. f, GW 20 primary human cortical sections immunostained for *FEZF2* (red), *TSHZ3* (green) and deep layer marker CTIP2 (magenta) are consistent with our meta-atlas data. Cells co-expressing CTIP2 and *TSHZ3* are highlighted in the 4×-zoom insets and white arrows. Box plot shows *FEZF2* and *TSHZ3* expression quantification in these two images (dots), confirming *TSHZ3* enrichment in deep layers. All significance values were calculated two-sided with Welch's *t*-tests. For all immunofluorescent images, scale bar is 100 μm (main panels) and 20 μm (insets). For all box plots, whiskers extend from minimum to maximum values of data; box extends from the 25th to the 75th percentile with line at the median. ChIP-seq, chromatin immunoprecipitation sequencing; DL, deep layer; mo., months; pctl, percentile; UL, upper layer.

To maximize consistency and minimize any batch effects, we leveraged a recently described chimeroïd approach that combines multiple pluripotent stem cell lines after their specification to a cortical identity²⁵. Three human pluripotent stem cell lines were first used to generate cortical organoids using our standard protocols (Methods), before dissociation into single-cell suspensions and reaggregation in the presence of lentiviral constructs encoding a fluorescent reporter and a *FEZF2*-targeting shRNA, a *TSHZ3*-targeting shRNA or both to drive double KD of both genes (Fig. 8a). To validate that the chimeroïd approach was successful, we performed fluorescence-activated cell sorting (FACS) to enrich for green fluorescent protein (GFP)⁺ (FEZF2 KD), mCherry⁺ (TSHZ3 KD) and GFP⁺/mCherry⁺ (2×-KD) cells (Fig. 8a and Extended Data Fig. 8a). Quantitative real-time polymerase chain reaction (qRT-PCR) validated the KD of the target genes in these enriched



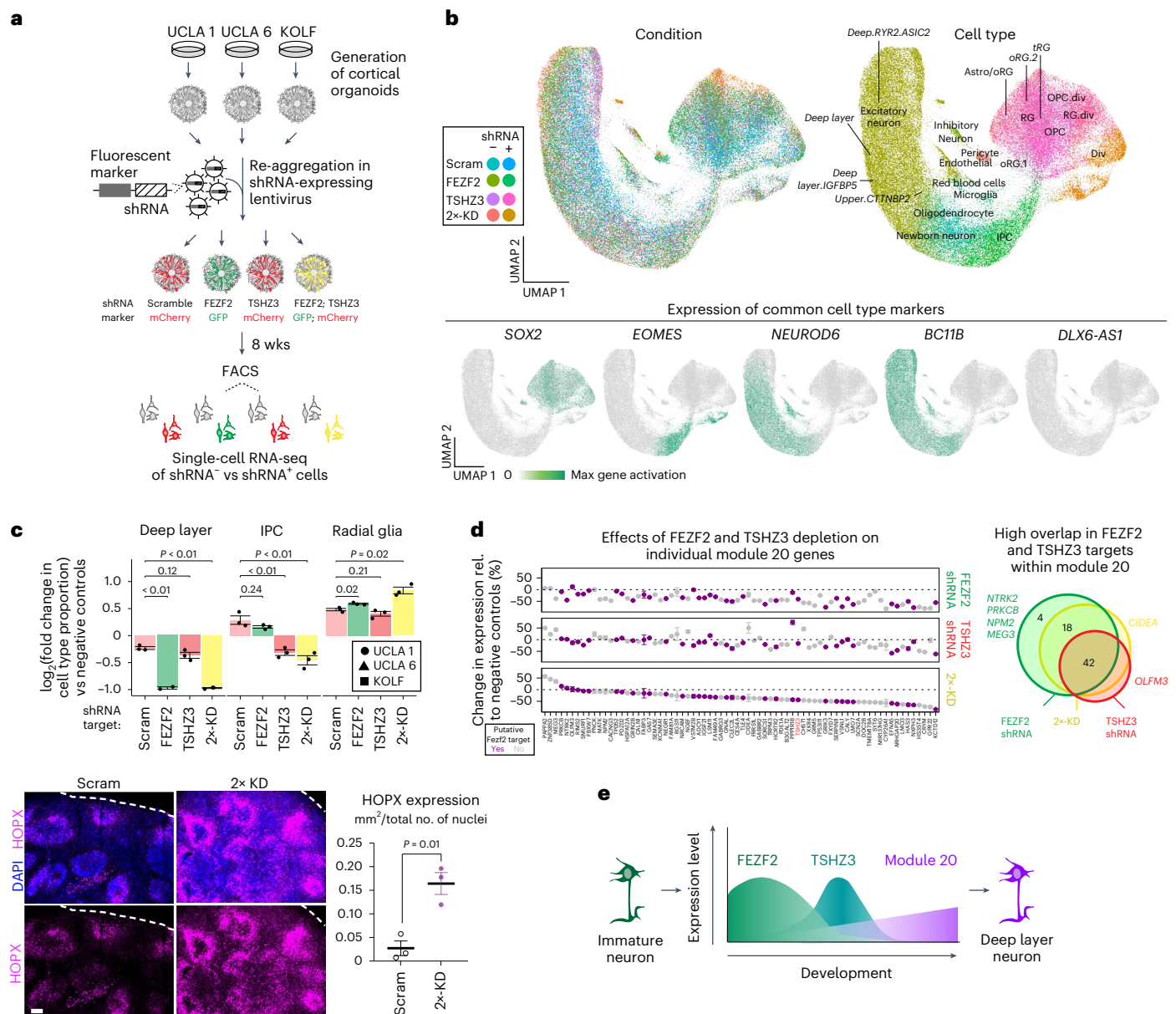


Fig. 8 | FEZF2 and TSHZ3 regulate module 20 to drive deep layer neuronal fates in human cortical chimeroids. **a**, Cortical organoids generated from three stem cell lines were dissociated and re-aggregated into cortical chimeroids²⁵ in the presence of fluorescent reporter-shRNA lentiviruses, including a 2x-KD condition (1:1 mixture of FEZF2 and TSHZ3 shRNAs) for enhanced KD. After 8 weeks, flow cytometry sorted cells based on shRNA expression. **b**, UMAPs show that 124,024 cells co-cluster regardless of shRNA condition, and projection onto the developmental meta-atlas confirmed appropriate cell type retention. Bottom UMAPs show expression of *SOX2* (progenitors), *EOMES* (IPCs), *NEUROD6* (excitatory neurons), *BCL11B* (deep layer neurons) and *DLX6-AS1* (inhibitory neurons). **c**, Double FEZF2/TSHZ3 KD decreased deep layer neurons and IPCs while increasing radial glia. Data points show fold change in cell type proportions in shRNA-expressing versus shRNA-negative cells for each cell line, with average values ± s.e.m. across all lines shown as bars. Changes in radial glia were corroborated by immunostaining of HOPX (outer radial glia marker), shown by

representative immunofluorescence images. HOPX⁺ area (mm²) was quantified and normalized by DAPI⁺ nuclei ($n = 3$ images, dots). Chimeroid boundaries are demarcated by dashed white lines (scale bar, 100 μ m). **d**, FEZF2 and TSHZ3 downregulation reduced the expression of most module 20 genes. Dot plots show average ± s.e.m. percent change in gene activity in shRNA-expressing versus shRNA-negative cells (23,782, 12,936 and 12,625 cells in the negative populations of FEZF2, TSHZ3 and 2x-KD conditions, respectively; 20,250, 14,056 and 7,099 cells in the positive populations, respectively). Similar effects were observed on module 20 genes that are putative FEZF2 targets (purple) versus non-targets (gray). A Venn diagram shows that 42 of the 74 module 20 genes are common targets of all three KD conditions. **e**, These experiments suggest a model in which waves of FEZF2 and TSHZ3 expression within deep layer neurons of the developing human cortex activate module 20, enabling maturation of deep layer neurons. OPC, oligodendrocyte progenitor cell; wks, weeks.

populations, and we proceeded with single-cell RNA sequencing using the particle-templated instant partition sequencing (PIP-seq) platform (Fluent Biosciences; Methods, Fig. 8b and Extended Data Fig. 8a).

Multiple lines of investigation, including immunofluorescence for CTIP2 (deep layer marker) and SOX2 (radial glia progenitor marker) as well as the mitochondrial content in the sequencing and

the expression of expected marker genes, verified the health of our chimeroids (Fig. 8b and Extended Data Fig. 8a), giving us confidence in the strength of the approach. Using reference single-nucleotide variants, we demultiplexed the PIP-seq data⁸¹ to retrieve biological replicates by sample in addition to the technical replicates across sorted samples (Extended Data Fig. 8b,c and Supplementary Table 28).

These biological replicates enabled rigorous analysis of our chimeroid KD experiment, and we observed a significant decrease of the target genes, which corresponded to a decrease of meta-module 20 activity (Extended Data Fig. 8d). Interestingly, these analyses also show that FEZF2 shRNAs can also reduce expression of *TSHZ3*, alluding to a model where FEZF2 activates module 20 in part through *TSHZ3*. We annotated the cell types in our dataset by leveraging projection strategies from the developmental meta-atlas (Methods, Fig. 8b, Extended Data Fig. 8e and Supplementary Tables 28 and 29), observing that multiple KD conditions altered the proportion of cell types in the glutamatergic lineage (Fig. 8c and Extended Data Fig. 8f). These changes in cell type proportion were most striking in the 2×-KD conditions, in which a decrease in both deep layer neurons and intermediate progenitor cells (IPCs) was accompanied by an increase in radial glia (Fig. 8c and Extended Data Fig. 8f). Compared to *FEZF2* or *TSHZ3* KD alone, the 2×-KD condition also induced a greater number of differentially expressed genes (DEGs) within both deep layer and radial glia cells (Extended Data Fig. 9a and Supplementary Table 30). Interestingly, 2×-KD increased the expression of genes involved in synaptic function in both deep layer neurons and radial glia cells. The ability of 2×-KD to regulate and significantly increase radial glia populations was further validated via immunostaining of the outer radial glia marker HOPX in our chimeroids (Fig. 8d). These findings suggest that depletion of both FEZF2 and TSHZ3 might impede the proper differentiation of radial glia into deep layer neurons, either by halting radial glia differentiation or by shifting radial glia toward alternative fates. In line with this hypothesis, we observed a cell cluster unique to 2×-KD cells that contains gene signatures that map to a mixture of developmental neuronal subtypes, including an RYR2/ASIC2⁺ deep layer subtype (Extended Data Fig. 9b). Although this population is marked by genes associated with mature neuronal synapses, we note that it has no clear analog in the adult meta-atlas. Together, these findings indicate that FEZF2 and TSHZ3 play a role in deep layer cell type specification in the human via regulation of genes identified in module 20 from our meta-atlas analysis.

To interrogate how FEZF2 and TSHZ3 might function in a gene regulatory network, we profiled the effects of these TFs on chromatin accessibility using single-cell multi-omic profiling of cells obtained from our chimeroid approach (Extended Data Fig. 10a). We confirmed that these samples also displayed clear expression of all three input cell lines; downregulation of *FEZF2*, *TSHZ3* and module 20 gene activity; and depletion of deep layer cells as observed in our transcriptomic experiments (Extended Data Fig. 10b and Supplementary Tables 31 and 32). Interestingly, these experiments demonstrated that TSHZ3 shRNAs had a unique ability to decrease FEZF2 and TSHZ3 promoter and gene body accessibility and the average chromatin accessibility of module 20 genes (Extended Data Fig. 10c). When we looked at each individual module 20 gene, depletion of *TSHZ3*, but not of *FEZF2*, significantly reduced chromatin accessibility for the majority of module members (Extended Data Fig. 10c). At the level of the whole genome, we observed that TSHZ3 continued to display a greater capacity to regulate chromatin accessibility than FEZF2: TSHZ3 is required for the chromatin accessibility of over 1,000 genes, and these targets are enriched for genes related to synaptic activity (Extended Data Fig. 10d and Supplementary Table 33). Coupled together, our data suggest that the combination of FEZF2 and TSHZ3 mediates the proper progression of radial glia toward a deep layer fate via the activation of module 20 genes, with FEZF2 acting as a transcription activator and TSHZ3 mediating chromatin accessibility (Fig. 8e).

Discussion

Our findings demonstrate how our integrative meta-atlas strategy can illuminate gene expression networks that drive cell fate specification from development to adulthood. Across tissues, but especially in the brain, development unfolds as an orchestrated sequence of cell type diversification, with different cell types, subtypes and developmental

trajectories progressing in parallel. This results in a continuum of cell states, and, although these states have been marked by the expression of TFs and other genes of interest, these cell type markers do not always connect to the nuanced biological processes that drive cell fate specification. Here we provide a searchable resource of 225 meta-modules that characterize stages of peak neurogenesis in the developing human brain (Supplementary Tables 5 and 8–10; UCSC Genome Browser: <https://dev-ctx-meta-atlas.cells.ucsc.edu>) and 299 modules that describe cell types in the adult human cortex (Supplementary Table 16; UCSC Genome Browser: <https://cells-test.gi.ucsc.edu/?ds=adult-ctx-meta-atlas>). These two meta-atlas and their corresponding meta-modules highlight how transcriptional programs that characterize cell subtypes during development diverge in the adult brain, consistent with previous reports of expansive transcriptional rewiring during development during the third-trimester and childhood periods^{22,48}. Our findings, therefore, provide further support for examining the developmental trajectories of cell subtypes from a perspective grounded in developmental data rather than beginning with the adult data where the transcriptional landscape is substantially altered and divergent from developmental processes.

Our meta-modules facilitate such analyses, including our identification and functional interrogation of deep layer-associated meta-module 20. Essential principles of the specification of these neuronal subtypes, particularly via the terminal differentiation factor *Fezf2*, have been elucidated through rodent studies, but the role of other TFs or additional gene networks in specifying this subtype identity has not been fully characterized. Interestingly, the relationship among meta-module 20, *Tshz3* and *Fezf2* during subtype specification was not precisely preserved in the mouse, with both meta-module 20 and *Tshz3* being more broadly expressed in mouse data than in human. By taking advantage of recent chimeroid models of the developing human brain, our data suggest a nuanced relationship among these factors, wherein FEZF2 might activate expression of module 20 genes in part via the ability of TSHZ3 to open the chromatin regions of module 20 members and mediate deep layer specification. Interestingly, this also includes the gene accessibility of *FEZF2*, although the impact on *FEZF2* gene expression remains limited and suggests that this accessibility is counteracted by other forces that lead to the downregulation of FEZF2 at later developmental stages. Although the transcriptional targets of TSHZ3 are not well described, TSHZ3 has been reported to complex with chromatin remodelers and act as a transcriptional repressor in the context of muscle stem cells⁸² and in adult rat cortical neurons⁸³. This suggests that meta-atlas analyses like the strategy used here can link genes such as *FEZF2* and *TSHZ3* that have not been closely functionally linked in human development, revealing nuanced relationships with potential implications in vulnerability to neurodevelopmental disorders.

These data, as well as our immunostaining results validating the spatiotemporal co-expression of other meta-module members, give us confidence in the validity and biological insight offered by our meta-module set. Our meta-module pipeline, therefore, provides an orthogonal method that can be coupled with other existing gene co-expression algorithms to optimally discover novel biology. These meta-modules not only provide a method of linking how adult cortical subtypes might be initiated and shaped in the developing human cortex but also provide a foundation for further functional interrogation of these meta-modules that will yield mechanistic insights.

The meta-atlas described here is, therefore, a fruitful resource for the field, and, to enable widespread access and interaction, we generated cell browsers⁸⁴ that integrate both baseline gene expression and meta-module activity in a searchable interface. Moreover, our meta-atlas strategy is versatile—amenable to the generation of larger meta-atlas that can incorporate future datasets from consortia or to the generation of meta-atlas for a variety of other complex biological systems.

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/s41593-025-01933-2>.

References

- Defelipe, J. The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Front. Neuroanat.* **5**, 29 (2011).
- Sakai, T. et al. Fetal brain development in chimpanzees versus humans. *Curr. Biol.* **22**, R791–R792 (2012).
- Pattabiraman, K., Muchnik, S. K. & Sestan, N. The evolution of the human brain and disease susceptibility. *Curr. Opin. Genet. Dev.* **65**, 91–97 (2020).
- Eze, U. C., Bhaduri, A., Haeussler, M., Nowakowski, T. J. & Kriegstein, A. R. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat. Neurosci.* **24**, 584–594 (2021).
- Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).
- Polioudakis, D. et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron* **103**, 785–801 (2019).
- Network, B. I. C. C. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).
- Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
- Bhaduri, A. et al. An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).
- Braun, E. et al. Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, eadf1226 (2023).
- Cameron, D. et al. Single-nuclei RNA sequencing of 5 regions of the human prenatal brain implicates developing neuron populations in genetic risk for schizophrenia. *Biol. Psychiatry* **93**, 157–166 (2023).
- Fan, X. et al. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* **28**, 730–745 (2018).
- Fan, X. et al. Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci. Adv.* **6**, eaaz2978 (2020).
- Herring, C. A. et al. Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428–4447 (2022).
- Ramos, S. I. et al. An atlas of late prenatal human neurodevelopment resolved by single-nucleus transcriptomics. *Nat. Commun.* **13**, 7671 (2022).
- Smith, R. S. et al. Early role for a Na⁺,K⁺-ATPase (ATP1A3) in brain development. *Proc. Natl Acad. Sci. USA* **118**, e2023333118 (2021).
- Velmeshev, D. et al. Single-cell analysis of prenatal and postnatal human cortical development. *Science* **382**, eadf0834 (2023).
- Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
- Ziffra, R. S. et al. Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).
- Gandal, M. J. et al. Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
- Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
- Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
- Parikshak, N. N. et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
- Cha, J. & Lee, I. Single-cell network biology for resolving cellular heterogeneity in human diseases. *Exp. Mol. Med.* **52**, 1798–1808 (2020).
- Anton-Bolanos, N. et al. Brain chimeroids reveal individual susceptibility to neurotoxic triggers. *Nature* **631**, 142–149 (2024).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
- Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
- Hie, B., Cho, H., DeMeo, B., Bryson, B. & Berger, B. Geometric sketching compactly summarizes the single-cell transcriptomic landscape. *Cell Syst.* **8**, 483–493 (2019).
- Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
- Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
- Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
- Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
- Xie, Z. et al. Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
- Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
- Caglayan, E. et al. Molecular features driving cellular complexity of human brain evolution. *Nature* **620**, 145–153 (2023).
- Jorstad, N. L. et al. Transcriptomic cytoarchitecture reveals principles of human neocortex organization. *Science* **382**, eadf6812 (2023).
- Lake, B. B. et al. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 6031 (2017).
- Ma, S. et al. Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257 (2022).
- Mathys, H. et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385 (2023).
- Morabito, S. et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).
- Nagy, C. et al. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nat. Neurosci.* **23**, 771–781 (2020).
- Pfisterer, U. et al. Identification of epilepsy-associated neuronal subtypes and gene expression underlying epileptogenesis. *Nat. Commun.* **11**, 5038 (2020).
- Schirmer, L. et al. Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature* **573**, 75–82 (2019).
- Siletti, K. et al. Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
- Zhu, K. et al. Multi-omic profiling of the developing human cerebral cortex at the single-cell level. *Sci. Adv.* **9**, eadg3754 (2023).
- Delgado, R. N. et al. Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* **601**, 397–403 (2022).
- Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, eaat8077 (2018).

48. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
49. Takeuchi, A. et al. Identification of Qk as a glial precursor cell marker that governs the fate specification of neural stem cells to a glial cell lineage. *Stem Cell Rep.* **15**, 883–897 (2020).
50. Baumert, R. et al. Novel phospho-switch function of delta-catenin in dendrite development. *J. Cell Biol.* **219**, e201909166 (2020).
51. Herrick, S., Evers, D. M., Lee, J. Y., Udagawa, N. & Pak, D. T. Postsynaptic PDLIM5/Enigma Homolog binds SPAR and causes dendritic spine shrinkage. *Mol. Cell. Neurosci.* **43**, 188–200 (2010).
52. Horiuchi, Y. et al. A polymorphism in the PDLIM5 gene associated with gene expression and schizophrenia. *Biol. Psychiatry* **59**, 434–439 (2006).
53. Horiuchi, Y. et al. Experimental evidence for the involvement of PDLIM5 in mood disorders in hetero knockout mice. *PLoS ONE* **8**, e59320 (2013).
54. Kato, T. et al. Gene expression and association analyses of *LIM (PDLIM5)* in bipolar disorder and schizophrenia. *Mol. Psychiatry* **10**, 1045–1055 (2005).
55. Buttgeriet, A. et al. Sall1 is a transcriptional regulator defining microglia identity and function. *Nat. Immunol.* **17**, 1397–1406 (2016).
56. Weng, Q. et al. Single-cell transcriptomics uncovers glial progenitor diversity and cell fate determinants during development and gliomagenesis. *Cell Stem Cell* **24**, 707–723 (2019).
57. Hoffman, G. E., Smith, M. S. & Verbalis, J. G. c-Fos and related immediate early gene products as markers of activity in neuro-endocrine systems. *Front. Neuroendocrinol.* **14**, 173–213 (1993).
58. Freeman, M. R. & Rowitch, D. H. Evolving concepts of gliogenesis: a look way back and ahead to the next 25 years. *Neuron* **80**, 613–623 (2013).
59. Gunn, T. M. et al. Identification and preliminary characterization of mouse *Adam33*. *BMC Genet.* **3**, 2 (2002).
60. Ushakov, V. S. et al. Heparan sulfate biosynthetic system is inhibited in human glioma due to *EXT1/2* and *HS6ST1/2* down-regulation. *Int. J. Mol. Sci.* **18**, 2301 (2017).
61. Paganini, L. et al. A *HS6ST2* gene variant associated with X-linked intellectual disability and severe myopia in two male twins. *Clin. Genet.* **95**, 368–374 (2019).
62. DeFelipe, J. et al. New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat. Rev. Neurosci.* **14**, 202–216 (2013).
63. Georgiou, C. et al. A subpopulation of cortical VIP-expressing interneurons with highly dynamic spines. *Commun. Biol.* **5**, 352 (2022).
64. Schuman, B. et al. Four unique interneuron populations reside in neocortical layer 1. *J. Neurosci.* **39**, 125–139 (2019).
65. Lin, H. C. et al. Promyelocytic leukemia zinc finger is involved in the formation of deep layer cortical neurons. *J. Biomed. Sci.* **26**, 30 (2019).
66. Usui, N. et al. Zbtb16 regulates social cognitive behaviors and neocortical development. *Transl. Psychiatry* **11**, 242 (2021).
67. Chen, B., Schaevitz, L. R. & McConnell, S. K. Fezl regulates the differentiation and axon targeting of layer 5 subcortical projection neurons in cerebral cortex. *Proc. Natl Acad. Sci. USA* **102**, 17184–17189 (2005).
68. Chen, J. G., Rasin, M. R., Kwan, K. Y. & Sestan, N. Zfp312 is required for subcortical axonal projections and dendritic morphology of deep-layer pyramidal neurons of the cerebral cortex. *Proc. Natl Acad. Sci. USA* **102**, 17792–17797 (2005).
69. Molyneaux, B. J., Arlotta, P., Hirata, T., Hibi, M. & Macklis, J. D. Fezl is required for the birth and specification of corticospinal motor neurons. *Neuron* **47**, 817–831 (2005).
70. Ip, B. K., Bayatti, N., Howard, N. J., Lindsay, S. & Clowry, G. J. The corticofugal neuron-associated genes *ROBO1*, *SRGAP1*, and *CTIP2* exhibit an anterior to posterior gradient of expression in early fetal human neocortex development. *Cereb. Cortex* **21**, 1395–1407 (2011).
71. Johnson, M. B. et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009).
72. Pletikos, M. et al. Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* **81**, 321–332 (2014).
73. Di Bella, D. J. et al. Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554–559 (2021).
74. Yao, Z. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241 (2021).
75. Lodato, S. et al. Gene co-regulation by *Fezf2* selects neurotransmitter identity and connectivity of corticospinal neurons. *Nat. Neurosci.* **17**, 1046–1054 (2014).
76. Caubit, X. et al. *TSHZ3* deletion causes an autism syndrome and defects in cortical projection neurons. *Nat. Genet.* **48**, 1359–1369 (2016).
77. Hevner, R. F. Layer-specific markers as probes for neuron type identity in human neocortex and malformations of cortical development. *J. Neuropathol. Exp. Neurol.* **66**, 101–109 (2007).
78. Arlotta, P. & Pasca, S. P. Cell diversity in the human cerebral cortex: from the embryo to brain organoids. *Curr. Opin. Neurobiol.* **56**, 194–198 (2019).
79. Bhaduri, A. et al. Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).
80. Uzquiano, A. et al. Proper acquisition of cell class identity in organoids allows definition of fate specification programs of the human cerebral cortex. *Cell* **185**, 3770–3788 (2022).
81. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
82. Faralli, H. et al. Teashirt-3, a novel regulator of muscle differentiation, associates with BRG1-associated factor 57 (BAF57) to inhibit myogenin gene expression. *J. Biol. Chem.* **286**, 23498–23510 (2011).
83. Kajiwara, Y. et al. FE65 binds Teashirt, inhibiting expression of the primate-specific caspase-4. *PLoS ONE* **4**, e5071 (2009).
84. Speir, M. L. et al. UCSC Cell Browser: visualize your single-cell data. *Bioinformatics* **37**, 4578–4580 (2021).
85. Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).

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) 2025

Methods

Ethics statement

All brain samples and induced pluripotent stem cell (iPSC) lines collected for analysis in this study were obtained with consent and donor anonymization. All of the procedures were reviewed and approved by the University of California, Los Angeles (UCLA) Institutional Biosafety Committee (IBC) and Human Pluripotent Stem Cell Research Oversight Committee (hPSCRO) (UCLA IBC approval number BUA-2020-091-025-A; UCLA hPSCRO approval number 2020-008-05).

Acquiring datasets and quality control

All R analyses for this study, including meta-atlas generation, were conducted in R versions 4.1.0 or 4.3.1.

Developmental meta-atlas. For all datasets but Polioudakis et al.⁶, Bhaduri et al.⁹ and Nowakowski et al.⁵, gene count matrices for individual cells and corresponding metadata were downloaded from the Gene Expression Omnibus (GEO), and cells from the same individual were combined. Polioudakis et al.⁶ data were downloaded from the browser within the publication (see below), and Bhaduri et al.⁹ and Nowakowski et al.⁵ data were received via personal communication but could also be downloaded from the UCSC Cell Browser⁸⁴. In the case of Smith et al.¹⁶, directories containing the 10x-provided matrix, *mtx*, *genes.tsv* and *barcodes.tsv* were downloaded from <https://figshare.com/s/64b648891e4817efb123> as originally described¹⁶. We then processed the 10x-derived count matrices and directories from each individual with a standard pipeline to generate Seurat objects (Seurat version 4 (ref. 26)), conducting a normalization of the counts as needed and filtering out cells with fewer than 500 genes detected and more than 5% of unique molecular identifiers (UMIs) mapping to mitochondrial genes. Genes detected in fewer than three cells were omitted. In cases where the original publication used more stringent criteria for these quality control measures, we defaulted to the original publication's settings (Supplementary Table 1). In the case of Polioudakis et al.⁶, after downloading raw counts (UMI) gene expression matrix and metadata from the original paper's indicated data browser, we adapted the quality control measures described by the authors, selecting for cells with the 95th percentile of total UMIs with fewer than 250 UMIs mapping to the mouse genome and less than 5% of UMIs mapping to mitochondrial genes. For consistency with the remainder of our meta-atlas, we then applied a stricter minimum for the number of genes detected per cell (increasing the minimum from 200 in the original paper to 500), and we did not place a maximum on the genes detected per cell, as the biologically relevant range of the number of genes per cell will vary between cell types. We then removed genes detected in fewer than three cells.

We found that the size of several individual Seurat objects (Supplementary Table 2) was computationally prohibitive. Thus, we subset each of these objects into two Seurat objects, with cells from the individual randomly distributed. In contrast, all 48 individuals in Nowakowski et al.⁵ were pooled into one Seurat object given its relatively small size.

Adult meta-atlas. With the exception of datasets for Gandal et al.²⁰, Ramos et al.¹⁵ and Schirmer et al.⁴³, Seurat objects or gene counts matrices for individual cells and corresponding metadata for each dataset were downloaded from the GEO or as otherwise instructed in the original publication (Supplementary Table 13). Seurat objects for the Gandal et al.²⁰, Ramos et al.¹⁵ and Schirmer et al.⁴³ datasets were obtained via personal communication from M. Gandal, S. Ramos and A. Zulji, respectively. All datasets for the adult meta-atlas are single-nuclei profiles.

Seurat objects for each individual were generated as described for the developmental meta-atlas (Seurat version 5 (ref. 86)), with the exception that the additional filtering based on the percent of mitochondrial genes per cell was applied to only a few datasets given the

relatively low mitochondrial RNA detected by single-nuclei sequencing (Supplementary Table 13). In the case of Siletti et al.⁴⁴, area-specific Seurat objects that corresponded to cortical areas were downloaded, and additional modifications were required for computation feasibility. For two individuals (H19.30.001 and H19.30.002), individual-level Seurat objects were generated by first extracting the counts matrix for each individual from each area-specific Seurat object. Features in the resulting matrices were converting Ensembl IDs to Gene Symbols, and the feature counts were summarized across genes to collapse data from different transcripts of the same gene. Seurat objects were generated from the resulting individual-level, area-specific counts matrix, and the five area-specific Seurat objects were merged into one object representing a single individual. The features in the resulting Seurat objects were filtered to retain just the genes that overlapped with H18.30.001 in the Bakken et al.³⁴ dataset. These objects were then randomly subset into eight objects. The H18.30.002 individual was subset by first randomly grouping the cell barcodes belonging to this individual into 11 subsets, which were then used as a guide to generate Seurat objects as detailed above. Features from the counts matrices of the Zhu et al.⁴⁵ datasets were also converted from Ensembl ID to Gene Symbols, and counts matrices were filtered to retain just the features that overlap with the Bakken et al.³⁴ dataset for consistency with the remainder of the meta-atlas.

Batch-corrected integration

We constructed our meta-atlases by integrating the indicated datasets (Supplementary Tables 1 and 13) using conventional Seurat pipelines (developmental meta-atlas, version 4; adult meta-atlas, version 5) with modifications to accommodate the size of the meta-atlases.

Developmental meta-atlas. Similarities between cells from different individual Seurat objects were identified and used as anchors to harmonize the data between the various individual Seurat objects; we used the reciprocal PCA method using a *k* of 20. To cluster the 599,211 cells in the resulting integrated meta-atlas, we implemented conventional Seurat version 4 pipelines. The *FindVariableFeatures* and *ScaleData* functions were used to prepare the data for PCA (*RunPCA* function), and significant principal components were identified using methods described in Shekhar et al.⁸⁷. These were used to run a graph-based clustering approach using the *FindNeighbors* and *FindClusters* functions, generating 69 clusters. We then identified cluster markers using differential gene expression analysis (*FindAllMarkers*, reporting only genes positively enriched in a cluster). For each cluster marker, we calculated a gene score metric^{9,79} that quantifies the enrichment and specificity of a gene to a given cluster with the following equation:

Gene score of gene A in cluster 1 = ((% of cells in cluster 1 expressing gene A) / (% of cells not in cluster 1 expressing gene A)) × log₂ fold change of average gene A expression between cluster 1 cell versus all other cells in the dataset

As in previous work^{9,79}, this gene score metric enables comparative analysis within the dataset while leveraging both fold change and specificity, especially when considering differences between cell types, which is a unique advantage of single-cell analysis not encapsulated by the average fold change alone. This approach is particularly useful for overcoming batch effects due to divergent technologies or sequencing depths—an important benefit joint analysis of datasets without uniform pre-processing.

Clusters with similar marker gene profiles are indicative of clusters with high biological similarity and could be grouped together for downstream analyses. To determine which clusters can be merged, we ran a self-correlation of the clusters based on cluster marker gene scores, combining clusters with an *R*² > 0.7. This generated 50 clusters, for which we calculated cluster markers as described above. Based on these markers, 48 clusters were assigned cell class, state, type and subtype identities, and two clusters were designated as outliers (Supplementary

Table 3). UMAP coordinates for this integrated dataset were then calculated using the RunUMAP function.

Adult meta-atlas. Integration of individuals for the adult meta-atlas took advantage of the sketch integration pipeline described for Seurat version 5. In brief, a random sampling of 20% of each individual was conducted to obtain a dataset of 520,013 cells, matching the scale of the developmental meta-atlas. After normalization, this object was split into layers centered by dataset, and FindVariableFeatures was applied to each layer. The SketchData function was then performed (ncells = 1,000, method = 'LeverageScore'), and the resulting sketch of the dataset was used to perform integration and dimensionality reduction as described for the developmental meta-atlas above. This results in a low-dimensional space, upon which the remainder of the 520,013 adult meta-atlas cells can be projected (ProjectIntegration, ProjectData and RunUMAP functions, dims = 1:30). To annotate cell types, we leveraged the label transfer pipeline described for Seurat version 5 (FindTransferAnchors and MapQuery functions), using the Jorstad et al.³⁶ dataset due to its hierarchical layer-specific and cell-type-specific annotations.

Comparison of integration methods

For computational feasibility, a random subset of 50,000 cells was used to benchmark across four different integration methods. Using Seurat version 5, this subset was split by dataset, and normalization and PCA were conducted on a per-dataset basis. PCA was run on this merged dataset before batch correction, first by identifying variable features (vst, nfeatures = 2,000) and scaling the data based on variable features. All downstream methods were conducted using 15 principal components, as calculated using methods described in Shekhar et al.³⁶. Integrations were conducted using the IntegrateLayers function in Seurat version 5 via reciprocal PCA (runtime of 106 s), Fast MNN (59 s) and Harmony (80 s) algorithms. For each of the resulting reductions, graph-based clustering approaches (FindNeighbors, FindClusters and RunUMAP functions) were used to cluster and visualize cells on UMAP space. Scanorama integration was performed using the 'reticulate' package (1.36.1) to import the Scanorama Python package⁸⁸ (python3 3.12.3).

Lineage analysis

To analyze cellular trajectories and infer pseudotime in our developmental meta-atlas, we used the Monocle 3 (ref. 30) (version 1.3.4) toolkit following initial clustering, dimensionality reduction and quality control steps (filtering out cells with low gene counts and high mitochondrial content), using Seurat version 4 in R version 4.3.2. The object was subset to exclude the small percentage of outlier cells as required for the Monocle 3 pipeline. Progenitor subtypes tRG and oRG.2 were selected as root cells for trajectory analysis, and cell trajectories were inferred by using the principal graph method implemented in Monocle 3. Cells were ordered in pseudotime along the constructed trajectories, and branch points were identified to explore lineage decisions. The trajectory data, including pseudotime values and principal graph projections, were extracted and exported for further analysis.

As an alternative method to infer cellular trajectories and pseudotime from our single-cell RNA sequencing data, we employed the Slingshot toolkit³¹ (version 2.7.0). The analysis was performed on a 2,999-cell subset of the developmental meta-atlas for computational feasibility. Slingshot automatically fit a principal curve through the data points, reflecting the developmental progression of cells along the inferred trajectory.

Meta-module generation

In parallel with this conventional integration analysis, we also performed a meta-module analysis pipeline to extract networks of genes that share expression patterns throughout the entire meta-atlas. These

analyses were adapted from clustering methods as in the weighted gene co-expression network analysis (WGCNA) pipeline and used the following packages: WGCNA²¹, factoextra, dynamicTreeCut and data.table.

Identification of individual-level gene signatures. We first identified the gene networks representative of each individual, conducting a clustering analysis for each of the Seurat objects generated for each individual.

For each individual Seurat object, we first performed a self-correlation of the object's normalized gene expression matrix, grouping cells based on their transcriptome. This analysis generated a correlation-based distance matrix, which we clustered to assign cells into specific clusters. We used hierarchical clustering for this step, because it provides a dendrogram of cells that allows for adaptive clustering depending on the biological question of interest. For example, although most individual Seurat objects were clustered at the highest resolution (hclust deep split option set at 4), Seurat objects generated from the Smith et al.¹⁶ and Bhaduri et al.⁹ datasets were significantly larger and clustered at lower resolution (hclust deep split option set at 3) (Supplementary Table 2) to avoid drastic changes in the number of clusters between these individuals. Our analyses can withstand the computational demand of hierarchical clustering due, in large part, to the relatively fewer cells per individual, and this approach worked well for both the developmental meta-atlas (~600,000 cells) as well as the substantially larger adult meta-atlas (~2.6 million cells). We note that, in cases of large numbers of cells within individuals, graph-based clustering of cells within each individual could also be applied.

The resulting cluster assignments for each individual were added as metadata in its accompanying Seurat object, and conventional Seurat commands were used to identify markers of these clusters. Gene scores for these cluster markers were calculated using the equation described above.

Generation of meta-modules from individual cluster markers.

From these individual gene signatures, we extracted modules of genes with shared expression patterns across the entire meta-atlas. We first aggregated the cluster markers of all the individuals in our meta-atlas and then retained cluster markers with the 90th percentile of gene scores across the entire meta-atlas. The resulting table—containing genes, the meta-atlas clusters for which these genes have been identified as cluster markers and their corresponding gene scores—was used to generate a distance matrix. Hierarchical clustering of this matrix was used to group genes based on their gene scores across all clusters in the meta-atlas. For this clustering, the dendrogram cut was set specifically at 1 to generate minimal module granularity, thereby generating meta-modules consisting of genes with similar expression patterns across all individuals represented in each meta-atlas (Supplementary Table 5).

Module activity scoring, binarization and visualization

We determined meta-module activity within each cell by devising a module activity score based on the average expression level of each gene in a meta-module. Specifically, activity in the cell was scored by first taking the sum of all normalized counts per million (CPM) detected for each meta-module gene and then dividing by the total number of meta-module genes to reduce bias toward modules containing large numbers of genes. We set the minimum activity score for a module to be considered active in a cell as the 90th percentile of all meta-module activity scores across the entire dataset, which we calculated for each of the seven datasets in our meta-atlas to account for variations in sequencing depth and other technical factors. This threshold enabled cells to be assigned as positive or negative for a given module.

Module activity scores and module positivity assignments for each cell were added to Seurat objects as metadata, enabling the visualization of these metrics using the FeaturePlot and DimPlot functions from

Seurat version 4. Heat maps (morpheus R package version 1.0-4) and bar charts (ggplot2 R package version 3.4.3) displaying the enrichment of cell types among cells positive for a given module were constructed by first determining the distribution of cell types among module-positive cells, which was then normalized to the distribution of cell types across the entire meta-atlas. The ggpubr package (version 0.6.0, <https://rpkgs.datanovia.com/ggpubr/>) was used to determine significance using two-sided Welch's *t*-test where indicated.

These scores proved to be highly versatile metrics that allowed for the calculation of meta-module activity in other datasets outside of our meta-atlas and can be used to measure the activity of a wide variety of gene networks.

Module specificity score

To quantify the ability of modules to represent specific cell subtypes, we developed a module specificity score analogous to our previously published gene score metric used to evaluate cluster markers. By quantitatively linking a module to its cell subtype specificity in a given dataset, the module specificity score facilitates the comparison of cell-subtype-specific module activity patterns across comparison datasets. Module specificity scores measure the relative activity of the module in a given cell subtype as well as the enrichment of a given cell subtype among module-positive cells, as follows: module specificity score of module A in cell subtype 1 = ((% of cells in cell subtype 1 with module activity above the 90th percentile) / (% of cells not in cell subtype 1 with module activity above the 90th percentile)) $\times \log_2$ fold change of average module A expression between cell subtype 1 versus all other cell subtypes in the dataset.

As in gene score calculations, negative module specificity scores indicate a de-enrichment of this module at the indicated cell subtypes (that is, the module is less active in the given cell subtype relative to all other cells). Similarly, the lack of any module-positive cells within a cell subtype (or, at times, throughout the whole dataset) will result in a module specificity score of NaN or 0. In cases where there was no overlap between meta-module genes and features detected in the comparison dataset, cell subtypes received a score of NaN for that module, which was converted to 0 for module specificity calculations. For visualization purposes, where we are interested primarily in modules that are enriched in a cell subtype, we, therefore, set a minimum of 0.

Module annotation

The enrichR R package (version 3.2 (ref. 33)) was used to aid in the biological annotation of our 225 developmental meta-modules by surveying term databases focused on signaling pathways (WikiPathway 2021 Human, KEGG 2021 Human and Elsevier Pathway Collection), transcriptional regulation (ChEA 2016, ENCODE and ChEA Consensus TFs from ChIP, TF Perturbations Followed by Expression and TRRUST Transcription Factors 2019) and Gene Ontology sets (GO Biological Process 2021, GO Molecular Function 2021 and GO Cellular Component 2021). Our annotations began with a survey of Gene Ontology terms enriched in each meta-module, with the greatest weight placed on enriched terms with an adjusted *P* < 0.01 (Supplementary Table 8). Extensive literature review was used as necessary to assign the most specific biological associations to each module (Supplementary Table 9).

Acquisition and analysis of comparison datasets

To analyze the single-nuclei transcriptomic dataset of adult human cortex, gene expression matrices, metadata and t-distributed stochastic neighbor embedding (t-SNE) coordinates were downloaded from the Allen Brain Cell Types Database (RRID: SCR_017266; <https://biccn.org/data>). This gene expression matrix was first filtered to only include nuclei that had passed relevant quality control metrics to merit inclusion in the pre-calculated t-SNE plot, resulting in 47,432 nuclei. The filtered matrix was then used to create a Seurat object, normalizing the counts and omitting genes detected in fewer than three cells and cells

with fewer than 500 genes detected. An updated version of this dataset (Jorstad et al.³⁶) was also included in our adult meta-atlas.

Gene expression matrices, metadata and UMAP coordinates for the developing mouse cortex profile were downloaded from the Broad Single Cell Portal, as referenced in Di Bella et al.⁷³. The gene expression matrix was filtered to only include cells for which meta-data were provided (49,024 cells) before Seurat object generation.

For the adult mouse cortex, gene expression matrices and meta-data were downloaded from the NeMO Archive for the BRAIN Initiative Cell Census Network (ID: nemo:dat-v8hcy). Given the computational cost of processing of all 1.1 million cells in this dataset, a random sampling of 100,000 cells from the data was performed using a Python script from an accompanying dataset described in the same paper (ID: nemo:dat-iy7gkp)⁷⁴. A Seurat object was generated from this subset, omitting genes detected in fewer than three cells. Cells from non-cortical areas were removed, generating 73,761 cells, which were then processed with standard pipelines for quality control, normalization and UMAP coordinate generation as described above.

For all comparison datasets, module activity scores and enrichment of cell types among module-positive cells were calculated as described above, with meta-module genes not present in the comparison dataset omitted from analysis. In cases where there was no overlap between meta-module genes and features detected in the comparison dataset, cells received an activity score of NaN for that module.

Gene co-expression analysis using WGCNA and SCENIC

WGCNA modules were generated from a randomly sampled 10,000-cell subset of the developmental meta-atlas for computational feasibility, using WGCNA (version 1.72.5 on R 4.3.1) as previously described⁷⁹. This same subset was also used for gene regulatory network reconstruction with pySCENIC. Auxiliary datasets, including hg38 regulatory feature rankings (feather files) and motif-to-TF annotations data, were obtained from the cistargetDB website (<https://resources.aertslab.org/cistarget>). pySCENIC analysis (version 0.12.1)^{32,89} followed standard workflow from the developers. In brief, regulatory interactions between TFs and potential target genes were inferred from the expression matrix with SCENIC-implemented GRNBoost2 to generate adjacency matrix, subsequently generating SCENIC modules. Next, motif enrichment and TF-regulon prediction were performed using SCENIC-implemented cisTarget.

Immunofluorescent validation of module activity in primary cortex

Brain dissections from GW 16 and GW 20 samples were obtained and flash frozen in optimal cutting temperature (OCT) as described in Bhaduri et al.⁹. Cryosections (16 μ m) were mounted onto treated glass slides, where they were fixed for 2 min with 4% paraformaldehyde and then washed 3 \times 10 min with PBS. Immunofluorescent staining for ZBTB16, VGAT and CCK and for QKI and SALL1 required an additional antigen retrieval step, in which sections were incubated with boiling citrate-based antigen retrieval solution (Vector Laboratories, H-3300-250) for 20 min and then washed with PBS. All slides were incubated in blocking buffer (3% BSA, 5% donkey serum and 0.1% Triton) for 30 min and then incubated overnight at 4 °C with the indicated primary antibodies diluted in blocking buffer. After a 3 \times 10-min wash in PBST (0.1% Triton in PBS), slides were incubated in the dark at room temperature for 2 h with 1:1,000 DAPI and 1:500 Alexa Fluor secondary antibodies (Thermo Fisher Scientific) diluted in blocking buffer. Primary antibodies included: FEZF2 (1:500; Abcam, ab214186), TSHZ3 (1:1500, gift from L.F.), CTIP2 (1:500; Abcam, ab18465), SATB2 (1:500; Abcam, ab51502), HS6ST2 (1:200; Abcam, ab122220), ADAM33 (1:250; Santa Cruz Biotechnology, sc-514055), ZBTB16/PLZF (1:250; Santa Cruz Biotechnology, sc-28319), VGAT (1:1,000; Millipore Sigma, AB5062P), CCK (1:1,000; Millipore Sigma, C2581), QKI (1:250; Millipore Sigma, MABN624MI), PDLIM5/ENH (1:250; Thermo Fisher Scientific, 388800),

SALL1 (1:400; Thermo Fisher Scientific, PA5-62057), c-FOS (1:400; Santa Cruz Biotechnology, sc-166940) and ZFP36L1/2/BRF1 (1:250; Cell Signaling Technology, 12306-1-AP). Alexa Fluor secondary antibodies from Thermo Fisher Scientific used included: mouse (488: A-31570, A-11001; 555: A-31572, A-11005; 568: A-31573, A-11004; 647: A-31574, A-11029); rat (488: A-21208, A-11007; 555: A-21209, A-11009; 568: A-21210, A-11008; 647: A-21211, A-11037); guinea pig (488: A-11073, A-11075; 555: A-21449, A-21450; 568: A-11075, A-11073; 647: A-21450, A-21449); rabbit (488: A-11008, A-11034; 555: A-21428, A-21207; 568: A-11011, A-21245; 647: A-21245, A-21207); and goat (488: A-11055, A-11034; 555: A-21432, A-11039; 568: A-11056, A-11040; 647: A-21435, A-11041).

Stained sections were imaged using a Zeiss LSM 880 with $\times 20$ objective, with the exception of c-FOS and ZFP36L1/2 stains, which were imaged using an EVOS M5000 with $\times 10$ objective. Gain, pinhole, laser power and other acquisition parameters were left constant for both GW 16 and GW 20 samples within the same staining panel. Maximum-intensity z-stack projections and tilescans were created using ZEN Blue (Zeiss). Fluorescence intensities were adjusted using FIJI software (RRID: SCR_002285) specifically for display purposes and not for quantification. Quantification was performed in Imaris version 10 by quantifying the number of cells or voxels positive for a certain marker using standardized thresholding based on fluorescence intensity. Numbers of marker-positive or co-localized cells/voxels were normalized to the number of DAPI⁺ spots for each region of interest. Subregions (that is, CP, IZ and VZ) were determined by a combination of DAPI density and layer-specific stains in parallel sections of cortical tissue.

Generation of lentiviral shRNA expression vectors

Lentiviral shRNA expression constructs were generated by first modifying the pLKO.1 TRC cloning vector (Addgene, 10878), digesting with BamHI and KpnI and replacing the puromycin resistance cassette with an mCherry coding sequence.

To clone each shRNA construct, the following oligos were ordered from Integrated DNA Technologies:

Forward:

5'-CCGG—target sequence (sense)—CTCGAG—target sequence (antisense)—TTTTTG-3'

Reverse:

5'-AATTCAAAA—target sequence (sense)—CTCGAG—target sequence (antisense)-3'

Oligos were ordered for each of the following target sequences: scrambled (scram) (5'-CCTAAGGTAAAGTCGCCCTCG-3'), FEZF2 1 (5'-CTGCTCAACATCTGCTCTCCG-3'), FEZF2 2 (5'-AGTGCGCCGAAACGTATTAA-3') and TSHZ3 (5'-CCCTTACATCAGCCAAATAA-3'). Oligos were annealed and then ligated into the pLKO.1-mCherry digested with AgeI and EcoRI.

Versions of these constructs harboring a GFP reporter gene were subsequently made by digesting pLKO.1-mCherry with BamHI and KpnI, followed by a HiFi Assembly (New England Biolabs, E2621) of the backbone with a GFP insert amplified from pLV hUbc-dCas9-T2A-GFP (Addgene, 53191) with Phusion Plus Green PCR Master Mix (Thermo Fisher Scientific, F632) using the following primers:

5'-CTCCCTCGTTGACCGAATCACCACCTCTCTCCCCAGGGGATCCACCGGAGCTTACCATGGTGAGCAAGGGCGAGGAGCT-3'
5'-TCCTTTTCGGTCCGGCGCTGCGGGTCTGTTGGGCGGGCGTTAACCGGTCTGTACAGCTCGT-3'

The PCR product was purified with NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, 740609) following the manufacturer's instructions before HiFi Assembly.

For each of these constructs, lentiviral supernatants were generated by transfecting one 10-cm plate of HEK293T cells at 90% confluency with 3 μ g of pMD2.G (Addgene, 12259), 9 μ g of psPAX2 (Addgene,

12260) and 12 μ g of pLKO.1 transfer vector using Lipofectamine 2000 (Thermo Fisher Scientific, 11668019). Medium was replaced after 24 h with 10% FBS and 1 \times penicillin–streptomycin (Thermo Fisher Scientific, 15140122) in high-glucose DMEM supplemented with GlutaMAX and pyruvate (Thermo Fisher Scientific, 10569010). Supernatants were collected two times at 24-h intervals, passed through a 0.45- μ m filter and concentrated with Lenti-X Concentrator (Takara, 631232). The resulting pellet was resuspended in mTesR+ (STEMCELL Technologies, 100-0276), and the lentiviruses were flash frozen in liquid nitrogen and stored at -80°C before use.

Large-scale production of lentiviral supernatants was generated for chimeroïd experiments by transfecting one T225 flask of HEK293T cells at 90% confluency with 11.25 μ g of pMD2.G (Addgene, 12259), 33.75 μ g of psPAX2 (Addgene, 12260) and 45 μ g of pLKO.1 transfer vector using Lipofectamine 2000 (Thermo Fisher Scientific, 11668019). Medium was replaced, and supernatants were harvested, filtered and stored as described above.

Generation and evaluation of FEZF2/TSHZ3 shRNA organoids

Infection of stem cells with shRNA constructs. Human embryonic stem cells (UCLA6 and UCLA1) and iPSC line CIRM20107 were authenticated at the source. All lines used for this study were between passages 12 and 24. Stem cells were thawed in mTesR+ media and plated onto six-well plates coated with growth-factor-reduced Matrigel (Corning, 354230). Cells were treated overnight with 10 μ M ROCK inhibitor Y-27632 (STEMCELL Technologies, 72304) to attenuate cell death. Stem cells were subsequently maintained in mTesR+, with media changed every 1–3 d. Cells were passaged at about 70% confluency, at which time cells were washed with PBS and incubated in ReleSR (STEMCELL Technologies, 05872) for 1 min at room temperature. The ReleSR was aspirated, and cells were incubated at 37°C for 5 min, after which mTesR+ was added to the well, and cells were manually lifted with cell lifters (Thermo Fisher Scientific, 08-100-240) and expanded onto Matrigel-coated six-well plates.

Organoid generation. At the second passage post-thaw, cells were infected with lentiviral supernatants encoding shRNAs. One approximately 70% confluent six-well well of stem cells was passaged as described above and expanded onto 4–6 six-well wells, each of which contained 5 μ M ROCK inhibitor, 10 μ g ml⁻¹ polybrene (Millipore Sigma, TR-1003-G) and 1:200 lentivirus in mTesR+ media. Media were changed 18 h after infection. To increase the proportion of mCherry-expressing human induced pluripotent stem cell (hiPSC) E cells, infected hiPSC E lines were subjected to a second round of infection in which cells were passaged at a 1:2 ratio. All lines were maintained and expanded with the standard protocols described above, undergoing a maximum of one passage between infection and organoid aggregation.

Organoids were based on protocols described in Kadoshima et al.⁹⁰ and Velasco et al.⁹¹. Stem cells were washed with PBS and dissociated into single-cell suspensions with a 5-min, 37°C incubation in Accutase (Millipore Sigma, A6964). Cells were scraped and pelleted for 5 min at 300g and then immediately resuspended in GMEM (Life Technologies, 11710-035) supplemented with 1 \times penicillin–streptomycin, 15% Knock-Out Serum (Life Technologies, 10828-028), 1 \times NEAA (Life Technologies, 11140050), 1 \times sodium pyruvate (Life Technologies, 11360070), 100 μ M β -mercaptoethanol (Life Technologies, 21985023), 20 μ M ROCK inhibitor Y-27632, 10 μ M SB431542 (Tocris, 1614) and 3 μ M IWRI-e (Cayman Chemical, 13659). Resuspended cells were transferred to 96-well V-bottom low-adhesion plates (S-Bio, MS-9096VZ) at 10,000 cells per well.

Throughout cell culture, media were changed every 2–3 d. For the first 6 d, organoids were also treated with 20 μ M ROCK inhibitor Y-27632. After 18 d, organoids were transferred to six-well ultra-low-adhesion plates (Corning, 07-200-601) and placed onto an orbital shaker rotating at 100 r.p.m. At this time, organoids were also

moved to DMEM/F12-GlutaMAX media (Life Technologies, 10565-018) supplemented with 1× N2 (Life Technologies, 17502-048), 1× CD Lipid Concentrate (Life Technologies, 11905-031) and 1× penicillin–streptomycin. Organoids generated from TSHZ3 shRNA-infected hiPSC E lines did not survive transfer and were discarded from the experiment.

At 35 d, organoid culture media were also supplemented with 1× 10% heat-inactivated FBS (Life Technologies, 10082147), 5 µg ml⁻¹ heparin (Sigma-Aldrich, H3149) and 0.5% Matrigel. Organoids were harvested for qRT–PCR and immunofluorescence analyses after 7 weeks in culture (50–52 d) and harvested for single-cell transcriptomics after 8 weeks (55–57 d).

Validation of *FEZF2* and *TSHZ3* KD with qRT–PCR. RNA was extracted from three organoids from each condition using an RNeasy Plus Mini Kit (Qiagen, 74136), in which organoids were lysed directly in the included lysis buffer supplemented with β-mercaptoethanol. Equivalent amounts of RNA were used in cDNA synthesis reactions performed using SuperScript IV VILO Master Mix (Thermo Fisher Scientific, 11756050), and qRT–PCR was conducted on a QuantStudio 5 (Applied Biosystems) with Power SYBR Green PCR Master Mix (Applied Biosystems, 4368708) and the following primer pairs:

GAPDH:
Forward: 5′-GGAGCGAGATCCCTCCAAAT-3′
Reverse: 5′-GGCTGTTGTCATACTTCTCATGG-3′
FEZF2:
Forward: 5′-AGGTGTGCGGCAAGGTGTTT-3′
Reverse: 5′-ACACGAACGGTCTGGCTCCG-3′
TSHZ3:
Forward: 5′-GGCGCGCAGCAGCCTATGTT-3′
Reverse: 5′-CTTGCCGAGGGCTCTCCAT-3′

Gene expression levels were normalized to *GAPDH* and then normalized to the average *GAPDH*-normalized expression level in scram shRNA-transduced organoids from three lines.

Evaluation of cortical organoid patterning. Organoids were fixed for 45 min in 4% paraformaldehyde, washed with 1× PBS and rehydrated with an incubation in 30% sucrose in PBS until saturated. Organoids were then embedded into cryomolds with a 1:1 ratio of 30% sucrose in PBS and OCT (Tissue-Tek, 14-373-65), and molds were stored frozen at –80 °C. Sections were cut on a cryostat at 20-µm thickness and then mounted onto treated glass slides for immunofluorescent staining with antigen retrieval as described above, using primary antibodies against CTIP2, SOX2 (1:500; Santa Cruz Biotechnology, sc-365823) and VSNL1 (1:100; OriGene, UM870034).

Stained organoids were imaged on an EVOS M5000 (Thermo Fisher Scientific) with ×10 objectives, and fluorescence intensities were later adjusted using FIJI software (RRID: SCR_002285). Image acquisition and adjustment parameters were left constant for all samples within the same staining panel. Quantification of VSNL1 staining was performed using Imaris software (version 10, RRID: SCR_007370) using three images obtained with a ×20 objective for each condition across three biological replicates, with the exception of TSHZ3 shRNA-infected cells (two replicates). In brief, the number of DAPI⁺ nuclei was identified by counting spots exceeding a set threshold of total DAPI intensity. The extent of VSNL1 expression was measured by calculating the total area of surfaces surpassing a minimum intensity of VSNL1. Parameters for object creation, such as voxel size and spot radius, were kept constant for all images, and parameters for object classification (minimum DAPI and VSNL1 intensities) were kept constant for images within the same biological replicate. To compare VSNL1 expression across all images, VSNL1⁺ surface area in an image was normalized to the number of DAPI⁺ nuclei, and significance was calculated with two-sided Welch's *t*-test.

PIP-seq capture and sequencing. At week 8, organoids from shRNA-transduced UCLA 6 and UCLA 1 lines were processed for single-cell transcriptomics using the PIP-seq T2 3′ Single Cell RNA Kit version 4.0 (Fluent Biosciences). Cortical organoids were dissociated into single-cell suspensions via 30-min incubation at 37 °C in papain and DNase in EBSS (Worthington, LK003150), shaking vigorously every 5–10 min. Samples were then triturated with manual pipetting 10 times and pelleted with a 5-min spin at 300g in 4 °C.

In the case of cells from UCLA 6 lines, pellets were resuspended in the provided Cell Resuspension Buffer (Fluent Biosciences), and 5,000 cells were targeted for capture. In contrast, cell pellets from UCLA 1 lines were additionally segregated based on mCherry expression for intra-organoid comparison of cells with versus without shRNA expression. Cells were resuspended in 2% FBS in PBS and then passed through a wet 35-µm cell strainer into round-bottom tubes for flow cytometry–assisted cell sorting (FACS). Samples were treated with DAPI (1:1,000) to identify dead cells and processed on a FACS Aria (BD Biosciences) configured with a 561-nm laser, a 595LP dichroic mirror and a 605/40-nm bandpass filter for mCherry detection and a 355-nm laser, a 450 LP dichroic mirror and a 515/30-nm bandpass filter for DAPI detection. FACS gates were set using uninfected cells, and mCherry[–] and mCherry⁺ sorted samples were pelleted once more and resuspended in 5 µl of 2% FBS in PBS. This process enabled the targeting of approximately 25,000 cells for capture, with the exception of mCherry⁺ cells from FEZF2 1 shRNA-transduced organoids (7,500 cells targeted).

Sequencing libraries were obtained from cells according to PIP-seq T2 instructions, using 16–17 cycles for cDNA amplification and 8–12 cycles for library amplification depending on the amount of DNA. Libraries were quantified on an Agilent 2100 Bioanalyzer and sequenced as per manufacturer recommendations on a NovaSeq S1 (UCLA 1) and S2 (UCLA 6) flow cell (UCSF Genomics Services Core).

Analysis of single-cell transcriptomic data. BCLs were obtained from the UCSF Genomics Services Core, and demultiplexed FASTQ files were generated (bcl2fastq version 2.20.0.422) and aligned to a custom reference containing the human genome (GRCh38), an mCherry sequence and genomes of four common mycoplasma species using PIPseeker software version 02.01.04 (Fluent Biosciences), with sensitivity set to 3 according to manufacturer recommendations. For each sample, count matrices were converted to Seurat objects using the Read10X function, Seurat version 4. Samples from each biological replicate (that is, derived from UCLA 6 or UCLA 1) were merged into one Seurat object; the Seurat pipelines described above were used to perform rigorous quality control (omit genes detected in <3 cells and cells with <500 genes detected and/or >5% of UMIs mapping to mitochondrial genes); and the lack of mycoplasma infection in our dataset (<1% of UMIs in the dataset after quality control mapping to mycoplasma genomes) was confirmed. Using Seurat as detailed above, we clustered cells based on their gene profiles, calculated and scored cluster markers, assigned cell type based on marker expression (Supplementary Tables 13–16) and calculated UMAP coordinates. Cells were also scored for module 20 activity as described above.

Effects of FEZF2 and TSHZ3 shRNAs on expression of *FEZF2*, *TSHZ3* and module 20 activity were evaluated by normalizing the average expression level in each condition to that of scram shRNA-transduced organoids (UCLA 6) or the mCherry[–] scram shRNA-transduced organoids (UCLA 1). The ggpubr package (version 0.6.0, <https://rpkgs.datanovia.com/ggpubr/>) was used to determine significance using two-sided Welch's *t*-test where indicated.

Molecular effects of these shRNAs were assessed by identifying genes positively or negatively enriched in indicated conditions with the FindMarkers function, Seurat version 4. Specifically, in the case of UCLA 6 lines, DEGs were identified between scram-transduced versus FEZF2/TSHZ3 shRNA-transduced organoids in each of the module 20-associated cell types (Deep Layer Neuron, NEFL.ZCCHC12.EN

and GPC6.FTX.EN). For UCLA 1 lines, DEGs were identified between mCherry⁻ versus mCherry⁺ cells from the same shRNA condition among deep layer neurons (Deep.Layer.1 and Deep.Layer.2 subtypes). The enrichR R package³³ (version 3.2) was used to identify Gene Ontology terms enriched among these DEG sets, surveying the term databases described above. Dot plots displaying the most significantly enriched terms for each DEG set were generated using the ggplot2 R package, version 3.4.3.

FEZF2/TSHZ3 KD in cortical chimeroids

Generation and maintenance of cortical chimeroids. Chimeroids were derived from cortical organoids of human embryonic stem cells (UCLA 6 and UCLA 1) and iPSC line KOLF, which were authenticated at the source. All lines used for this study were between passages 7 and 25 and cultured as described above. After a maximum of two passages, each cell line was used to generate and maintain cortical organoids using our standard protocols detailed above (10,000–13,000 cells per well).

At day 19, all organoids were dissociated and re-aggregated to generate cortical chimeroids using protocols based on Anton-Bolanos et al.²⁵. In brief, organoids from each of the three cell lines were dissociated via 30-min incubation at 37 °C in papain and DNase in EBSS (Worthington, LK003150), shaking vigorously every 5–10 min. Samples were then triturated with manual pipetting 10 times and pelleted with a 5-min spin at 300g in 4 °C. Cell pellets were resuspended in GMEM supplemented with 1× penicillin–streptomycin, 15% KnockOut Serum (Life Technologies, 10828-028), 1× NEAA (Life Technologies, 11140050), 1× sodium pyruvate (Life Technologies, 11360070), 100 μM β-mercaptoethanol (Life Technologies, 21985023), 20 μM ROCK inhibitor Y-27632, 10 μM SB431542 (Tocris, 1614) and 3 μM IWRI-e (Cayman Chemical, 13659), mirroring conditions of the initial organoid aggregation to maximize cell health. A mixture of single-cell suspensions was then generated, consisting of 50% of KOLF, 42% of UCLA 6 and 8% of UCLA 1 cells based on cell concentration. Low organoid yield for UCLA 1 cells resulted in the low representation of this line in the final chimeroid mixture, although we note that the single-cell transcriptomic data suggest that these cells recovered to represent 34% of cells in the final chimeroid population (Supplementary Fig. 18c).

The chimeroid single-cell suspension was then transferred to 96-well V-bottom low-adhesion plates (S-Bio, MS-9096VZ) at 40,000 total cells per well, in which cells were also incubated with diluted lentiviral supernatants (1:1,000 of the scram shRNA/mCherry virus; 1:100 of either of the FEZF2 1 shRNA/GFP virus or the TSHZ3 shRNA/mCherry virus; or 1:200 each of the FEZF2 1 shRNA/GFP and the TSHZ3 shRNA/mCherry viruses). FEZF2 shRNA 1 proved to be the most potent of the two FEZF2 shRNAs (Supplementary Figs. 9 and 10) and was, thus, used for chimeroid experiments. After 20 h, 100 μl of media on the re-aggregating chimeroids was replaced with 100 μl of DMEM/F12-GlutaMAX media (Life Technologies, 10565-018) supplemented with 1× N2 (Life Technologies, 17502-048), 1× CD Lipid Concentrate (Life Technologies, 11905-031) and 1× penicillin–streptomycin to resume organoid maturation. Media were changed again the next day, and, at 23 d, organoids were transferred onto six-well ultra-low-adhesion plates (Corning, 07-200-601) and placed onto an orbital shaker rotating at 100 r.p.m.

Beginning at 35 d, organoid culture media were also supplemented with 1× 10% heat-inactivated FBS (Life Technologies, 10082147), 5 μg ml⁻¹ heparin (Sigma-Aldrich, H3149) and 0.5% Matrigel. From day 70 onwards, Matrigel was increased to 1% in the organoid culture media, which were also supplemented with 1× B-27 (Life Technologies, 17504-044).

Evaluation of cortical organoid patterning. Chimeroid health and patterning were assessed by immunostaining at week 5, at which point chimeroids were fixed, embedded and stained as described above

using primary antibodies against CTIP2 (1:500; Abcam, ab18465) and SOX2 (1:500; Santa Cruz Biotechnology, sc-365823). Stained organoids were imaged on an EVOS M5000 (Thermo Fisher Scientific) with ×10 objectives, and fluorescence intensities were later adjusted using Fiji software (RRID: SCR_002285). Image acquisition and adjustment parameters were left constant for all samples within the same biological replicate.

In a similar manner, the effects of FEZF2 and TSHZ3 shRNAs on chimeroid cell type specification were evaluated by immunostaining at week 9 using primary antibodies against HOPX (1:500; Proteintech, 11419-1-AP). Quantification of HOPX staining was performed using Imaris software (version 10, RRID: SCR_007370) using images obtained with a ×10 objective for each condition across three biological replicates as described in the above organoid experiments.

Validation of FEZF2 and TSHZ3 KD with qRT-PCR. Levels of *FEZF2* and *TSHZ3* in cortical chimeroids were assessed by qRT-PCR as described above for the organoid experiments, with the following exceptions. At week 7, three chimeroids from each condition were sorted using flow-assisted cell sorting. To maximize cell numbers for subsequent qRT-PCR, double-negative populations from each sort were pooled into the same collection tube. Target populations (mCherry⁺ from TSHZ3 shRNA conditions and GFP⁺ from FEZF2 shRNA conditions) were isolated, and mCherry⁺ and GFP⁺ cells obtained from the double KD conditions, but displayed a single KD, were also pooled with those from the respective single KD populations.

PIP-seq capture, sequencing and data analysis. Chimeroids were harvested at week 8 for single-cell sequencing using the PIP-seq T2 3' Single Cell RNA Kit version 4.0 PLUS (Fluent Biosciences). Three chimeroids from the single shRNA conditions and 33 chimeroids from the double KD conditions were prepared for flow-assisted cell sorting as described above for the organoid KD experiments. Target populations were sorted on a FACS Aria (BD Biosciences) configured with a 561-nm laser, a 595LP dichroic mirror and a 605/40-nm bandpass filter for mCherry detection; a 481-nm laser, a 505LP dichroic mirror and a 530/30-nm bandpass filter for GFP detection; and a 355-nm laser, a 450 LP dichroic mirror and a 515/30-nm bandpass filter for DAPI detection. Sorted samples were pelleted and resuspended in 2% FBS in PBS to obtain an estimated 5,000 cells per microliter. This enabled the targeting of approximately 25,000 cells for capture. To maximize our ability to obtain high-quality single-cell profiles, we conducted at least two PIP captures for each sorted cell population. Sequencing libraries were obtained from cells according to PIP-seq T2 instructions, using 16 cycles for cDNA amplification and 8–12 cycles for library amplification depending on the amount of DNA. Libraries were quantified on an Agilent 2100 Bioanalyzer and sequenced as per manufacturer recommendations on a NovaSeq X Plus flow cell (UCLA Broad Stem Cell Research Center Sequencing Core).

Demultiplexed FASTQs were obtained from the UCLA Sequencing Core. As in the organoid KD experiments, FASTQ files were aligned to a custom reference containing the human genome (GRCh38), genomes of four common mycoplasma species and the mCherry sequence, the GFP sequence or both, as dictated by the sorted population (PIPseeker, version 3). To maximize our ability to call single-cell droplets in each capture, sensitivity was set to 5, and the resulting count matrices were used as inputs for CellBender⁹² (remove-background function, version 0.3.2). These counts matrices were then used to generate Seurat objects for each capture (Seurat version 5, min.cells = 3), and doublets were excluded using DoubletFinder⁹³ (version 2.0.4). All samples were then merged into one Seurat object using rigorous quality control standards as detailed above (omitting cells with <500 genes detected and/or >5% of UMIs mapping to mitochondrial genes).

Cell clusters and cluster markers, UMAP projections and module 20 activity calculations were performed as described above.

To annotate cell types in the chimeroid data, the label transfer pipeline described for Seurat version 5 (FindTransferAnchors and MapQuery functions) was used to transfer cell subtypes from the developmental meta-atlas onto the chimeroid dataset.

Effects of FEZF2 and TSHZ3 shRNAs on expression of *FEZF2*, *TSHZ3*, module 20 and individual module 20 members were evaluated by first normalizing the expression level among the cells of each condition to the average expression level measured in the corresponding shRNA-negative cells. The average, normalized expression levels between shRNA-expressing cells versus shRNA-negative cells were then compared. In the case of *FEZF2*, *TSHZ3* and module 20 activity, these comparisons were conducted on a per-cell-line basis. The *ggpubr* package (version 0.6.0, <https://rpkgs.datanovia.com/ggpubr/>) was used to determine significance using Welch's *t*-test where indicated.

Molecular effects of these shRNAs were assessed by identifying genes positively enriched in indicated conditions with the FindMarkers function, Seurat version 5. Specifically, DEGs were identified between scram shRNA-expressing cells versus cells expressing FEZF2 shRNA, TSHZ3 shRNA or both. These comparisons were conducted among deep layer neuronal subtypes, IPCs and radial glial subtypes. The *enrichR* R package (version 3.2) was used to identify Gene Ontology terms enriched among these DEG sets, surveying the term databases described above. Dot plots displaying the most significantly enriched terms for each DEG set were generated using the *ggplot2* R package, version 3.4.3.

Demultiplexing of cell lines from transcriptomic data. To distinguish between cell lines in our transcriptomic datasets, single-nucleotide polymorphism (SNP)-based genomic probabilities were generated for each of our three cell lines using Genomic Diversity Array. These were converted to reference VCF files (plink version 1.9, maf = 0.05), from which multi-allelic variants were filtered out using SAMtools (version 1.21) before use as reference genotypes to demultiplex individual cell lines in our final chimeroid dataset using Vireo⁸¹ (run via Demuxafy, version 3). For PIP-seq data, sorted BAM files containing cell barcodes in the header as required by Vireo were generated using the barcode function of PIPseeker (version 3) with the `-sorted-bam` option. Cells were assigned to donor cell line identities using the `best_singlet` results from the Vireo pipeline.

10x single-cell multi-ome capture, sequencing and gene expression data analysis. At week 10, chimeroids were harvested for single-cell multi-omic profiling using a Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Kit CG000338 Rev F (10x Genomics). Eight chimeroids from the FEZF2-shRNA condition and nine chimeroids from the TSHZ3-shRNA condition were prepared for flow-assisted cell sorting as described above, with over 100,000 cells obtained per sort. Nuclei were extracted as per the 10x Nuclei Isolation Protocol (CG000365), using a 3-min incubation in lysis buffer and three 1-ml washes with wash buffer. Sequencing libraries were obtained from cells according to manufacturer instructions, using 7–9 cycles for ATAC library amplification, 6–9 cycles for cDNA amplification and 14–16 cycles for gene expression library amplification depending on the amount of DNA. Libraries were quantified on an Agilent 2100 Bioanalyzer and sequenced as per manufacturer recommendations on a NovaSeq X Plus flow cell (UCLA Broad Stem Cell Research Center Sequencing Core).

Demultiplexed FASTQs were obtained from the UCLA Sequencing Core. As in the chimeroid PIP-seq experiments, FASTQ files were aligned to a custom reference containing the human genome (GRCh38), genomes of four common mycoplasma species and the mCherry sequence, the GFP sequence or both, as dictated by the sorted population (*cellranger-arc*, version 2.2.0). These counts matrices were then used to generate Seurat objects for each capture using rigorous

quality control standards as detailed above (omitting cells with <500 genes detected and/or >5% of UMIs mapping to mitochondrial genes). Seurat objects across all experimental conditions were merged, and cell type labels were transferred from the developmental meta-atlas as described above. Evaluation of the effects of shRNAs on *FEZF2*, *TSHZ3* and module 20 gene activity was conducted as described for the chimeroid PIP-seq assays.

Generation of ATAC Seurat objects. For each experimental condition, ATAC peak/cell matrices and the fragments file were loaded using the *Read10X* function in Seurat version 5. Chromatin assays and Seurat objects were subsequently created using the counts from the Peaks assay (min. cells = 100, min.features = 200). Gene annotations were then extracted from EnsDb and added to the object. Quality control metrics (nucleosome signal scores, transcription start site (TSS) enrichment scores, fraction of counts in blacklist regions and fraction of reads in peaks) were then calculated using the built-in functions in Signac (version 1.14). Cells with *nCount_peaks* <3,000 and >150,000, nucleosome signal scores >2, TSS enrichment scores <1, fraction of counts in blacklist regions >0.015 and fraction of reads in peaks <0.5 were omitted. Individual Seurat objects were then merged using the *Merge* function in Seurat.

To evaluate the effects of FEZF2 and TSHZ3 shRNAs on chromatin accessibility, a gene activity matrix was created using the Signac Gene-Activity function on the merged Seurat object. The resulting matrix was added to the Seurat object as a new assay (GeneActivity) and normalized using *LogNormalize*, with the median *nCount_GeneActivity* as the scale factor. Evaluation of the effects of shRNAs on *FEZF2*, *TSHZ3* and module 20 chromatin accessibility was conducted using this Gene Activity assay similar to what was described for the chimeroid PIP-seq assays above. Genes with differential chromatin accessibilities between shRNA-expressing and shRNA-negative cells of the same condition were also identified using the FindMarkers function, and the *enrichR* R package (version 3.2) was used to identify Gene Ontology terms enriched among these differentially accessible peak sets. Dot plots displaying the most significantly enriched terms for each DEG set were generated using the *ggplot2* R package, version 3.4.3.

Statistical analysis

No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications^{25,79}. Data distribution was assumed to be normal, but this was not formally tested. Randomization was not applicable for this study as it was imperative to keep track of which experimental samples were which to ensure sample integrity. Data collection and analysis were not performed blinded to the conditions of the experiments. Blinding was not possible for experiments given the use of different fluorescent reporters to demarcate conditions and the need to ensure provenance of publically available datasets used in the meta-atlas. Immunofluorescent stainings were conducted to validate results from computational analyses, and conclusions were restricted to binary conclusions rather than subtle, graded comparisons between conditions. Data points from original publications were not included for meta-atlas generation if they did not meet quality control thresholds (see 'Acquiring datasets and quality control' subsection).

Reporting summary

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

Data availability

Our developmental and adult meta-atlases are browsable in the UCSC Genome Browser (<https://dev-ctx-meta-atlas.cells.ucsc.edu> and <https://adult-ctx-meta-atlas.cells.ucsc.edu>), in which our meta-atlas is also available as a Seurat object for download.

Original datasets for the developmental meta-atlas are accessible as follows: Bhaduri et al.⁹ (UCSC Cells–Neocortex: <https://dev-brain-regions.cells.ucsc.edu>), Fan et al.¹² (GEO Series [GSE103723](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103723)), Fan et al.¹³ (GEO Series [GSE120046](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120046)), Nowakowski et al.⁵ (UCSC Cells–Cortex Development: <https://cortex-dev.cells.ucsc.edu>), Polioudakis et al.⁶ (<http://solo.bmap.ucla.edu/shiny/webapp/>), Smith et al.¹⁶ (Figshare dataset: <https://figshare.com/s/64b648891e4817efb123>) and Zhong et al.¹⁸ (GEO Series [GSE104276](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104276)).

Original datasets for the adult meta-atlas are accessible as follows: Bakken et al.³⁴ (Human M1 10x: <https://portal.brain-map.org/atlas-and-data/rnaseq/human-m1-10x>), Caglayan et al.³⁵ (GEO Series [GSE192773](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192773)), Gandal et al.²⁰ (<https://doi.org/10.1038/s41586-022-05377-7>), Herring et al.¹⁴ (GEO Series [GSE168408](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168408)), Jorstad et al.³⁶ (Human Multiple Cortical Areas Smart-Seq: <https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq>), Lake et al.³⁷ (GEO Series [GSE97942](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97942)), Ma et al.³⁸ ([GSE207334](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207334)), Mathys et al.³⁹ (Aging Brain Data: http://compbio.mit.edu/ad_aging_brain/), Morabito et al.⁴⁰ (<https://doi.org/10.1038/s41588-021-00894-z>), Nagy et al.⁴¹ (GEO Series [GSE144136](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144136)), Pfisterer et al.⁴² (Epilepsy19: <https://khh.bric.ku.dk/Epilepsy19/>; metadata from <https://github.com/khodosevichlab/Epilepsy19>), Ramos et al.¹⁵ (<https://doi.org/10.1038/s41467-022-34975-2>), Schirmer et al.⁴³ (<https://doi.org/10.1038/s41586-019-1404-z>), Siletti et al.⁴⁴ (CellxGene Collection: <https://cellxgene.cziscience.com/collections/283d65eb-dd53-496d-adb7-7570c7caa443>), Velmshch et al.¹⁷ (Pre-Postnatal Cortex RNA: <https://pre-postnatal-cortex.cells.ucsc.edu>) and Zhu et al.⁴⁷ (CELLxGENE Collection: <https://cellxgene.cziscience.com/collections/ceb895f4-ff9f-403a-b7c3-187a9657ac2c>). Raw data, count matrices and relevant meta-data for the single-cell profiles of FEZF2/TSHZ3 KD organoids ([GSE242779](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE242779)) and chimeroids ([GSE278573](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE278573)) were deposited in the GEO.

Code availability

Custom scripts for meta-module generation, activity scoring and specificity scoring are available at <https://github.com/BhaduriLab/dev-ctx-meta-atlas>.

References

86. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
87. Shekhar, K. et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166**, 1308–1323 e1330 (2016).
88. Hie, B. L., Kim, S., Rando, T. A., Bryson, B. & Berger, B. Scanorama: integrating large and diverse single-cell transcriptomic datasets. *Nat. Protoc.* **19**, 2283–2297 (2024).
89. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
90. Kadoshima, T. et al. Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc. Natl Acad. Sci. USA* **110**, 20284–20289 (2013).
91. Velasco, S. et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
92. Fleming, S. J. et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).

93. 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).

Acknowledgements

We would like to thank the members of the Bhaduri Laboratory for their insightful advice and comments on the study. We would also like to thank the members of the Broad Stem Cell Research Center Flow Cytometry Core for their help in isolating cells for this project; C. Julian (UCSF Institute for Human Genetics Genomics Services Core) and S. Feng (UCLA Broad Stem Cell Research Center Sequencing Core) for help with running sequencing; and L. Fasano for generously sharing the antibody against TSHZ3. This study was generously funded by ROONS111731 from the National Institutes of Health (NIH) (National Institute of Neurological Disorders and Stroke), R01MH132689 from the NIH (National Institute of Mental Health (NIMH)), the Young Investigator Award from the Brain & Behavior Research Foundation, the Alfred P. Sloan Foundation, the Rose Hills Foundation, the Klingenstein-Simons Fellowship from the Esther A. & Joseph Klingenstein Fund and the Simons Foundation, the NIH BRAIN Initiative Cell Atlas Network (UM1MH130991) and the Ablon Scholar Award (to A.B.). Additional funding was provided to P.R.N. (UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Training Program, UCLA Intercampus Medical Genetics Training Program (USHHS Ruth L. Kirschstein Institutional National Research Service Award no. T32GM008243)), to C.V.N. (T32 NS048004, Predoctoral Fellowship in association with the Training Grant in Neurobehavioral Genetics), to R.K. (T32 GM145388, Cell and Molecular Biology Training Program) and to M.H. (NIMH BRAIN, NIMH RF1MH132662, NHGRI U24HG002371 and CIRM DISCO-14514 (with A.B.)).

Author contributions

The study was designed by A.B., P.R.N. and E.F. P.R.N., D.A., A.M., C.V.N., S.W., V.G. and R.L.K. performed experiments. Analysis was performed by P.R.N., E.F., A.M., J.Y., S.W. and A.B. The cell browser was designed and implemented by B.W. and M.H. The manuscript was written by P.R.N. and A.B., with input and edits from all authors.

Competing interests

The authors declare no competing interests.

Additional information

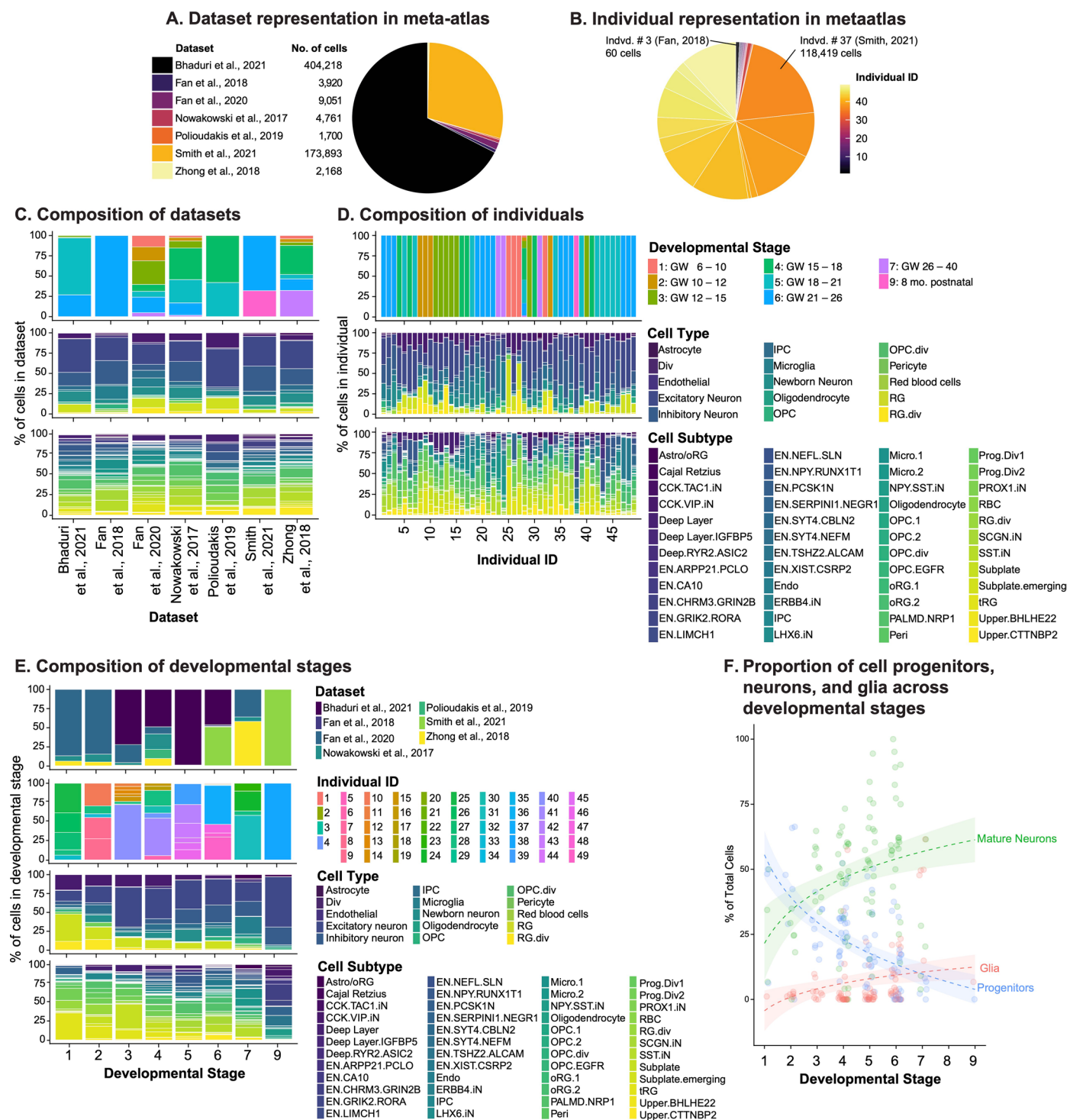
Extended data is available for this paper at <https://doi.org/10.1038/s41593-025-01933-2>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41593-025-01933-2>.

Correspondence and requests for materials should be addressed to Aparna Bhaduri.

Peer review information *Nature Neuroscience* thanks the anonymous reviewers for their contribution to the peer review of this work.

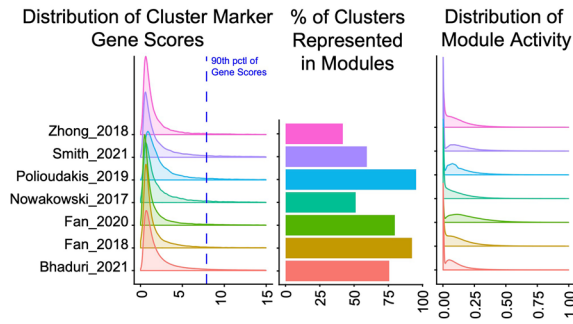
Reprints and permissions information is available at www.nature.com/reprints.



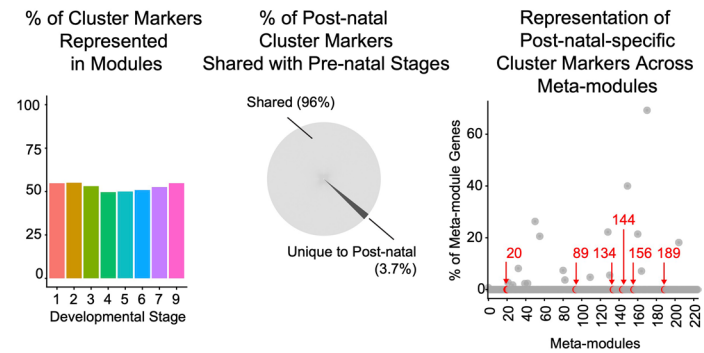
Extended Data Fig. 1 | Composition of the human developing neocortical meta-atlas. a) Our meta-atlas comprises mainly of cells from the Bhaduri et al.⁹ and Smith et al.¹⁶ datasets, with cells from five remaining datasets that increase the scope of developmental stages represented in our analyses. **b)** These datasets yield 96 individuals, with an individual from Smith et al.¹⁶ contributing the greatest number of cells. **c)** and **d)** The seven datasets (**c**) and 96 individuals (**d**) in our meta-atlas represent a unique set of developmental stages (top bar graphs), while still retaining cell type and subtype compositions expected for cortical tissues (middle and bottom bar graphs, respectively). **e)** Datasets and individuals in our meta-atlas span multiple stages during development (top

two bar graphs), as delineated previously based on biological landmarks (Kang et al.⁴⁸). The representation of cell types and subtypes across developmental stages in our meta-atlas change as expected (bottom two bar graphs), with early stages showing an enrichment in radial glia and progenitor subtypes and mature neurons and glial cells emerging in later timepoints. **f)** The percentage of progenitor cells, mature neurons, and glia in each individual change as expected across developmental stages. Cell type proportions for the indicated subtypes in each sample plotted as individual points, with trend lines showing logarithmic regression (95% confidence interval).

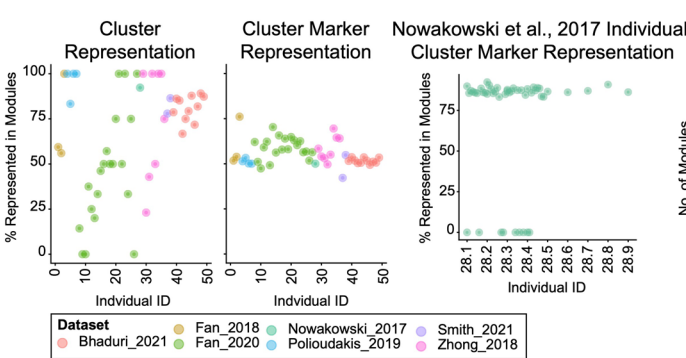
A. All developmental datasets are represented in the generation of meta-modules.



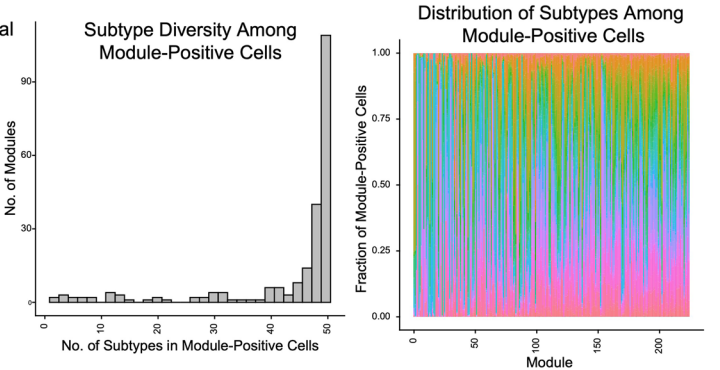
B. Comparable cluster marker representation of all developmental stages during module generation.



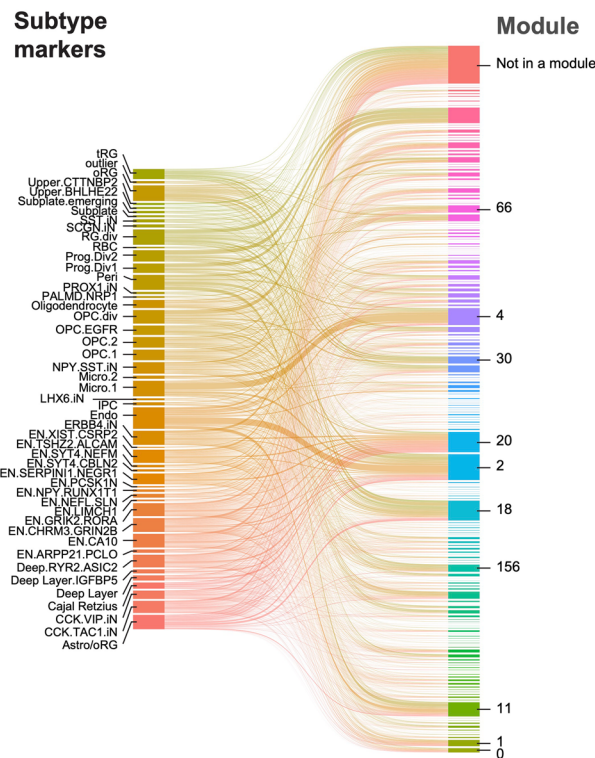
C. Representation of individuals in meta-module generation.



D. Module-positive cells represent a diversity of subtypes



E. Meta-modules described gene programs distinct from that of meta-atlas subtype marker genes.

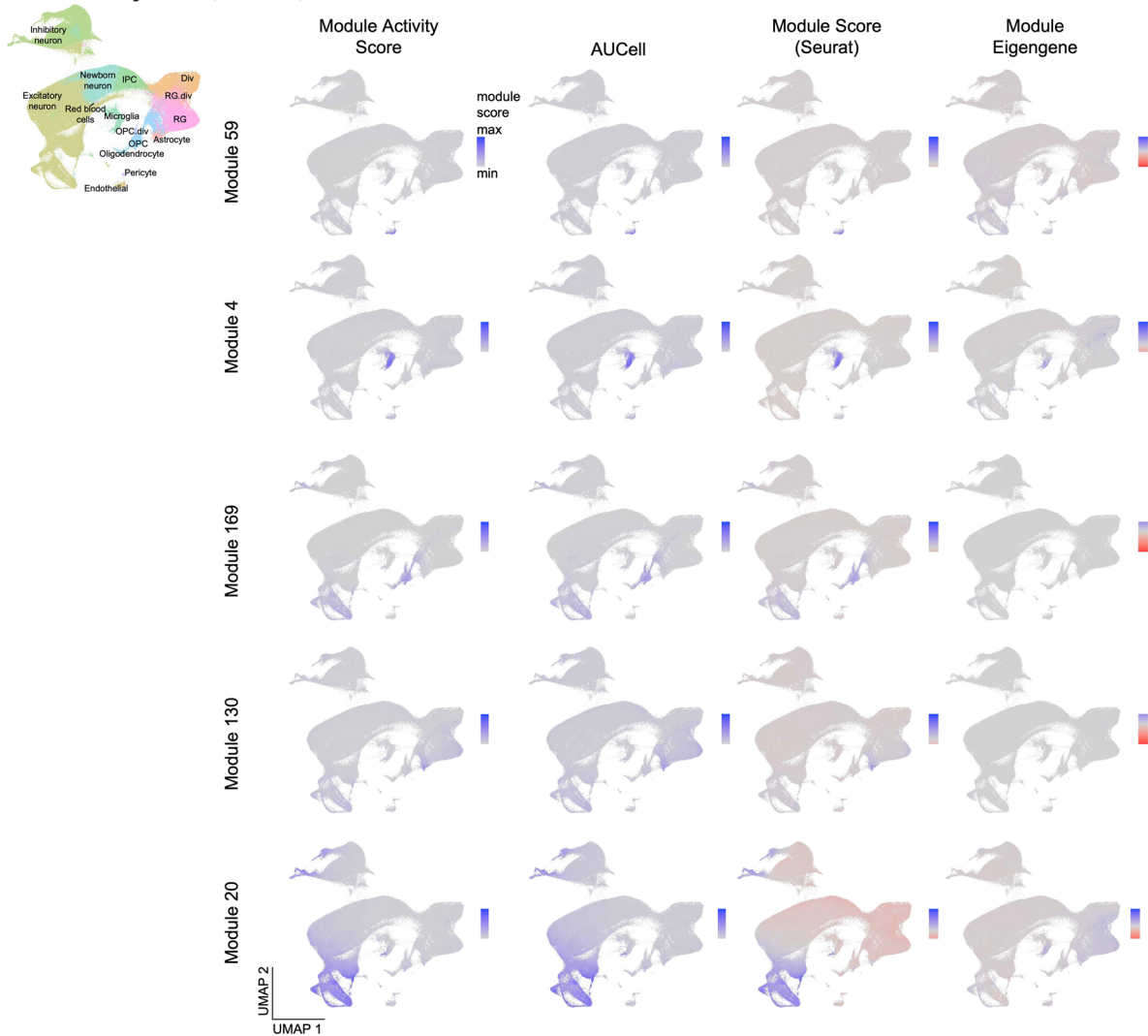


Extended Data Fig. 2 | See next page for caption.

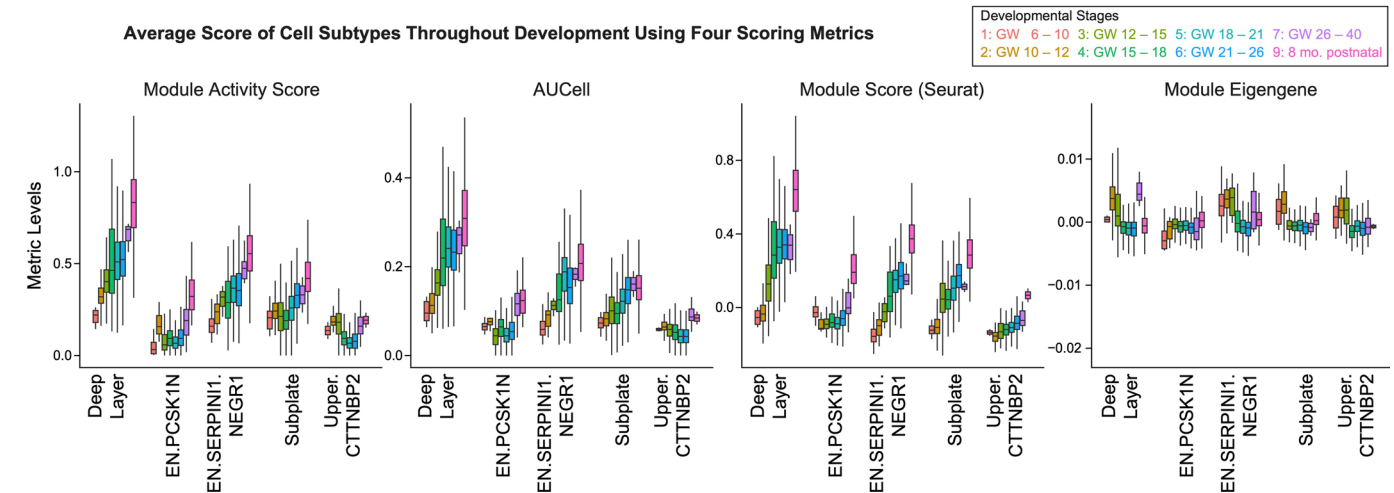
Extended Data Fig. 2 | Dataset, individual, and subtype representation in meta-modules. **a)** All seven datasets are represented in the generation of meta-modules. Clusters within datasets displayed similar gene score distributions (left), resulting in at least 40% of the clusters in each dataset being represented in the meta-modules (middle). Datasets also show comparable distributions of module activity overall (right). **b)** Developmental stages are equally represented in the meta-modules, with limited bias from pre-natal vs post-natal stages. Almost 50% of cluster markers of each developmental stage are represented in the meta-modules (left), with almost 96% overlap between the markers of post-natal vs pre-natal stages (middle). Only 18 of 225 modules harbor post-natal-specific markers (right), and notably none of these include our example modules (highlighted in red). **c)** Individual cluster markers well-represented in meta-modules. While individuals contributed clusters to meta-module generation to varying degrees (left), the cluster markers of each individual were represented in meta-modules at comparable levels (middle). While 9 individuals from Nowakowski 2017 did not produce cluster markers potentially due in part

to their relatively low number of cells (< 103), over 75% of cluster markers for the remaining individuals were represented in modules (right), suggesting no features of the 48 individuals in this dataset were lost due to merging and subsequent module construction. **d)** Subtype diversity among module-positive cells, defined as cells displaying the top 90th percentile of module activity within its respective dataset. Histogram of subtypes represented across module-positive cells (left) and bar plot showing the subtype proportions within each module (right) demonstrate the large diversity of cell subtypes within module-positive cells, and the ability of modules to reflect gene programs orthogonal to conventional cell type-centric annotations. **e)** Sankey plot demonstrating the representation (if present) of integrated meta-atlas cluster markers (left) among meta-modules (right). Genes marking the same subtype were often included in different meta-modules, or not included in meta-module generation at all, further highlighting the ability of meta-modules to elucidate biological features not fully encapsulated by cluster markers.

A. Cell type-specific activity patterns in select modules are consistent across expression-based metrics such as Module Activity Score, AUCell, and Module Score.

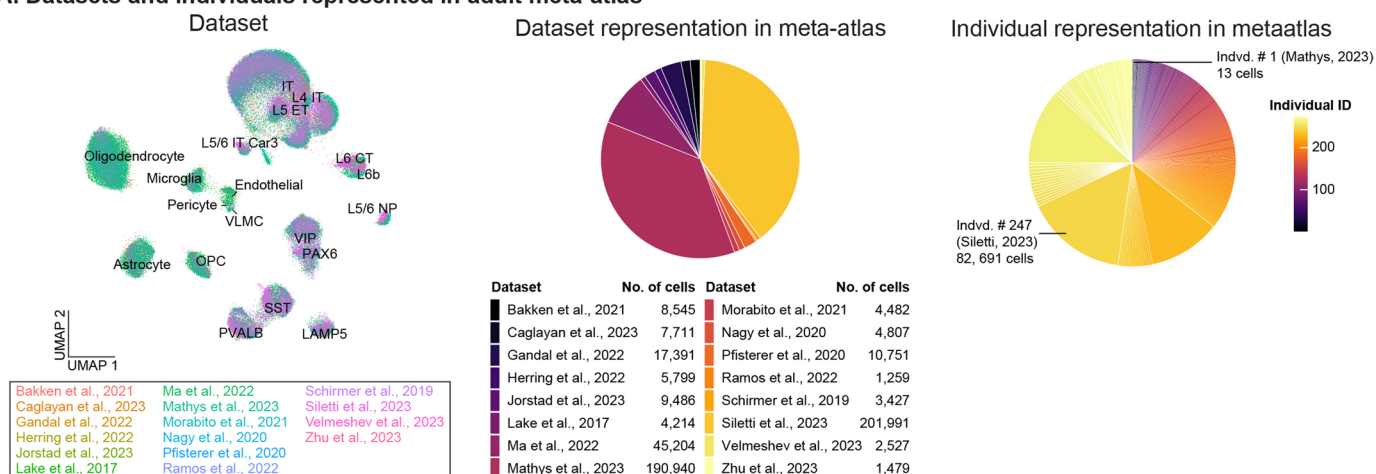
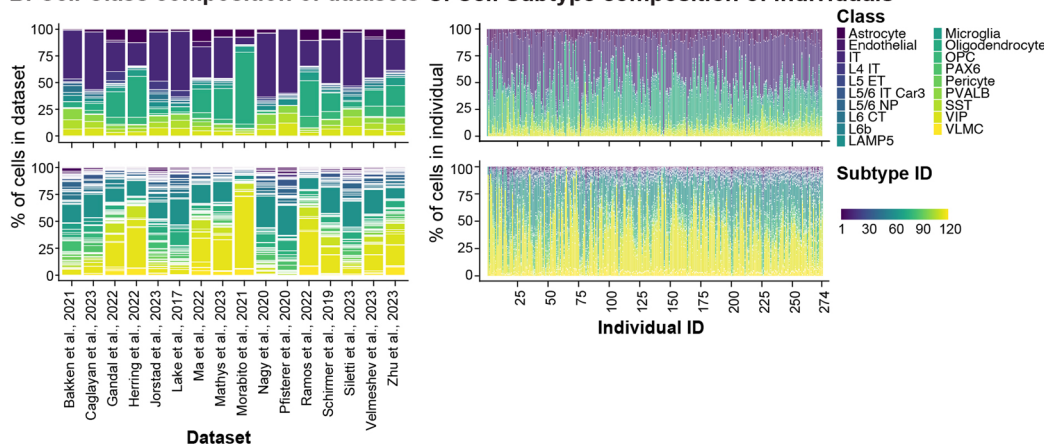
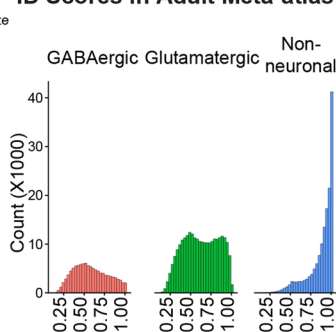
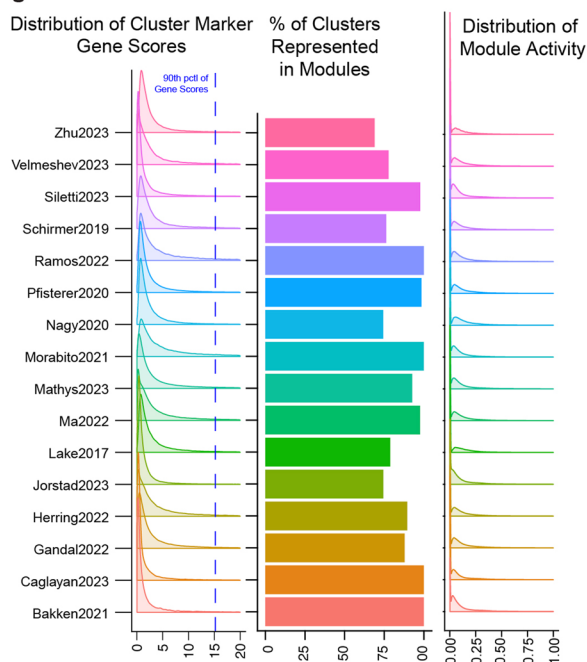
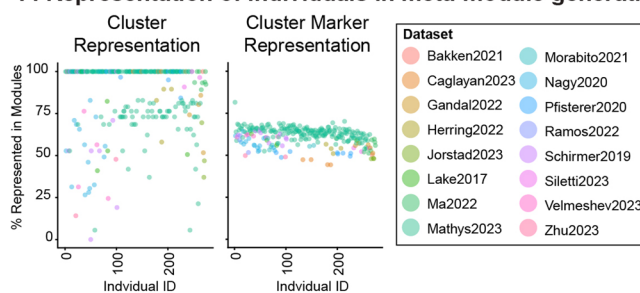
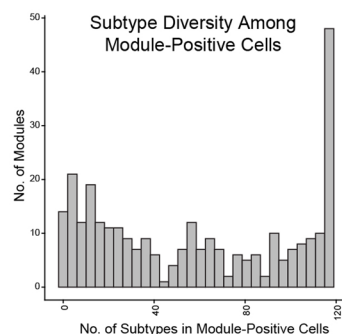


B. Deep layer-specific patterns of module 20 activity are still present in AUCell and Module Score calculations.



Extended Data Fig. 3 | Module activity scoring is comparable with existing gene co-expression metrics. a and b) Selected modules with cell type-specific activity patterns as measured by our module activity score demonstrated comparable activity patterns when evaluated using AUCell or Module Score. Module expression patterns evaluated by ModuleEigengene differed from these three metrics, likely due to its unique focus on other metrics besides

gene expression. **a)** UMAPs showing scores for module activity, AUCell, Module Score, and Module Eigengene across the developmental meta-atlas. **b)** Boxplots comparing the scores of these four metrics across developmental stages in deep layer and other neuronal subtypes. With the exception of Module Eigengene, deep layer neurons display an early, sustained increase in activity across all scoring algorithms.

A. Datasets and individuals represented in adult meta-atlas**B. Cell Class composition of datasets C. Cell Subtype composition of individuals****D. Distribution of Predicted ID Scores in Adult Meta-atlas****E. All developmental datasets are represented in the generation of meta-modules.****F. Representation of individuals in meta-module generation.****G. Module-positive cells represent a diversity of subtypes**

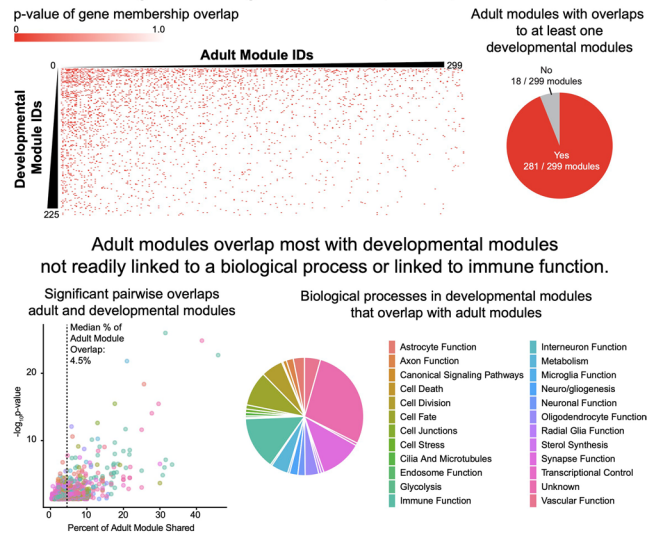
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Composition of the adult human neocortical meta-atlas. **a**) Cells from different datasets integrate together (left, UMAP), with the majority of adult meta-atlas cells originating from the Mathys et al.³⁹ and Siletti et al.⁴⁴ datasets and the remaining 14 datasets contributing cells that increase the diversity of individuals sampled in our meta-atlas (middle). In total, our atlas contains 274 individuals, within an individual from Siletti et al.⁴⁴ contributing the greatest number of cells (right). **b–d**) The 16 datasets (**b**) and 274 individuals (**c**) in our meta-atlas represent cell type and subtype compositions expected for cortical tissues (middle and bottom bar graphs, respectively), as annotated using cell type label transfer from the Jorstad et al.³⁶ profile of the adult cortex (also included in the adult meta-atlas) (**d**). **e**) All 16 datasets are represented in the generation of meta-modules. Clusters within datasets displayed similar gene

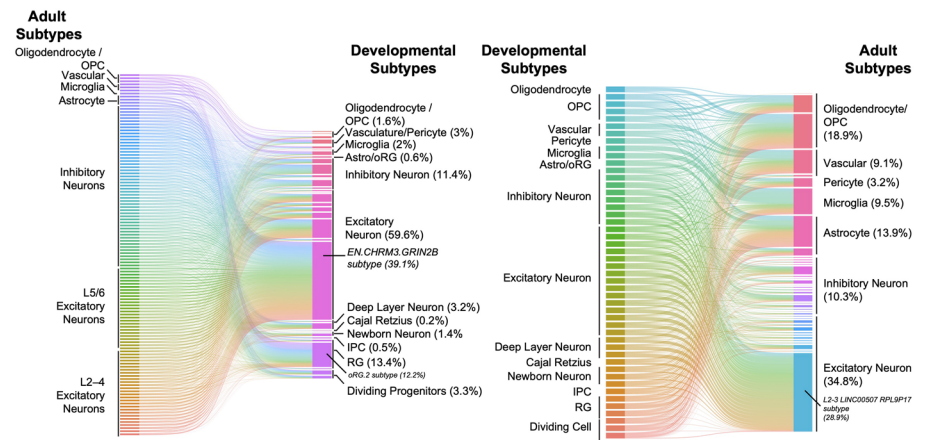
score distributions (left), resulting in over 50% of the clusters in each dataset being represented in the meta-modules (middle). Datasets also show comparable distributions of module activity overall (right). **f**) Individual cluster markers well-represented in meta-modules. While individuals contributed clusters to meta-module generation to varying degrees (left), the cluster markers of each individual were represented in meta-modules at comparable levels (middle). **g**) Histogram of subtypes represented across module-positive cells (cells displaying the top 90th percentile of module activity within its respective dataset) demonstrates the diversity of subtypes represented within each module-expressing population, similar to what was observed in the developmental meta-atlas.

B. Overlap in gene composition between adult and developmental meta-modules

Statistical significance of gene membership overlap between modules



D. Cell subtype-specificity of meta-modules diverge between development and adulthood.



The figure consists of two chord diagrams side-by-side, illustrating the distribution of adult subtypes within developmental and adult modules for Radial Glia and Astro/oRG.

Left Diagram: Modules Representative of Developmental Radial Glia

Adult Subtypes (Left to Right):

- VLMC (0.7%)
- Pericyte (4.7%)
- Oligodendrocyte / OPC (22%)
- Microglia (5.3%)
- Inhibitory Neuron (6.1%)
- Excitatory Neuron (6.1%)
- Endothelial (11.3%)
- Astrocyte (44%)

Developmental Modules (Left): Represented by a blue bar.

Adult Modules (Left): Represented by a red bar.

Marked by:

- Dev Modules Only (Blue)
- Adult Modules Only (Red)
- Both Module Types (Brown)

Right Diagram: Modules Representative of Developmental Astro/oRG

Adult Subtypes (Left to Right):

- Pericyte (4%)
- Oligodendrocyte / OPC (20%)
- Microglia (4%)
- Inhibitory Neuron (4%)
- Excitatory Neuron (2%)
- Endothelial (2%)
- Astrocyte (64%)

Developmental Modules (Left): Represented by a blue bar.

Adult Modules (Left): Represented by a red bar.

Marked by:

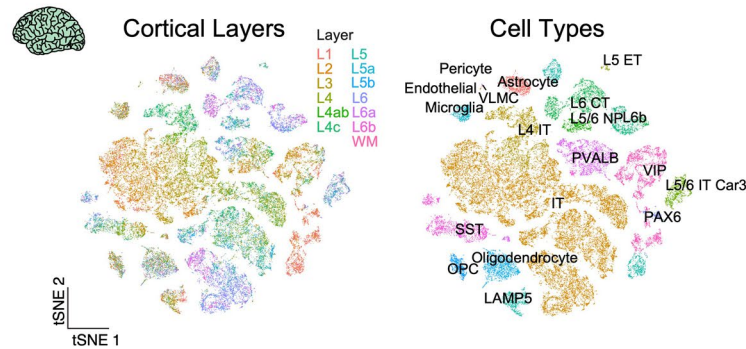
- Dev Modules Only (Blue)
- Adult Modules Only (Red)
- Both Module Types (Brown)

Nature Neuroscience

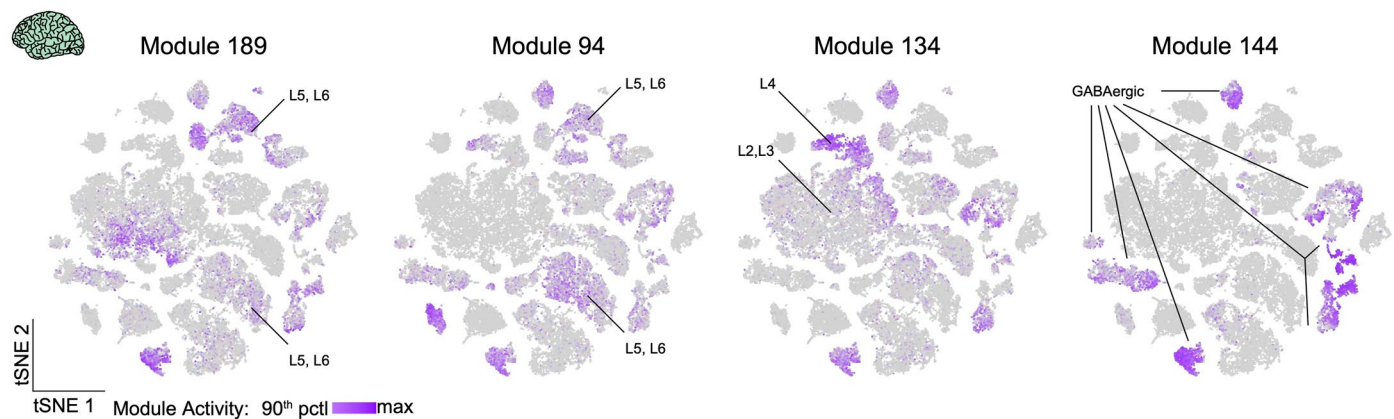
Extended Data Fig. 5 | Meta-modules allow comparison between developing and adult cortical cell types. **a)** Regardless of whether modules are derived from the developmental or the adult meta-atlas, adult subtypes typically harbor one or two strikingly specific modules while developmental cell types are represented by groups of modules with comparable specificity scores. Heat maps show module specificities of developmental modules on adult subtypes (top) and adult modules on developmental subtypes (bottom), with the average module specificity score and number of positively-scoring modules per subtype quantified in Fig. 3d. **b)** Evaluation of statistically significant overlap in gene composition across all developmental and adult modules. Heat map displays the p-values of the percent of genes in the indicated adult module (columns) that are shared in the indicated developmental module (rows). Statistical significance calculated using hypergeometric test. Though most adult modules share a significant overlap with at least one developmental module (top pie chart), the extent of overlap is overall low (bottom scatter plot). Adult modules overlap with developmental modules representing a variety of biological processes (bottom pie chart), most notably immune or synapse function. **c)** Developmental neuronal subtypes vary in their ability to be characterized by developmental vs adult meta-modules. Bar plots show the proportion of developmental vs adult modules among the 50 meta-modules with the highest specificity score within the indicated subtypes. Whereas upper layer excitatory neuronal subtypes are characterized primarily by adult modules, deep layer subtypes are

represented largely by developmental modules. Among inhibitory neurons, CCK+ subtypes are represented equally by developmental and adult modules, whereas SST-expressing interneuron identity are driven largely by adult modules. **d)** Cell subtype-specificity of meta-modules diverge between development and adulthood. (left) Sankey plot tracks the 50 meta-modules with the highest specificity score for each adult subtype. For each of these modules, lines link their cell subtype-specificity in the adult meta-atlas to the developmental cell subtype for which this module displays the highest specificity. Percentages show the proportion of modules displaying the greatest specificity to the indicated cell subtype – for example, most modules (39.1%) are specific to the EN.CHRM3.GRIN2B subtypes in development, regardless of their cell subtype-specificity in the adult. (right) Analogous Sankey plot linking the 50 meta-modules with the highest specificity score for each developmental subtype and their cell subtype-specificity in the adult meta-atlas. **e)** Meta-modules specific to progenitors in development can mark distinct sets of cell types in the adult. Sankey plots show the top 50 modules (regardless of source) that are specific to either radial glia (left) or astrocyte/outer radial glia (right) in the developmental meta-atlas, and their subsequent cell type-specificities in the adult. Most of these modules (44–64%) are specific to astrocytes in the adult. But while there is some convergence, notably among adult glial and non-neuronal cells, developmental modules mark a distinct set of neuronal cells versus adult meta-modules.

A. Distribution of cortical layers and cell types in a previously published atlas of adult cortical layers and areas (Jorstad et al., 2023).



B. Developmental modules linked to neuronal subtype specification are broadly expressed in the human adult cortex, with enrichments to specific neuronal classes or cortical layers.

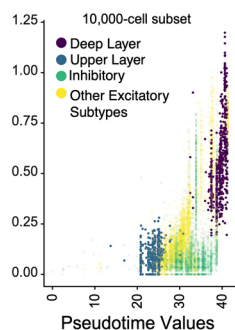


Extended Data Fig. 6 | Comparison to adult transcriptomic profiles annotated by layer reveals modules linked to layer-specific cell fates. Our module activity metric enables the assessment of meta-module expression in a transcriptomic atlas of the adult human cortex, that included layer-dissected and sequenced cell types. **a)** UMAPs show the cortical layers, cell types and subtypes represented in this single-nuclei dataset comprising of 49,495 nuclei spans the middle temporal gyrus, anterior cingulate cortex, and primary sensory cortices (visual, motor, somatosensory, and auditory). **b)** UMAPs showing the layer- and cell type-specific activities of modules 189, 94, 134, and 144 in the adult human

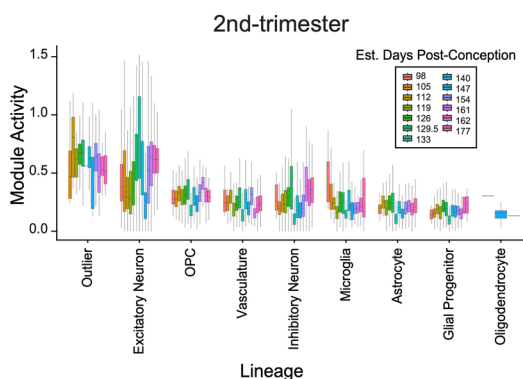
cortex. Modules 189, 94, and 134 is highest among glutamatergic neuronal subtypes, consistent with their activity among excitatory neuronal subtypes in our developing meta-atlas. These modules also display the greatest activity in distinct layers: module 189 and 94 dominant among deeper layers and module 134 most active in upper layers, in accordance with the later activation of module 134 among developing excitatory neurons. In contrast, cells with the greatest module 144 activity comprise almost exclusively of GABAergic cells, mirroring the enrichment of this module among inhibitory neurons in the developing human cortex.

A. Module 20 activity most prominent in excitatory vs inhibitory neurons.

Module Activity Across Pseudotime Values

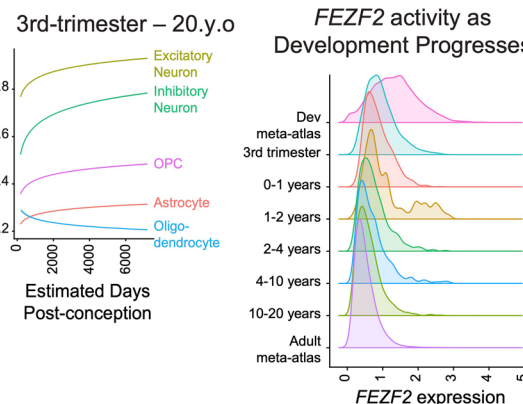


B. Module 20 activity patterns retained in Velmeshev et al., 2023 profile of the human cortex.

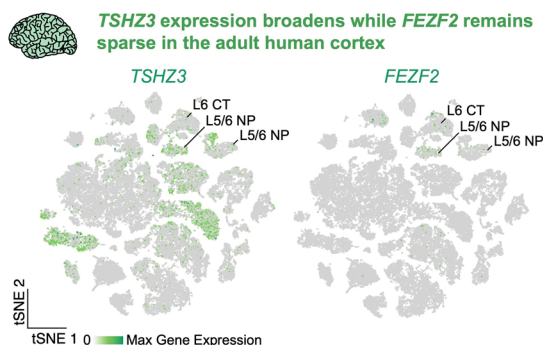


C. FEZF2 Levels Throughout Lifespan

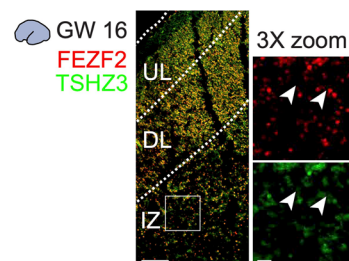
Broad decreases in FEZF2 activity as Development Progresses



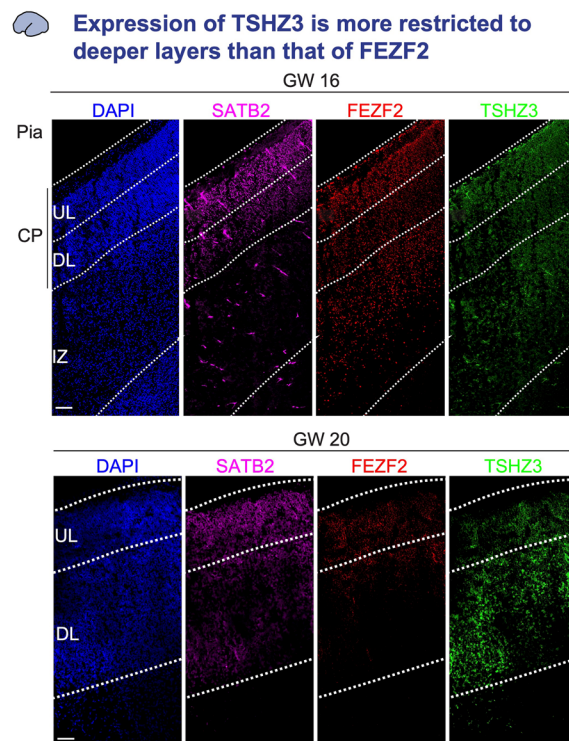
D. TSHZ3, FEZF2, and Module 20 activity in adult human cortex



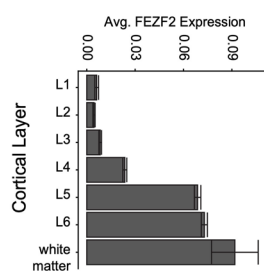
E. Lower prevalence of FEZF2+TSHZ3+ cells outside of the cortical plate.



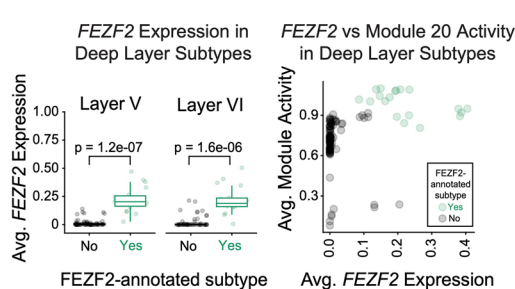
F. FEZF2 and TSHZ3 expression the developing human cortex relative to upper layer marker SATB2.



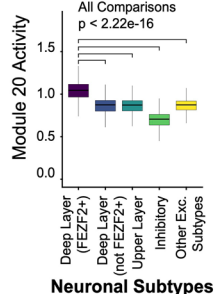
FEZF2 Levels Throughout Cortical Plate of Adult Human Cortex



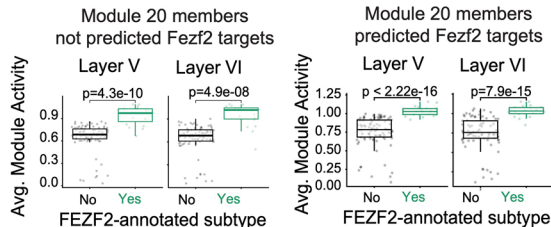
Deep layer cells of adult cortex include subtypes with high FEZF2 and Module 20 Activity



Module 20 Activity in the Adult Meta-atlas



Module 20 members predicted to be FEZF2 targets display activity patterns similar to members that are not



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

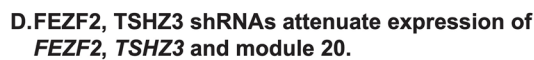
Extended Data Fig. 7 | Module 20, FEZF2, and TSHZ3 expression in the developing human cortex. **a)** Analysis of module 20 activity across pseudotime in a 10,000-cell subset of the developmental meta-atlas demonstrates that deep layer neurons are enriched in cells with both high module 20 activity and pseudotime values (greater maturity). In contrast, other excitatory subtypes and all inhibitory neuronal subtypes reflected a range of maturation states. While module 20 activity displayed a maturation-dependent increase among excitatory neurons, activity remained low in the majority of inhibitory subtypes irrespective of pseudotime, suggesting that the effects of module 20 are more prominent amongst excitatory vs inhibitory neuronal subtypes. **b)** The predominance of module 20 activity in excitatory neurons is further corroborated in the Velmeshev et al.¹⁷ dataset, both during second trimester (boxplots) and third trimester throughout young adulthood (regression models, 95% confidence interval). **c)** Across the timepoints profiled in our developmental meta-atlas, the Velmeshev et al.¹⁷ dataset, and our adult meta-atlas, FEZF2 expression broadly decreases within excitatory neuronal lineages with the exception of FEZF2 reactivation in a subset of cells during the 1–2 year period. **d)** Expression of FEZF2, TSHZ3, and Module 20 in the adult cortex. Top UMAPs highlight the expression of TSHZ3 in the Jorstad et al.³⁶ adult cortex dataset, in which TSHZ3 broadens relative to its expression during development. In contrast, FEZF2 expression in the adult is sparse and limited to a subset of deep layer cells (middle). Boxplots show the average log₁₀-normalized CPMs for FEZF2 among these subtypes and all other deep layer subtypes, with the average activity of each individual subtype displayed as a scatter plot. The

adjacent scatter plot further shows that these FEZF2-annotated deep layer subtypes exhibit both high FEZF2 expression and module 20 activity, in contrast to other layer 5/6 cells. Enrichment of Module 20 activity in FEZF2+ deep layer neurons was also observed in our adult meta-atlas (bottom left boxplot), though Module 20 members predicted to be Fezf2 targets did not display preferential enrichment to these cell types as measured in the Jorstad et al.³⁶ dataset (bottom right boxplots). Significance calculated with two-sided Welch's t-tests. There were $n = 19$ and $n = 90$ subtypes annotated with versus without FEZF2 in layer V, respectively, and $n = 18$ and $n = 85$ subtypes annotated with versus without FEZF2 in layer VI. **e** and **f** Immunofluorescent staining for FEZF2 (red) and TSHZ3 (green) in primary human cortical tissues. **e** In GW 16, FEZF2 and TSHZ3 do not co-express in the intermediate zones, as shown by 3x zoom-insets and white arrows. **f** Co-staining with the upper layer marker SATB2 (magenta) shows the presence of FEZF2+ neurons adjacent SATB2+ neurons at GW 16 (top), suggesting that at this stage, delineated boundaries between laminar fates have yet to be refined. In contrast, TSHZ3 expression is highest in cortical plate and subplate areas below SATB2+ neurons at GW16. Staining for these factors in the GW 20 primary human cortex (bottom) displayed the expected decreases in FEZF2 levels and a clearer distinction between SATB2+ and TSHZ3+ neurons, further suggesting that TSHZ3 is more restricted to deep layer neurons than is FEZF2. For all immunofluorescent images, dashed lines demarcate upper layers (UL), deep layers (DL) and intermediate zone (IZ), distinguished based on accompanying SATB2 or CTIP2 co-stains and well-established variations in cell densities across the human cortex. Scale bar = 100 μ m.

<i>FEZF2</i> and <i>TSHZ3</i> Expression in Sorted KD Chimeroids	% of Target Cells in Total Population	Reporter Gene Levels in Sorted Populations	Prevalence of Mitochondrial Genes
---	--	---	--------------------------------------



Distribution of technical replications within cortical chimeroids Distribution of individual cell lines within cortical chimeroids



Extended Data Fig. 8 | Single-cell transcriptomic profiling of FEZF2 and TSHZ3 depletion in cortical chimeroids. a) Several methods were used to evaluate the shRNA expression and cell health within our chimeroid model. Flow-assisted cell sorting (FACS) and subsequent RT-qPCR of week 7 cortical chimeroids (left) demonstrate the ability of FEZF2 and TSHZ3 shRNAs to attenuate expression of their target genes. Data generated from the sort of three organoids per condition, with double-negative populations from each sort pooled into the same collection tube to obtain enough cells for RT-qPCR analysis. Target populations (mCherry+ from TSHZ3 shRNA conditions, GFP+ from FEZF2 shRNA conditions) were isolated, and for the double KD the mCherry+ and GFP+ cells were also pooled with those from the respective single-KD conditions. Bar plots show the average \pm s.e.m. of the relative gene expression for each condition across three technical replicates. Bar plots to the right show the percent of target cells in the total population of each sort. Given the low (1.1%) rates of target cells in the 2x-KD condition, 33 chimeroids were used for FACS and subsequent transcriptomic profiling for these condition (compared to 3 chimeroids each for the remaining 3 shRNA conditions). Right violin plots demonstrate that across all sorted populations that were used for single-cell transcriptomic analysis, mCherry and GFP reporter expression was observed in the appropriate samples and all conditions displayed relatively low percentage of mitochondrial genes per cell. These data were indicative of cell health in our chimeroid samples, which we also assessed with immunostaining for CTIP2 (deep layer) and SOX2 (progenitors) in cortical chimeroids at week 5 (representative images, right). **b)** Multiple captures were collected for each of the 8 sorted populations (shRNA-expressing vs shRNA-negative cells for each of the 4 shRNA conditions) to maximize the number of cells obtained for single-cell transcriptomics. Pie chart demonstrates the relatively even distribution of these source samples, and UMAP highlights that cells cluster together with minimal batch effect from this multiple-capture strategy. **c)** SNP-based genomic probabilities profiled by

Genomic Diversity Array were used as input to demultiplex individual cell lines in our final chimeroid dataset using Vireo. Individual cell lines represented an even proportion of cells within the chimeroid at the time of single-cell transcriptomic harvest (pie chart, left), and cell lines were able to segregate into transcriptomic clusters with minimal batch effect (UMAP, right). **d)** UMAPs show the enrichment of FEZF2, TSHZ3, and Module 20 in deep layer neuronal subtypes, despite the relatively sparse FEZF2 and TSHZ3 expression in our dataset. Comparing FEZF2, TSHZ3, and module 20 expression in shRNA-expressing versus shRNA-negative cells from the same organoid revealed that FEZF2 and TSHZ3 shRNAs induced modest decrease in the expression of their target genes as well as in the activity of module 20. All of these phenotypes were exacerbated with the introduction of both FEZF2 and TSHZ3 shRNAs, and were consistent across all three cell lines, which are represented in the bottom plots as individual data points. Data points show, for each line and in each shRNA condition, the average activity of FEZF2 (left), TSHZ3 (middle) or module 20 activity (right) in shRNA-negative cells versus shRNA-expressing cells. Data normalization was conducted on a per-cell line basis, with points normalized to the average value detected for shRNA-negative cells of the indicated experimental condition. The average values \pm s.e.m measured across all three cell lines are represented as bar plots. Significance calculated with two-sided Welch's t-tests. **e)** Cells in chimeroid dataset were annotated by transferring cell subtype labels from the developmental meta-atlas (right histograms). UMAPs highlight the clustering of chimeroid cells by cell subtype, which is further supported by the expression of common cell type markers such as outer radial glia (HOPX), deep layer and neuronal markers (CRYM, MEF2C, STMN2). Additionally, markers of cell types not expected in chimeroid models were absent (astrocyte, GFAP; upper layer, SATB2). **f)** Cell subtype composition across all experimental conditions highlight the decrease in deep layer subtypes (red outlines) most prominently as a result of FEZF2 shRNAs or 2x-KD conditions.

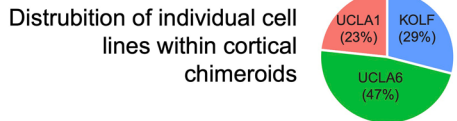
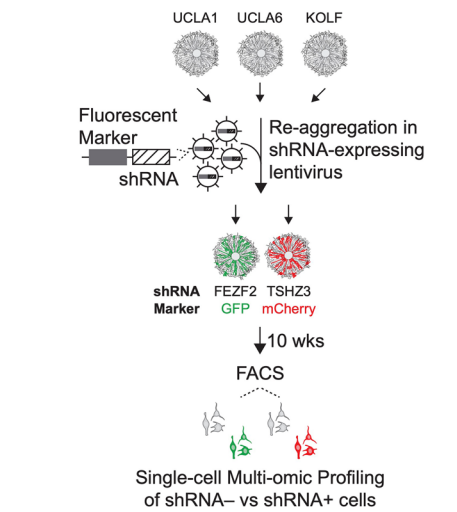
FEZF2 and TSHZ3 depletion
have additive impact
on DEG number and effect size



Extended Data Fig. 9 | Single-cell transcriptomic profiling of FEZF2 and TSHZ3 depletion in cortical chimeroids. a) Differential gene expression (DEG) analysis comparing the molecular effects of FEZF2 shRNA, TSHZ3 shRNA, and 2x-KD conditions to that of scram shRNA. Double-knockdown had additive effects on both the number of DEGs and the magnitude of gene expression changes relative to scram shRNA-expressing cells (left boxplots), with 2x-KD shRNA affecting the expression of synaptic genes both in deep layer neurons and radial glia. Bar graphs (left) plot the number of differentially expressed genes (DEGs) in each comparison, for each cell type, while boxplots show that the average gene scores of these DEGs are significantly higher in the 2x-KD conditions relative to single knock-down conditions. In deep layer cells, there were 305 DEGs, 421 DEGs, and 469 DEGs for FEZF2-shRNA, TSHZ3-shRNA, and 2x-KD conditions. In radial glia, there were 264 DEGs, 128 DEGs, and 370 DEGs for FEZF2-shRNA, TSHZ3-shRNA, and 2x-KD conditions. Significance calculated with two-sided Welch's t-tests. Gene ontology terms that are enriched within these DEG sets are displayed in dot

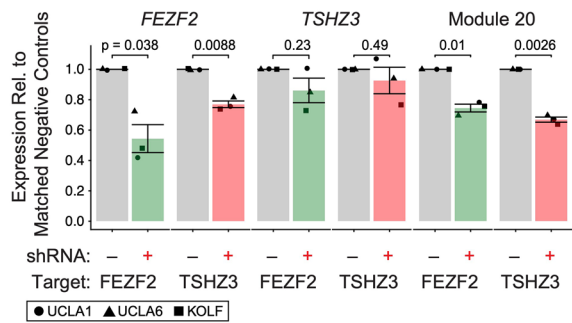
plots (left). Individual terms are shown with adjusted p-value represented as dot color and the percent of term-associated genes that are identified as DEGs. **b)** Emergence of a unique neuronal cell type in 2x-KD conditions. Top right UMAP shows that cells in the cluster 15 (generated using standard Louvain clustering using Seurat) are composed almost entirely of 2x-KD cells (inset), further shown by middle pie chart displaying composition of experimental conditions among cluster 15 cells. Cluster 15 cells are predominantly composed of a deep layer subset (left pie chart), but also contains other excitatory neuronal subtypes. The neuronal identity of this cluster is further supported by gene ontology term enrichment analysis of cluster 15 markers, demonstrating the prevalence of synaptic complexes. Individual terms are shown with adjusted p-value represented as dot color and the percent of term-associated genes that are identified as DEGs. Of note, an analog for cluster 15 cells in the adult meta-atlas was not identified.

A. Single-cell multi-omic analysis of *FEZF2* / *TSHZ3* depletion in 3-line human cortical chimeroids.

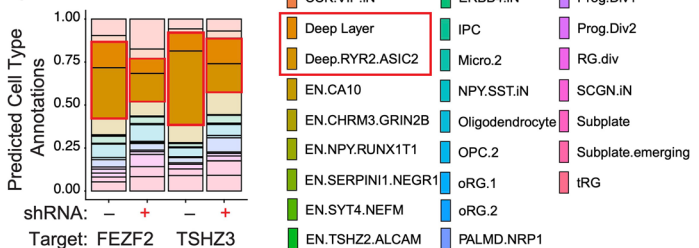


B. Effects of *FEZF2* and *TSHZ3* depletion on gene activity and cell types retained in multi-omic assay.

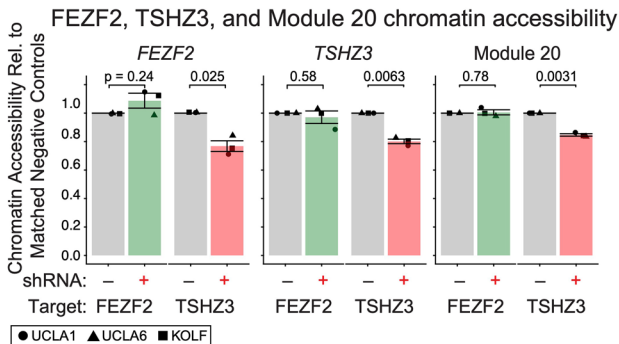
FEZF2, *TSHZ3*, and Module 20 activity measured in Multi-ome Assay



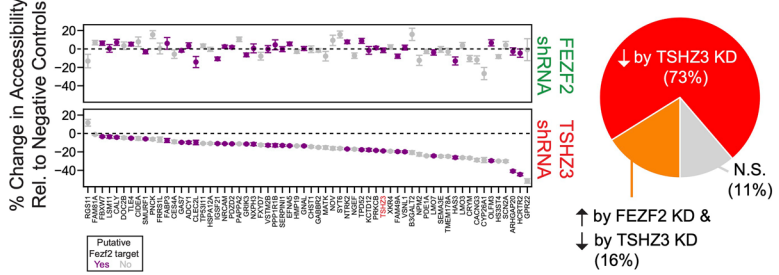
Cell Type Composition Across Experimental Conditions



C. *TSHZ3* – not *FEZF2* – is required for chromatin accessibility of *FEZF2*, *TSHZ3* and Module 20.

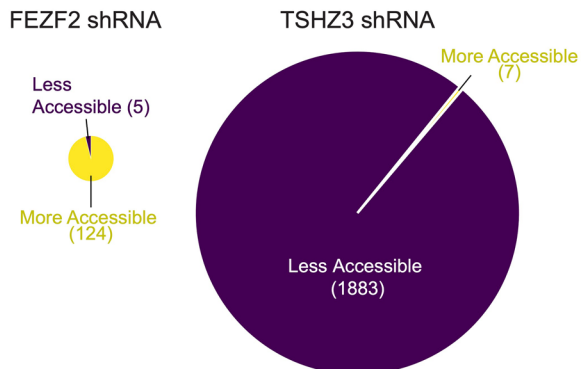


Effects of *FEZF2* and *TSHZ3* Depletion on Chromatin Accessibility Module 20 Genes

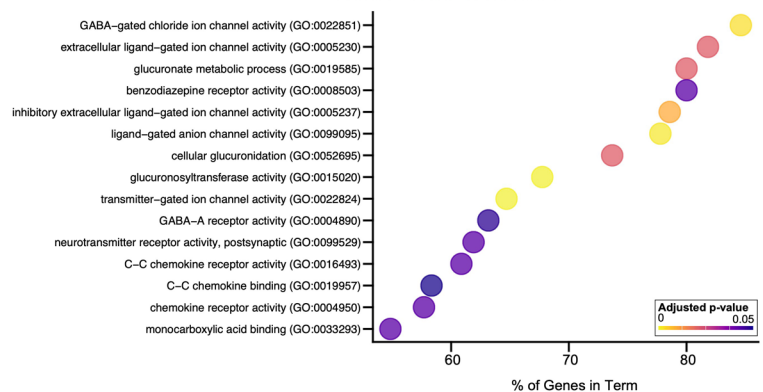


D. *TSHZ3* is required for the chromatin accessibility of neuronal and synaptic genes.

Effects of *FEZF2* and *TSHZ3* Depletion on Chromatin Accessibility



Gene Ontology Terms Enriched in Chromatin Regions with Decreased Accessibility Under *TSHZ3* shRNA



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Multi-omic analysis of FEZF2 and TSHZ3 depletion.

a) Human cortical chimeroids infected with either FEZF2 or TSHZ3 shRNA were used to simultaneously evaluate the effects of these transcription factors on the chromatin accessibility and gene expression of module 20 members. FEZF2/TSHZ3-depleted chimeroids were generated as described in Fig. 7a. Individual cell lines were demultiplexed as in Supplementary Fig. 18c, again exhibiting representation of all cell lines in the resulting multi-omic dataset (bottom pie chart). **b)** Phenotypes observed in our single-cell transcriptomic analysis were largely recapitulated in the transcriptomic signatures of our multi-omic analysis. Across all three cell lines, FEZF2 and TSHZ3 shRNAs induced modest decreases in the expression of their target genes as well as in the activity of module 20. Data points represent each cell line, showing for each shRNA condition, the average activity of *FEZF2* (left), *TSHZ3* (middle) or module 20 activity (right) in shRNA-negative cells versus shRNA-expressing cells. Data normalization was conducted on a per-cell line basis, with points normalized to the average value detected for shRNA-negative cells of the indicated experimental condition. The average values \pm s.e.m measured across all three cell lines are represented as bar plots. Significance calculated with two-sided Welch's t-tests. Bottom composition plot shows the distribution of cell subtypes in the multi-omic dataset, as annotated via label transfer from developmental meta-atlas. As in our transcriptomic profiling, FEZF2 and TSHZ3 shRNAs were sufficient to decrease deep layer subtypes (outlined in red). **c)** TSHZ3 – not – FEZF2 – is required for the chromatin accessibility of *FEZF2*, *TSHZ3*, and Module 20 genes. Chromatin accessibility of the individual *FEZF2* and *TSHZ3* genes, as well as average accessibility for Module 20 genes, were first evaluated in shRNA-expressing vs shRNA-negative cells. Chromatin accessibility is defined as the accessibility of the promoter and gene body regions associated with each gene. Bar plots (left) show that only TSHZ3 shRNAs significantly decreased chromatin accessibility for these genes, a phenotype consistent across all three cell lines. Data for each line shown as individual points, which were normalized to the average value

detected for shRNA-negative cells of the indicated experimental condition on a per-cell line basis. The average values \pm s.e.m measured across all three cell lines are represented as bar plots, and significance calculated with two-sided Welch's t-tests. Expanding this analysis to each individual Module 20 gene further corroborated the importance of TSHZ3, but not FEZF2, to the chromatin landscape of these genes. Right dot plots show that while TSHZ3 shRNAs led to a concomitant decrease in chromatin accessibility for nearly all Module 20 genes, FEZF2 shRNAs had little-to-no effect on the majority of Module 20. Data shown are the average \pm s.e.m percent change in chromatin accessibility in shRNA-expressing cells vs shRNA-negative cells of the indicated experimental conditions, with significance calculated as two-sided Welch's t-test. There were 549 cells and 328 cells in the shRNA-negative populations of the FEZF2 shRNA and TSHZ3 shRNA conditions, respectively. There were 517 and 5,342 cells in the shRNA-positive populations of the FEZF2 shRNA and TSHZ3 shRNA conditions, respectively. Consistent with our transcriptomic profiling in Fig. 7e, down-regulated Module 20 genes include TSHZ3 (red), however shRNAs displayed comparable effects on both module 20 genes that are putative *Fefz2* targets (purple) versus those that are not (grey). These data are summarized in the right pie chart, showing that TSHZ3 shRNA significantly reduces the chromatin accessibility of Module 20 genes. **d)** Left pie charts comparing the effects of FEZF2 and TSHZ3 shRNAs on the chromatin accessibility beyond that of Module 20 genes show that TSHZ3 shRNAs reduce chromatin accessibility on a large scale. The 1,883 genes that have significantly less accessible chromatin regions in the presence of TSHZ3 shRNAs are characterized by genes involved in synaptic activity, as shown by gene ontology term enrichment analysis. Differentially accessible chromatin (DAC) analysis was conducted between shRNA-expressing cells vs shRNA-negative cells of the same organoid. Gene ontology terms that are enriched within these DAC sets are displayed in dot plots (left). Individual terms are shown with adjusted p-value represented as dot color and the percent of term-associated genes that are identified as DACs.

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 No software was used for the purpose of data collection.

Data analysis The following software was used for data analysis:

R (versions 4.1.0 or 4.3.1)
 Seurat (v4, v5)
 Monocle3 (v1.3.4)
 ggpubr (v0.6.0)
 EnrichR (v3.2)
 ggplot2 (v3.4.3)
 PIPseeker (v02.01.04, v3)
 bcl2fastq (v2.20.0.422)
 FIJI (RRID:SCR_002285)
 Imaris (v10, RRID:SCR_007370)
 Slingshot (v2.7.0)
 WGCNA (v1.72.5)
 factoextra
 dynamicTreeCut
 data.table
 morpheus R package (v1.0-4)
 reticulate (v1.36.1)

scanorama
 pySCENIC (v0.12.1)
 cistargetDB
 DoubletFinder (v2.0.4)
 plink (v1.9)
 samtools (v1.21)
 Vireo (run via Demuxafy v3)
 cellranger-arc (v2.2.0)
 Signac (v1.14)

Custom code was used in this paper and is publicly available at Github: <https://github.com/BhaduriLab/dev-ctx-meta-atlas>

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](#)

Our developmental and adult meta-atlases are browsable in the UCSC Genome Browser:

(<https://dev-ctx-meta-atlas.cells.ucsc.edu> and <https://adult-ctx-meta-atlas.cells.ucsc.edu>), in which our meta-atlas is also available as a Seurat object for download. Original datasets for the developmental meta-atlas are accessible as follows: Bhaduri et al., 2021 (UCSC Cells - Neocortex: <https://dev-brain-regions.cells.ucsc.edu>), Fan et al., 2018 (GEO Series GSE103723), Fan et al., 2020 (GEO Series GSE120046), Nowakowski et al., 2017 (UCSC Cells - Cortex Development: <https://cortex-dev.cells.ucsc.edu>), Polioudakis et al., 2019 (<http://solo.bmap.ucla.edu/shiny/webapp/>), Smith et al., 2021 (Figshare - Dataset: <https://figshare.com/s/64b648891e4817efb123>), and Zhong et al., 2018 (GEO Series GSE104276).

Original datasets for the adult meta-atlas are accessible as follows: Bakken et al., 2021 (Human M1 10x: <https://portal.brain-map.org/atlas-and-data/rnaseq/human-m1-10x>), Caglayan et al., 2023 (GEO Series GSE192773), Gandal et al., 2022 (DOI: 10.1038/s41586-022-05377-7), Herring et al., 2022 (GEO Series GSE168408), Jorstad et al., 2023 (Human Multiple Cortical Areas Smart-Seq: <https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq>), Lake et al., 2017 (GEO Series GSE97942), Ma et al., 2022 (GSE207334), Mathys et al., 2023 (Aging Brain Data: http://compbio.mit.edu/ad_aging_brain/), Morabito et al., 2021 (DOI: 10.1038/s41588-021-00894-z), Nagy et al., 2020 (GEO Series GSE144136), Pfisterer et al., 2020 (Epilepsy19: <https://khh.bric.ku.dk/Epilepsy19/>; metadata from: <https://github.com/khodosevichlab/Epilepsy19>), Ramos et al., 2022 (DOI: 10.1038/s41467-022-34975-2), Schirmer et al., 2019 (DOI: 10.1038/s41586-019-1404-z), Siletti et al., 2023 (CellxGene Collection: <https://cellxgene.cziscience.com/collections/283d65eb-dd53-496d-adb7-7570c7caa443>), Velmeshev et al., 2023 (Pre-Postnatal Cortex RNA: <https://pre-postnatal-cortex.cells.ucsc.edu>), and Zhu et al., 2023 (Cellxgene Collection: <https://cellxgene.cziscience.com/collections/ceb895f4-ff9f-403a-b7c3-187a9657ac2c>).

Raw data, count matrices, and relevant meta-data for the single-cell profiles of FEZF2 / TSHZ3 KD organoids (GSE242779) and chimeroids (GSE278573) were deposited in the NCBI Gene Expression Omnibus.

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	Human data was used from publicly available sources and metadata can be retrieved from source datasets if available.
Reporting on race, ethnicity, or other socially relevant groupings	No reporting on race, ethnicity or other socially relevant groupings were used.
Population characteristics	No population characteristics were used.
Recruitment	No participants were recruited.
Ethics oversight	All brain samples and iPSC cell lines collected for analysis in this study were obtained with consent and donor anonymization. All of the procedures were reviewed and approved by the UCLA IBC and hPSCRO (UCLA Institutional Biosafety Committee approval number BUA-2020-091-025-A; UCLA Human Pluripotent Stem Cell Research Oversight Committee approval number 2020-008-05).

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](https://www.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	No sample size calculation was performed for meta-atlas generation and analysis due to the use of publicly available datasets. As is a typical standard in the field, all experiments were replicated in biological duplicate or triplicate where possible and immunostainings were performed in duplicate based upon human tissue availability, with one tissue piece used for each age represented. For single-cell transcriptomic/multiomic experiments, the basis of outputs on single cells restricted our analysis to the number of high-quality cells obtained per capture.
Data exclusions	<p>Developmental meta-atlas: For all datasets but the Polioudakis et al., 20196, Bhaduri et al., 20219, Nowakowski et al., 20175, gene count matrices for individual cells and corresponding metadata were downloaded from the Gene Expression Omnibus, and cells from the same individual were combined. Polioudakis et al data was downloaded from the browser within the publication (see below) and Bhaduri et al and Nowakowski et al were received via personal communication, but could also be downloaded from the UCSC Cell Browser⁸⁴. In the case of Smith et al., 202116, directories containing the 10X-provided matrix.mtx, genes.tsv, and barcodes.tsv were downloaded from https://figshare.com/s/64b648891e4817efb123 as originally described¹⁶. We then processed the 10X-derived count matrices and directories from each individual with a standard pipeline to generate Seurat objects (Seurat v426), conducting a normalization of the counts as needed and filtering out cells with < 500 genes detected and > 5% of UMIs mapping to mitochondrial genes. Genes detected in < 3 cells were omitted. In cases where the original publication used more stringent criteria for these quality control measures, we defaulted to the original publication's settings (Table S1). In the case of Polioudakis et al., 2019, after downloading raw counts (UMI) gene expression matrix and metadata from the original paper's indicated data browser, we adapted the quality control measures described by the authors, selecting for cells with the 95th percentile of total UMIs with < 250 UMIs mapping to the mouse genome, and < 5% of UMIs mapping to mitochondrial genes. For consistency with the remainder of our meta-atlas, we then applied a stricter minimum for the number of genes detected per cell (increasing the minimum from 200 in the original paper to 500), and we did not place a maximum on the genes detected per cell, as the biologically relevant range of the number of genes per cell will vary between cell types. We then removed genes detected in < 3 cells.</p> <p>Adult meta-atlas: With the exception of datasets for Gandal et al., 202220, Ramos et al., 202215, and Schirmer et al., 201943, Seurat objects or gene counts matrices for individual cells and corresponding metadata for each dataset were downloaded from the Gene Expression Omnibus or as otherwise instructed in the original publication (Table S13). Seurat objects for the Gandal et al., 2022, Ramos et al., 2022, and Schirmer et al., 2019 datasets were obtained via personal communication from M. Gandal, S. Ramos, and A. Zulji, respectively. All datasets for adult meta-atlas are single-nuclei profiles. Seurat objects for each individual were generated as described for the developmental meta-atlas (Seurat v5), with the exception that the additional filtering based on the percent of mitochondrial genes per cell was applied to only a few datasets given the relatively low mitochondrial RNA detected single-nuclei sequencing (Table S13). In the case of Siletti et al., 2023, area-specific Seurat objects that corresponded to cortical areas were downloaded and additional modifications were required for computation feasibility. For two individuals (H19.30.001 and H19.30.002), individual-level Seurat objects were generated by first extracting the counts matrix for each individual from each area-specific Seurat object. Features in the resulting matrices were converting Ensembl IDs to Gene Symbols, and the feature counts were summarized across genes to collapse data from different transcripts of the same gene. Seurat objects were generated from the resulting individual-level, area-specific counts matrix, and the 5 area-specific Seurat objects for were merged into one object representing a single individual. The features in the resulting Seurat objects were filtered to retain just the genes that overlapped with the H18.30.001 in the Bakken et al., 2021 dataset. These objects were then randomly subset into 8 objects. The H18.30.002 individual was subset by first randomly grouping the cell barcodes belonging to this individual into 11 subsets, which were then used as a guide to generate Seurat objects as detailed above. Features from the counts matrices of the Zhu et al., 2023 datasets were also converted from Ensembl ID to Gene Symbols and counts matrices were filtered to retain just the features that overlap with the Bakken et al., 2021 dataset for consistency with the remainder of the meta-atlas.</p>
Replication	All experiments were replicated in duplicate or triplicate where possible, immunostainings were performed in duplicate based upon human tissue availability, with one tissue piece used for each age represented.
Randomization	Randomization was not applicable for this study as it was imperative to keep track of which experimental samples were which to ensure sample integrity.
Blinding	Blinding was not possible for experiments given the use of different fluorescent reporters to demarcate conditions and the need to ensure provenance of publically available datasets used in the meta-atlas. Immunofluorescent stainings were conducted to validate results from computational analyses, and conclusions were restricted to binary conclusions rather than subtle, graded comparisons between conditions.

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Primary antibodies: FEZF2(ab214186, abcam, Lot 1006049, Clone N/A); TSHZ3(N/A, Lot N/A, Clone N/A: Non-commercial guinea-pig anti-TSHZ3 antibody produced in-house by Alistair Garratt's laboratory was raised against mouse amino acids 557-664, cloned as a His-tagged fusion protein in pET14b); CTIP2(ab18465, abcam, Lot 2101034078, Clone 25B6); SATB2(ab51502, abcam, Lot GR3281721-13, Clone SATBA4B10); HS6ST2(ab122220, abcam, Lot GR148036-25, Clone N/A); ADAM33(sc-514055, Santa Cruz Biotechnology, Lot E1716, Clone A-3); ZBTB16/PLZF(sc-28319, Santa Cruz Biotechnology, Lot I1022, Clone D-9); VGAT(AB5062P, Millipore Sigma, Lot 3974593, Clone N/A); QKI(MABN624, Millipore Sigma, Lot 3506608, Clone N147/6); PDLIM5/ENH(388800, Thermo Fisher, Lot RC234862, Clone N/A); SOX2(sc-365823, Santa Cruz Biotechnology, Lot C2523, Clone E-4); VSNL1(UM870034, OriGene, Lot W001, Clone UMAB115); SALL1(PA5-62057, Thermo Fisher, Lot ZG4382825A, Clone N/A), c-FOS(sc-16694, Santa Cruz Biotechnology, Lot C2523, Clone E-8), and ZFP36L1/2/BRF1(12306-1-AP, Cell Signaling Technology, Lot 00082509, Clone N/A).

Polyclonal AlexaFluor secondary antibodies from Thermo Fisher used: mouse (488: # A-31570, A-11001; 555: # A-31572, A-11005; 568: # A-31573, A-11004; 647: # A-31574, A-11029); rat (488: # A-21208, A-11007; 555: # A-21209, A-11009; 568: # A-21210, A-11008; 647: # A-21211, A-11037); guinea pig (488: # A-11073, A-11075; 555: # A-21449, A-21450; 568: # A-11075, A-11073; 647: # A-21450, A-21449); rabbit (488: # A-11008, A-11034; 555: # A-21428, A-21207; 568: # A-11011, A-21245; 647: # A-21245, A-21207); goat (488: # A-11055, A-11034; 555: # A-21432, A-11039; 568: # A-11056, A-11040; 647: # A-21435, A-11041).

Validation

In-house TSHZ3 antibody was described and validated extensively in Caubit, X. et al. TSHZ3 deletion causes an autism syndrome and defects in cortical projection neurons. Nat Genet 48, 1359-1369, doi:10.1038/ng.3681 (2016).

Validation statements for all commercial antibodies listed above can be found through the following links to the manufacturer's website:

Primary Antibodies:

FEZF2 (ab214186, Abcam): <https://www.abcam.com/fezf2-antibody-ab214186.html>
 CTIP2 (ab18465, Abcam): <https://www.abcam.com/ctip2-antibody-ab18465.html>
 SATB2 (ab51502, Abcam): <https://www.abcam.com/satb2-antibody-ab51502.html>
 HS6ST2 (ab122220, Abcam): <https://www.abcam.com/hs6st2-antibody-ab122220.html>
 ADAM33 (sc-514055, Santa Cruz Biotechnology): <https://www.scbt.com/p/adam33-antibody-a-3>
 ZBTB16/PLZF (sc-28319, Santa Cruz Biotechnology): <https://www.scbt.com/p/plzf-antibody-d-9>
 VGAT (AB5062P, Millipore Sigma): <https://www.sigmaaldrich.com/catalog/product/sigma/ab5062p>
 QKI (MABN624, Millipore Sigma): <https://www.sigmaaldrich.com/catalog/product/sigma/mabn624>
 PDLIM5/ENH (388800, Thermo Fisher): <https://www.thermofisher.com/order/catalog/product/PA5-62057>
 SOX2 (sc-365823, Santa Cruz Biotechnology): <https://www.scbt.com/p/sox2-antibody-e-4>
 VSNL1 (UM870034, OriGene): <https://www.origene.com/catalog/antibodies/primary-antibodies/um870034/vsnl1-antibody-umab115>
 SALL1 (PA5-62057, Thermo Fisher): <https://www.thermofisher.com/order/catalog/product/PA5-62057>
 c-FOS (sc-16694, Santa Cruz Biotechnology): <https://www.scbt.com/p/c-fos-antibody-e-8>
 ZFP36L1/2/BRF1 (12306-1-AP, Cell Signaling Technology): <https://www.cellsignal.com/products/antibodies/zfp36l1-2-brf1-antibody-12306>

Secondary Antibodies:

Mouse Antibodies

A-31570: <https://www.thermofisher.com/order/catalog/product/A-31570>
 A-11001: <https://www.thermofisher.com/order/catalog/product/A-11001>
 A-31572: <https://www.thermofisher.com/order/catalog/product/A-31572>
 A-11005: <https://www.thermofisher.com/order/catalog/product/A-11005>
 A-31573: <https://www.thermofisher.com/order/catalog/product/A-31573>
 A-11004: <https://www.thermofisher.com/order/catalog/product/A-11004>
 A-31574: <https://www.thermofisher.com/order/catalog/product/A-31574>
 A-11029: <https://www.thermofisher.com/order/catalog/product/A-11029>

Rat Antibodies

A-21208: <https://www.thermofisher.com/order/catalog/product/A-21208>
 A-11007: <https://www.thermofisher.com/order/catalog/product/A-11007>
 A-21209: <https://www.thermofisher.com/order/catalog/product/A-21209>
 A-11009: <https://www.thermofisher.com/order/catalog/product/A-11009>
 A-21210: <https://www.thermofisher.com/order/catalog/product/A-21210>
 A-11008: <https://www.thermofisher.com/order/catalog/product/A-11008>

A-21211: <https://www.thermofisher.com/order/catalog/product/A-21211>
 A-11037: <https://www.thermofisher.com/order/catalog/product/A-11037>

Guinea Pig Antibodies

A-11073: <https://www.thermofisher.com/order/catalog/product/A-11073>
 A-11075: <https://www.thermofisher.com/order/catalog/product/A-11075>
 A-21449: <https://www.thermofisher.com/order/catalog/product/A-21449>
 A-21450: <https://www.thermofisher.com/order/catalog/product/A-21450>
 A-11075: <https://www.thermofisher.com/order/catalog/product/A-11075>
 A-11073: <https://www.thermofisher.com/order/catalog/product/A-11073>
 A-21450: <https://www.thermofisher.com/order/catalog/product/A-21450>
 A-21449: <https://www.thermofisher.com/order/catalog/product/A-21449>

Rabbit Antibodies

A-11008: <https://www.thermofisher.com/order/catalog/product/A-11008>
 A-11034: <https://www.thermofisher.com/order/catalog/product/A-11034>
 A-21428: <https://www.thermofisher.com/order/catalog/product/A-21428>
 A-21207: <https://www.thermofisher.com/order/catalog/product/A-21207>
 A-11011: <https://www.thermofisher.com/order/catalog/product/A-11011>
 A-21245: <https://www.thermofisher.com/order/catalog/product/A-21245>
 A-21245: <https://www.thermofisher.com/order/catalog/product/A-21245>
 A-21207: <https://www.thermofisher.com/order/catalog/product/A-21207>

Goat Antibodies

A-11055: <https://www.thermofisher.com/order/catalog/product/A-11055>
 A-11034: <https://www.thermofisher.com/order/catalog/product/A-11034>
 A-21432: <https://www.thermofisher.com/order/catalog/product/A-21432>
 A-11039: <https://www.thermofisher.com/order/catalog/product/A-11039>
 A-11056: <https://www.thermofisher.com/order/catalog/product/A-11056>
 A-11040: <https://www.thermofisher.com/order/catalog/product/A-11040>
 A-21435: <https://www.thermofisher.com/order/catalog/product/A-21435>
 A-11041: <https://www.thermofisher.com/order/catalog/product/A-11041>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Lines UCLA1 (female, aka NIHhESC-10-0058) and UCLA6 (male, aka NIHhESC-11-0089) were obtained from the UCLA Stem Cell Core, CIRM 20107 (female) was obtained from the CIRM Fujifilm repository, and KOLF was obtained from Jackson Laboratories. HEK-293T cells were purchased from Takara (Lenti-X™ 293T Cell Line, Cat # 632180).
Authentication	Each pluripotent cell line was karyotyped prior to use in the lab and was validated by the UCLA Stem Cell Core or the CIRM Repository prior to distribution for authenticity. Lenti-X™ 293T Cell Lines are validated as per manufacturer website: https://www.takarabio.com/documents/Certificate%20of%20Analysis/632180/632180-120716.pdf .
Mycoplasma contamination	All lines tested negative for mycoplasma and all newly generated data used in this study was also aligned to a mycoplasma genome to verify the lack of contamination.
Commonly misidentified lines (See ICLAC register)	None

Plants

Seed stocks	<i>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.</i>
Novel plant genotypes	<i>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.</i>
Authentication	<i>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, mosaicism, off-target gene editing) were examined.</i>