Article

Cell-type-resolved mosaicism reveals clonal dynamics of the human forebrain

https://doi.org/10.1038/s41586-024-07292-5

Received: 22 October 2023

Accepted: 11 March 2024

Published online: 10 April 2024

Check for updates

Changuk Chung[1,2,11], Xiaoxu Yang[1,2,3,11], Robert F. Hevner[4,5], Katie Kennedy[6], Keng Ioi Vong[1,2], Yang Liu[1,2], Arzoo Patel[1,2], Rahul Nedunuri[1,2], Scott T. Barton[7], Geoffroy Noel[8], Chelsea Barrows[1,2], Valentina Stanley[1,2], Swapnil Mittal[1,2], Martin W. Breuss[9], Johannes C. M. Schlachetzki[1,10], Stephen F. Kingsmore[2] & Joseph G. Gleeson[1,2]✉

Debate remains around the anatomical origins of specific brain cell subtypes and lineage relationships within the human forebrain[1–7]. Thus, direct observation in the mature human brain is critical for a complete understanding of its structural organization and cellular origins. Here we utilize brain mosaic variation within specific cell types as distinct indicators for clonal dynamics, denoted as cell-type-specific mosaic variant barcode analysis. From four hemispheres and two different human neurotypical donors, we identified 287 and 780 mosaic variants, respectively, that were used to deconvolve clonal dynamics. Clonal spread and allele fractions within the brain reveal that local hippocampal excitatory neurons are more lineage-restricted than resident neocortical excitatory neurons or resident basal ganglia GABAergic inhibitory neurons. Furthermore, simultaneous genome transcriptome analysis at both a cell-type-specific and a single-cell level suggests a dorsal neocortical origin for a subgroup of DLX1$^+$ inhibitory neurons that disperse radially from an origin shared with excitatory neurons. Finally, the distribution of mosaic variants across 17 locations within one parietal lobe reveals that restriction of clonal spread in the anterior–posterior axis precedes restriction in the dorsal–ventral axis for both excitatory and inhibitory neurons. Thus, cell-type-resolved somatic mosaicism can uncover lineage relationships governing the development of the human forebrain.

Forebrain development is under the control of morphogens and transcription factors that mediate patterning of three-dimensional structures including the hippocampus, cortex and basal ganglia[8–12]. Although clonal dynamics in the mouse forebrain have been investigated using single-cell viral barcoding[13,14], studies in humans are limited to clonal relationships at broad spatial levels[15–18]. Given that the human brain comprises diverse cell types originating from various sources that intermingle, and eventually reside in proximal locations, direct mapping of clonal dynamics within distinct human forebrain cell types is essential.

Although most forebrain cells are thought to originate from radial glia that line the telencephalic lateral ventricles[19–21], observations in rodents instead have suggested ventral telencephalic progenitors as a source of GABAergic cortical inhibitory neurons[1], supported by subsequent investigations in non-human primate and human fetal tissue[3,5,13,22–24]. However, conflicting findings have suggested a recently evolved, potentially primate-specific dorsal telencephalic source of inhibitory neurons based on marker staining or single-cell lineage tracing in cultured human fetal brain tissues[2,4,6,7,25,26]. Yet, none of these studies has directly observed lineage relationships of inhibitory neurons within the fully developed human brain, leaving this longstanding debate unresolved.

Postzygotic mutations transmit faithfully to daughter cells that distribute in mosaic patterns, referred to as mosaic variants. The human brain, like other tissues, acquires mosaic variants in part due to rapid expansions of initial founder cell pools[16,17], resulting in clonal lineages sharing mosaic variants that can vary in bulk allele fraction (a portion of alternative alleles among total alleles) depending on cell mixing. As neurogenesis predominantly occurs during brain development, the distribution of mosaic variants in adults using mosaic variant barcode analysis (MVBA) can reveal clonal dynamics and lineage relationships that probably originated during embryogenesis[16–18].

Such approaches face challenges when dealing with small cell populations, such as cortical inhibitory neurons, due to technical limitations in obtaining adequate sample quantity from postmortem human tissues, and in deriving high-quality mosaic variant call sets. To overcome these challenges, we developed a methanol-fixed nuclei sorting (MFNS) protocol, which was added to previous protocols for cell-type-specific MVBA in bulk, sorted nuclei and single nuclei, allowing DNA isolation from high-quality intact nuclei suitable for library preparation.

[1]Department of Neurosciences, University of California San Diego, La Jolla, CA, USA. [2]Rady Children's Institute for Genomic Medicine, San Diego, CA, USA. [3]Department of Human Genetics, University of Utah, Salt Lake City, UT, USA. [4]Sanford Consortium for Regenerative Medicine, La Jolla, CA, USA. [5]Department of Pathology, UC San Diego School of Medicine, University of California, San Diego, La Jolla, CA, USA. [6]BioSkryb Genomics Inc., Durham, NC, USA. [7]Division of Medical Education, School of Medicine, University of California, San Diego, La Jolla, CA, USA. [8]Division of Anatomy, School of Medicine, University of California, San Diego, La Jolla, CA, USA. [9]Department of Pediatrics, Section of Genetics and Metabolism, University of Colorado School of Medicine, Aurora, CO, USA. [10]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA. [11]These authors contributed equally: Changuk Chung, Xiaoxu Yang. ✉e-mail: jogleeson@health.ucsd.edu

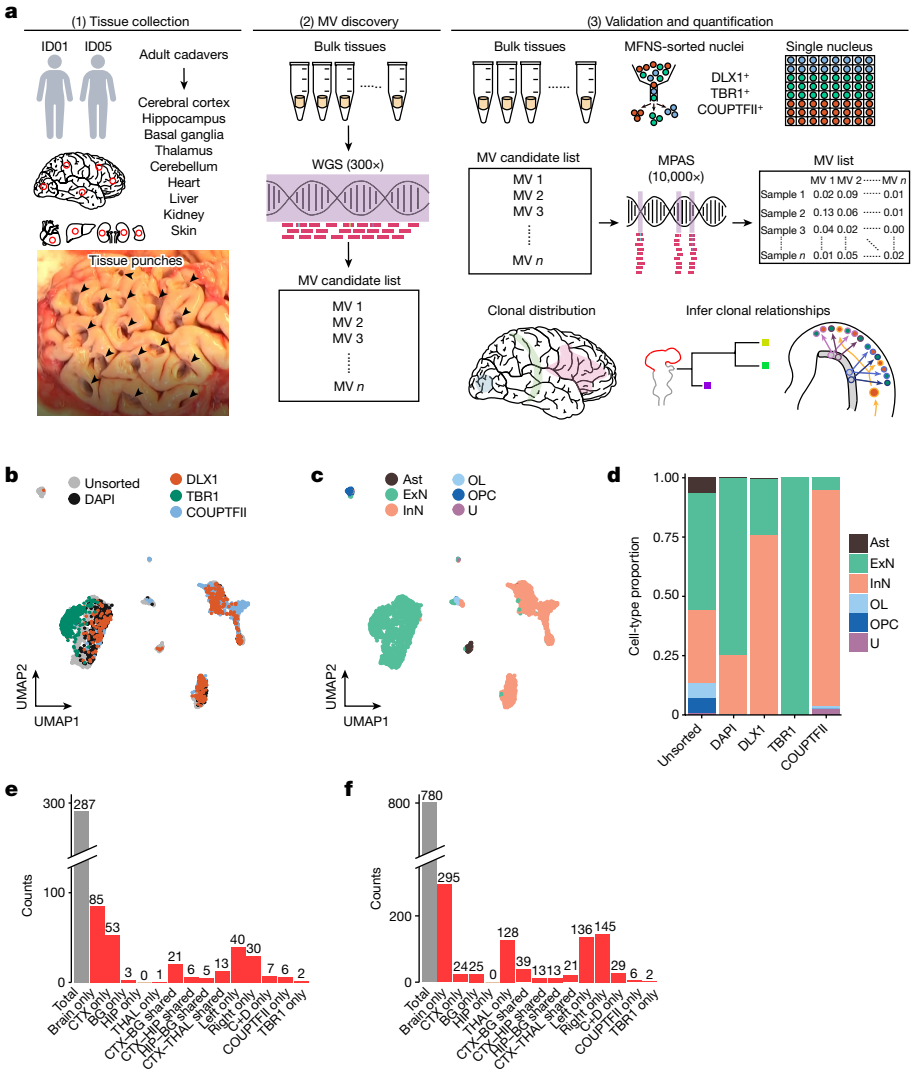


Fig. 1 | Comprehensive cMVBA identifies cell-type-resolved and region-specific mosaic variants. a, cMVBA workflow overview consists of three phases: (1) tissue collection, which involves cadaveric organs accessed for tissue punches from the organs listed (the red circles and black arrowheads indicate punch locations with 8 mm diameter in the organs and frontal lobe) for mosaic variant (MV) detection. (2) A subset of the bulk tissue punches undergoes 300× whole-genome sequencing (WGS) followed by best-practice MV calling pipelines to generate a list of MV candidates. (3) DNA from each punch bulk tissue, MFNS samples or individual nuclei are subjected to validation and quantification of MV candidates via MPAS. Allele fractions of the validated MVs from MPAS are used to determine clonal dynamics of different neuronal cell types and reconstruct features of brain development. Schematics in panel **a** were created with BioRender (https://biorender.com). **b**, Uniform manifold approximation and projection (UMAP) plot from single-nucleus RNA sequencing with MFNS-sorted or MFNS-unsorted cortical nuclear pools ($n = 3,322$). Sorted nuclear groups are labelled with distinct colours. **c**, Cortical cell types based on marker expression and differentially expressed genes in each cluster. Ast, astrocyte; ExN, excitatory neuron; InN, inhibitory neuron; OL, oligodendrocyte; OPC, oligodendrocyte precursor cell; U, undefined. **d**, Cell-type proportion within each sorted cortical nuclear population. **e**,**f**, Number of MVs categorized by location detected in donor ID01 (**e**) or ID05 (**f**) (Supplementary Data 4). 'Brain-only' MVs (that is, detected only in brain tissue, but not other organs) including subtypes are in red. 'C+D only' refers to brain-only MVs exclusively detected in both COUPTFII$^+$ and DLX1$^+$ populations but not the other cell types. BG, basal ganglia; CTX, cortex; HIP, hippocampus; THAL, thalamus.

In addition, we utilized a recently developed single-cell multi-omics approach to generate DNA genotypes and RNA transcriptomes from the same cell[27], deconvolving lineages and mapping cell-type-specific clonal dynamics within the human brain.

Identification of brain mosaic variants

Deep sequencing (300×) of a single biopsy detects dozens of clonal mosaic variants including single-nucleotide variants and small insertions or deletions (indels)[28,29]. To examine genomic relationships across distinct cell types in human brains, we further improved a previous protocol, requiring at least 50,000 nuclei[17], for lower cell number input (more than 200 nuclei) and greater cell-type diversity through MFNS, termed cell-type-specific MVBA (cMVBA), comprising three phases (Fig. 1a; Methods). In the 'tissue collection' phase, we revisited a subset of samples collected from a previously published donor ID01 (ref. 17). We also newly biopsied (8-mm punch) each lobe of the neocortex, basal ganglia, hippocampus, thalamus and cerebellum from ID01 or from an ascertained new donor (ID05) (Extended Data Fig. 1 and Supplementary Data 1). We also collected non-brain tissue including the heart, liver, both kidneys, adrenal and skin to define mosaic variant distribution across the body. In the 'mosaic variant discovery' phase, a subset of bulk tissues (10 and 32 tissues from ID01 and ID05, respectively; Supplementary Data 1) underwent 300× whole-genome sequencing, followed by state-of-the-art mosaic variant calling and filtering

based on established methods (Methods). In the 'validation and quantification' phase, we first prepared DNA samples extracted from bulk tissue and sorted nuclear populations, or amplified DNA from single nuclei (Supplementary Data 1). NEUN$^+$, DLX1$^+$, TBR1$^+$, NEUN$^-$/LHX2$^+$, OLIG2$^+$, NEUN$^+$/DARPP32$^+$ and PU.1$^+$ nuclear pools represented pan-neurons, GABAergic inhibitory neurons, excitatory neurons, astrocytes, oligodendrocytes, medium spiny neurons and microglia, respectively[17,30]. To isolate intact DNA from underrepresented cell types such as cortical inhibitory neurons, we implemented MFNS by screening several antibodies targeting DLX1 (a pan-inhibitory neuronal marker), TBR1 (an excitatory neuronal marker) and COUPTFII (a caudal ganglionic eminence (CGE)-derived inhibitory neuronal marker; encoded by *NR2F2*) (Methods), and confirmed that each antibody labels a particular cell type within the biopsy (Extended Data Fig. 2a–c).

We further confirmed the cell-type composition of DLX1$^+$, TBR1$^+$ and COUPTFII$^+$ nuclear pools compared with unsorted and DAPI$^+$ nuclear pools with single-nucleus RNA sequencing (Fig. 1b). Cell types of each cluster were identified based on marker expression patterns (Fig. 1c and Extended Data Fig. 2d,e). DLX1$^+$ and COUPTFII$^+$ pools contained mostly inhibitory neurons (more than 75%), and almost 100% of TBR1$^+$ nuclear pools were confirmed as excitatory neurons (Fig. 1d). Of note, COUPTFII$^+$ nuclei were mostly in a subcluster of inhibitory neuronal clusters, whereas DLX1$^+$ nuclei covered all inhibitory neuronal clusters, suggesting that nuclei sorted by COUPTFII antibody reflect a subset (approximately 38%) of DLX1$^+$ nuclei highly expressing *NR2F2* (Extended Data Fig. 2f). Cell-type identities of the DLX1$^+$ and TBR1$^+$ nuclei were further confirmed as inhibitory and excitatory neurons, respectively, by comparing DNA methylation patterns from reference methylomes of inhibitory and excitatory neurons across marker gene regions[31] (Extended Data Fig. 2g–j).

Using a total of 321 samples from ID01 and 147 samples (bulk or sorted nuclei) from ID05, we conducted ultra-deep massive parallel amplicon sequencing (MPAS) of each candidate mosaic variant (average coverage of approximately 10,000×). This step served two functions: (1) providing orthogonal validation for the mosaic variant within the sample, and (2) providing accurate assessment of allele fraction of each detected variant, allowing for downstream analyses (Extended Data Fig. 3a,b and Supplementary Data 2 and 3). A total of 287 and 780 mosaic variant candidates detected in whole-genome sequencing were thus positively validated and quantified in ID01 (Extended Data Fig. 3c) and ID05 (Extended Data Fig. 3d), respectively (Methods), and were subsequently used to annotate mosaic variants according to brain region and cell type (Fig. 1e,f, Extended Data Fig. 4a,b and Supplementary Data 4). Of note, we captured 2.7 times more mosaic variants in ID05 than ID01 due to the improvement of the used tissue biopsy method. We next performed mutational signature analysis using 368 brain-specific somatic single-nucleotide variants detected in ID01 or ID05 (Extended Data Fig. 4c). As expected, clock-like mutations such as signatures 1 and 5 were major components of the mutational spectrum, reflecting developmental origins. The allele fraction distribution of organ-shared mosaic variants showed absence of a peak at 25% allele fraction, implying asymmetric clonal branching during early embryonic development consistent with previous observations[32,33] (Extended Data Fig. 4d,e). The proportion of 'brain-only' mosaic variants showed a similar distribution in ID01 and ID05, accounting for 29.6% (85 of 287) and 37.8% (295 of 780) of total mosaic variants, respectively. A total of 7 and 29 mosaic variants were exclusively found in DLX1$^+$ or COUPTFII$^+$ but not in TBR1$^+$ neurons (C+D only) in ID01 and ID05, respectively. We observed similar trends of mosaic variant hemispheric restriction and microglia distribution as we and others have recently reported[17,34,35] (Extended Data Fig. 5). This suggests that the cMVBA pipeline reports anatomical-specific and cell-type-specific mosaic variants that can be used to profile clonal dynamics and reconstruct lineage relationships of specific cell types.

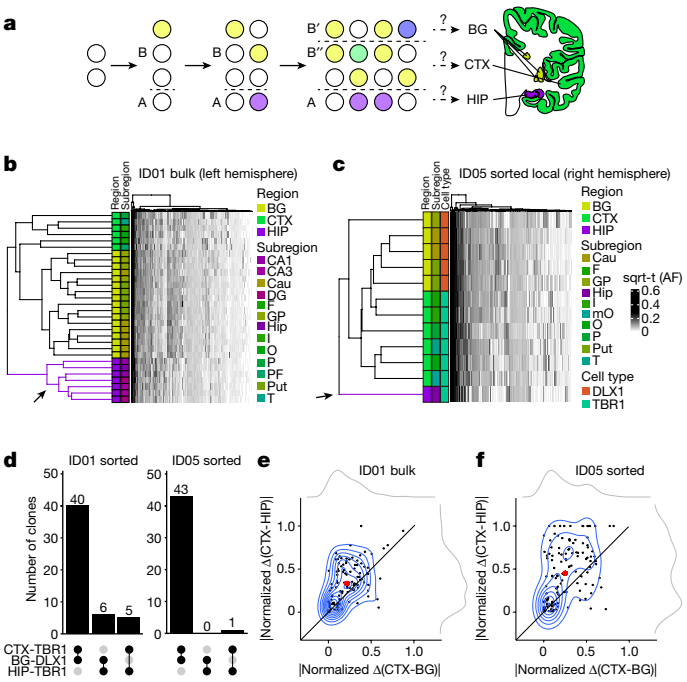


Fig. 2 | Human hippocampal lineage diverges from the cortex and basal ganglia. a, Model of clonal dynamics in forebrain anlage. Cells restricted to anlage A, which acquire a new clone (purple), are rarely present in anlage B. Later, B diverges into B′ and B″, which acquire new clones (blue and green) but still share more clones (yellow) than with anlage A. This analysis was applied to the geographies of the BG, CTX and HIP. The question marks indicate hypotheses to be examined. **b**,**c**, Heatmaps with 30 bulk samples from left hemispheric CTX, BG and HIP (*y* axis; **b**) or 12 selected sorted cell types (*y* axis; **c**), compared with MVs identified in at least two samples (146 MVs, *x* axis in **b** or 131 MVs, *x* axis in **c**), depicted in **d**. The dendrograms on the right show greater HIP lineage separation (purple, arrow) than the CTX or BG (green and yellow, respectively) either using bulk tissue (**b**) or sorted nuclei (**c**), suggesting HIP earlier lineage restriction. Cau, caudate; DG, dentate gyrus; GP, globus pallidus; Hip, hippocampal tissue in which the hippocampal subregion is not specified; I, insular cortex; mO, medial occipital cortex; O, occipital cortex; P, parietal cortex; PF, prefrontal cortex; Put, putamen; sqrt-t (AF), square-root-transformed allele fraction; T, temporal cortex. **d**, Counts of shared MVs across the CTX, BG or HIP within sorted nuclear pools showing many more shared MVs between the CTX and BG compared with the HIP in both donors (permutation $P < 0.001$). **e**,**f**, Contour plots of informative 113 and 131 MVs from **b** and **c** (blue) and two kernel density estimation plots (grey). The axes show absolute normalized allele fraction difference for each MV averaged across all samples from the respective tissues (CTX, HIP and BG). The black line is the identity line; the black dots are individual MVs; and the large red dot is the average across all MVs, suggesting that allele fraction differences are smaller between the CTX and BG than the HIP.

Genetic similarity of forebrain parts

The telencephalon derives from the most rostral part of the neural tube, subsequently committing to the cortex, basal ganglia and hippocampus, which comprise major structures of the adult forebrain. As we observed substantially fewer brain-only mosaic variants shared between the hippocampus and other telencephalic structures such as the cortex and basal ganglia (Fig. 1e,f), we hypothesized that the hippocampal founder cells are restricted in lineage compared with other brain regions (Fig. 2a). Just as clustering based on haplotype allele frequency patterns can assess genetic structures among populations[36,37], clustering of biopsied samples in an individual of mosaic variant allele fractions can be used to infer clonal relationships. We thus performed hierarchical clustering using allele fractions in bulk samples from ID01. We found that hippocampal samples strongly

clustered away from cortical and basal ganglia samples, suggesting that hippocampal progenitors are clonally more distinct from cortical or basal ganglia cells (Fig. 2b).

Forebrain structures contain heterogeneous cell types, not only derived from local progenitors but also cells migrating from distant brain regions[1,38,39]. To exclude the possibility that migrating cells contributed to these findings, we repeated hierarchical clustering, this time restricting analysis to only locally originating cell types, that is, excitatory (TBR1$^+$ nuclei) in the hippocampus or cortex and inhibitory neurons (DLX1$^+$ nuclei) in the basal ganglia. Hierarchical clustering with Manhattan distances of allele fractions in sorted nuclei samples from ID05 (Fig. 2c) and ID01 (Extended Data Fig. 6a) replicated the genomic similarity of hippocampal TBR1$^+$ clones from cortical TBR1$^+$ and basal ganglia DLX1$^+$ clones, confirming that cortical excitatory neurons show allele fraction patterns more like inhibitory neurons in the basal ganglia than excitatory neurons in the hippocampus. Furthermore, cortical TBR1$^+$ nuclei shared significantly more mosaic variants with basal ganglia DLX1$^+$ nuclei (40 and 43 for ID01 and ID05, respectively) than with hippocampal TBR1$^+$ nuclei (5 and 1 for ID01 and ID05, respectively) (Fig. 2d). The allele fraction variation was greater between the cortex and the hippocampus than between the cortex and the basal ganglia, exhibiting an average vector of data points above the identity line (Fig. 2e,f and Extended Data Fig. 6b). Together, this analysis suggests that the clonality of hippocampal progenitors is unlikely a result of migration of cells from other forebrain structures, and instead the results of locally proliferative cells within the hippocampal anlage.

Clonal dynamics of inhibitory neurons

Although in vitro analysis of neuronal progenitors from dorsal human brain cortical tissue has shown the potential to develop into inhibitory neurons[6], direct evidence for a dorsal origin of cortical inhibitory neurons in the mature human brain is lacking. cMVBA allowed for comparison of genomic similarity between different classes of cortical inhibitory neurons and other cell types. Hierarchical clustering was carried out for allele fractions measured in DLX1$^+$, TBR1$^+$ and COUPTFII$^+$ nuclear pools isolated from widespread sampling from cortical areas in both ID01 and ID05 (Fig. 3a). In four different cortical hemispheres, most of the COUPTFII$^+$ nuclear pools (that is, CGE-derived inhibitory neurons that distribute across cortical areas[6]) were exclusively clustered together, whereas most DLX1$^+$ and TBR1$^+$ nuclear pools in the same punch were clustered together (Fig. 3b,c). Bootstrap analysis further statistically validated that many of these dendrogram clusters appeared very unlikely to have arisen by chance (Extended Data Fig. 7a,b). Even after simulating removal of approximately 25% of the excitatory neuronal component from DLX1$^+$ nuclear pools (Fig. 1d) using computational deconvolution (Methods), most of DLX1$^+$ and TBR1$^+$ nuclear pools at the same cortical lobe remained clustered together apart from COUPTFII$^+$ nuclear pools (Extended Data Fig. 7c,d). Conversely, we simulated TBR1$^+$ nuclear contamination into COUPTFII$^+$ nuclear pools at a similar level (25%) and compared with DLX1$^+$ nuclear pools (Extended Data Fig. 7e,f). The 'contaminated' COUPTFII$^+$ nuclear pools continued to group together with the original COUPTFII samples, rather than clustering with the TBR1$^+$ nuclear pools in the same lobe, unlike how the deconvolved DLX1$^+$ nuclei did (Extended Data Fig. 7c,d), indicating that deconvolved DLX1$^+$ nuclei may have a fraction of dorsal clones sharing mosaic variants with local TBR1$^+$ neurons.

Next, we measured allele fraction correlations between mosaic variants within each cell type (Supplementary Data 5). As expected, mosaic variant clustering confirmed hemispheric restrictions of all three cell types (DLX1$^+$, TBR1$^+$ and COUTPFII$^+$ nuclei). In addition, mosaic variants in TBR1$^+$ nuclei showed many small subclusters enriched in particular lobes, in contrast to widespread distributions of COUPTFII$^+$ nuclei. This result was replicated in four independent hemispheres from both donors. A similar pattern was also observed in the global DLX1$^+$ nuclear pools, although with less intensity. The data suggest a wide distribution of COUPTFII$^+$ ventrally derived cortical inhibitory neuronal clones through tangential migration to the dorsal telencephalon as reported in mice[40], whereas focal distribution of TBR1$^+$ and at least some DLX1$^+$ neurons distribute radially from a dorsal telencephalic, probably radial glial, source. Collectively, if we exclude the very unlikely scenario that cortical-derived excitatory and inhibitory neurons originate from ventrally derived progenitor cells that migrate to specific locations together, the result suggests that a substantial portion of dorsal clones have the potential to differentiate into both cortical excitatory and inhibitory neurons.

Next, we investigated distributions of individual mosaic variants across cell types and cortical areas by mapping each mosaic variant allele fraction onto a 'lolliplot' and a geographical map called a 'geoclone' (Supplementary Data 6). As an example, mosaic variant 1-64512024-C-T was inhibitory neuron specific, showing notable uniform enrichment across cortical areas in COUPTFII$^+$ and DLX1$^+$ but not in TBR1$^+$ nuclear pools (Fig. 3d). This suggests that a subset of nuclei in DLX1$^+$ pools distribute in patterns similar to COUPTFII$^+$ nuclei. The mosaic variant 10-116196503-C-T was present in all three cell types, but with less allele fraction variation in COUPTFII$^+$ nuclei than in the other two cell types (Fig. 3e), suggesting a wide distribution across cortical regions and tangential migration from a ventral origin. Furthermore, the mosaic variant 16-69679204-G-C was locally enriched in both DLX1$^+$ and TBR1$^+$ nuclei of the same cortical lobes but below the level of detection in COUPTFII$^+$ nuclei, implying that a subset of DLX1$^+$ cells share locally proliferating dorsal telencephalic origins with TBR1$^+$ cells (Fig. 3f). We found some mosaic variants (that is, the mosaic variant 2-71776656-C-A) that were distinctively enriched in DLX1$^+$ nuclei distributed to both prefrontal lobes of ID01 (Fig. 3g), suggesting that some inhibitory neurons or their progenitors can populate both hemispheres. However, prefrontal samples were not available in ID05, and more data will be required to confirm these results.

To examine whether COUPTFII$^+$ clones are more uniformly distributed than DLX1$^+$ or TBR1$^+$ clones, standard deviations of allele fractions of shared mosaic variants in different cortical areas were calculated for the three neuronal populations (Fig. 3h,i). COUPTFII$^+$ nuclei standard deviations were significantly lower than the other two cell types in both ID01 and ID05, suggesting that the distribution of CGE-derived inhibitory neuronal clones is wider than excitatory neurons, whereas cortical pan-inhibitory neuronal clones showed a patchy distribution in a manner similar to excitatory neuronal clones.

We next estimated the proportion of dorsally derived cells within DLX1$^+$ nuclei, using the least squares method, to identify the proportion of dorsal and ventral origins that best approximates the allele fractions of DLX1$^+$ nuclei (Extended Data Fig. 7g). This estimation assumes that TBR1$^+$ and COUPTFII$^+$ nuclei exclusively originate from dorsal and ventral origins, respectively (Methods). The mean proportion of dorsally derived clones across lobes and individuals was 59.142% (95% confidence interval 51.037–67.247%). This implies that more than half of cortical inhibitory neurons may derive from dorsal progenitors.

Dorsal origins of inhibitory neurons

The previous data could not exclude the possibility that the observed genomic similarity between TBR1$^+$-sorted and DLX1$^+$-sorted nuclear pools derived in part from a rare cell type or from pool contaminants from the sorting protocol. We thus conducted single-cell simultaneous DNA + RNA sequencing (ResolveOME; see Methods), incorporating primary template-directed amplification coupled with single-nuclear MPAS with full-transcript single-nucleus RNA sequencing, in individual NEUN$^+$ nuclei from the right frontal and temporal cortex in ID05. This allowed for a priori identification of both cell types and mosaic variants in the same cell[27] (Fig. 4a and Supplementary Data 2 and 7). After basic quality control, uniform manifold approximation and projection

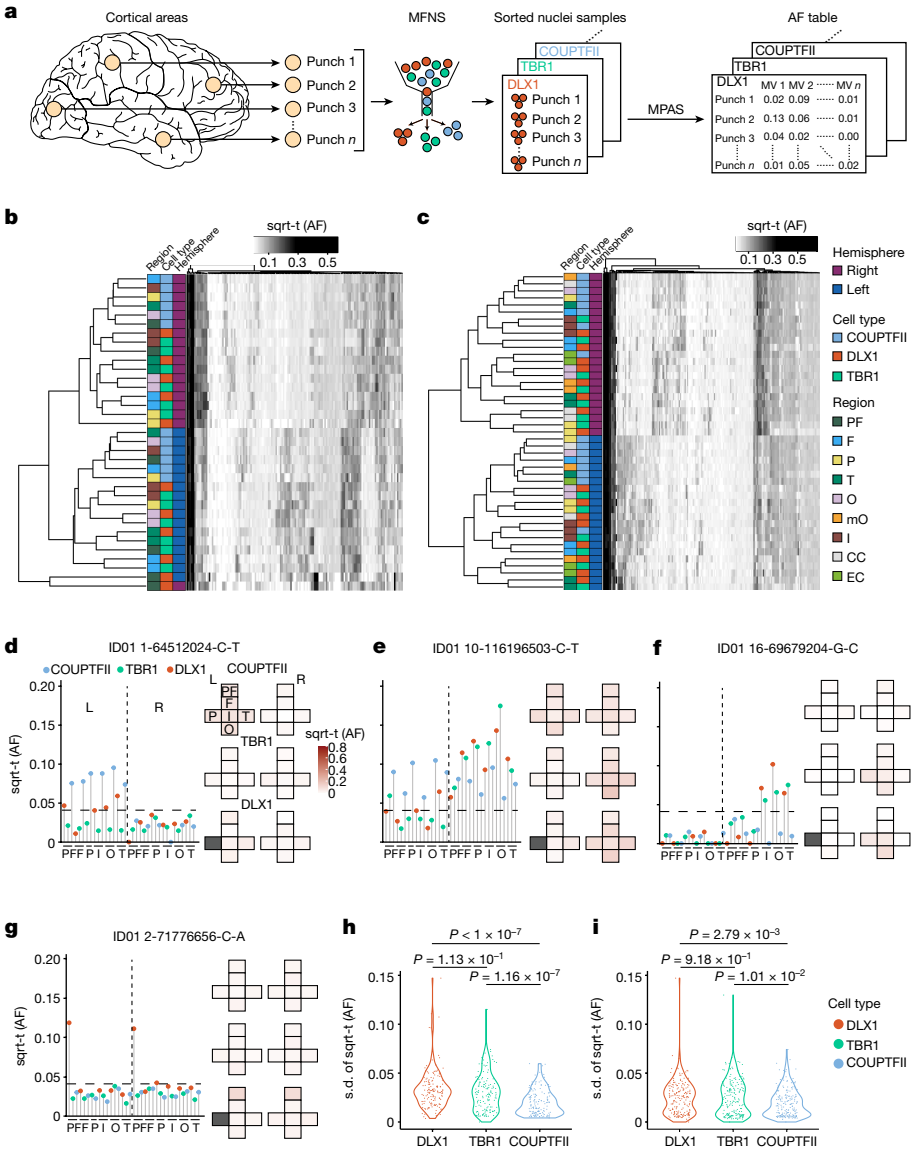


Fig. 3 | Clonal dynamics of cortical excitatory and inhibitory neurons. **a**, cMVBA workflow uses MFNS nuclei for MPAS assessment of allele fractions (AFs) in cortical punches. **b**,**c**, Heatmaps of sorted nuclei based on AFs of 146 informative shared MVs from the cortex of ID01 (**b**) and 186 from the cortex of ID05 (**c**) (*x* axis) compared with colour-coded hemisphere, cell type or region. CC, cingulate cortex; EC, entorhinal cortex; F, frontal cortex. The dendrograms indicate that subcortically derived COUPTFII$^+$ nuclear pools cluster together (teal), whereas DLX1$^+$ and TBR1$^+$ nuclear pools derived from the same region cluster together. **d**–**g**, Lolliplots comparing regions (*x* axis) with sqrt-t (AF) (*y* axis) for representative MVs. The height of the individual lollipop indicates AF; the dot colours indicate cell type; and the dashed lines indicate threshold. Next to each lolliplot is the 'geoclone' representation of sqrt-t (AF) shaded intensity (pink) from tissue where detected. The grey boxes indicate not sampled. L, left; R, right. **h**,**i**, Standard deviation (s.d.) of sqrt-t (AFs) for 146 and 186 MVs in the three different cell types in donor ID01 (**h**) and ID05 (**i**), respectively. Each dot represents a single MV measured in 34 versus 45 punches across the neocortex in ID01 and ID05, respectively. One-way ANOVA with Tukey's multiple comparison test with adjusted *P* values.

(UMAP) clustering with a reference dataset[41,42] distinguished between cortical excitatory and inhibitory neurons, along with a few minor non-neuronal cell types (Extended Data Fig. 8a–c and Supplementary Data 10). The detection frequency of mosaic variants in single-cell genotyping was positively correlated with allele fractions in sorted populations, as expected (Extended Data Fig. 8d–f). Informative mosaic variants were detected from a total of 85 excitatory and 33 inhibitory neurons, allowing direct observation of single-cell level mosaic variant distribution (Fig. 4b) with 5.863% and 1.052% of false-negative and false-positive rates, respectively (Methods).

These genotypes allowed for the assessment of shared mosaic variants in individual excitatory or inhibitory neurons in individual nuclei across various brain regions. We observed numerous excitatory and inhibitory neurons in the same lobe carrying mosaic variants exclusively in the frontal lobe (F only) or temporal lobe (T only). As an example, the mosaic variant 13-69308268-A-G, detected in both excitatory and inhibitory neurons of the right frontal but not the temporal lobe (Fig. 4b), was enriched in the TBR1$^+$ and DLX1$^+$ nuclear pools with high allele fractions but below the level of detection in the COUPTFII$^+$ nuclear pool (Fig. 4c). Furthermore, overall distribution patterns of this mosaic variant in TBR1$^+$ and DLX1$^+$ nuclei across cortical lobes were very similar to each other, but distinct from that of COUPTFII$^+$, supporting that this mosaic variant is shared between locally born and cortical resident excitatory and inhibitory neurons but underrepresented within cortical inhibitory neurons derived from ventral telencephalic sources (Fig. 4c).

We further generated a phylogeny tree by considering alleles on each of the genomic positions as 'pseudo-sequence' for each sample, and

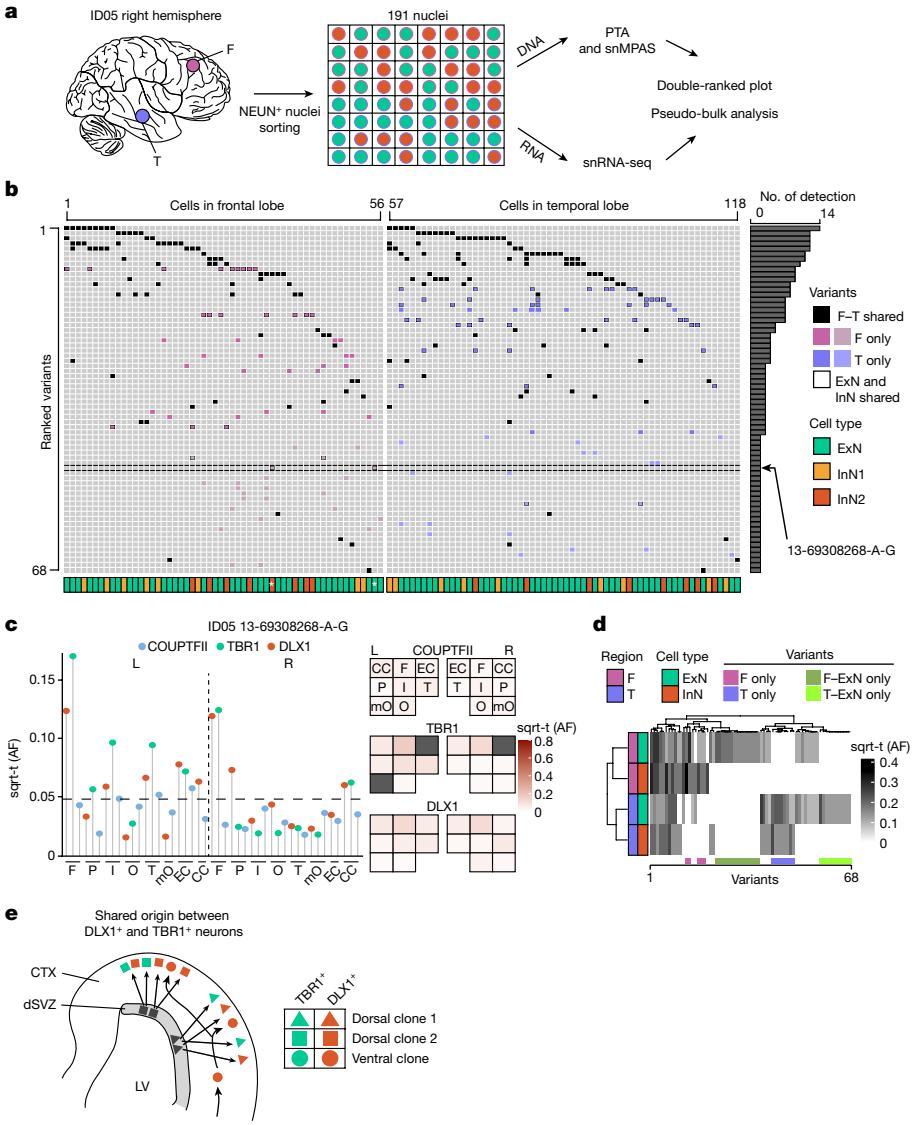


Fig. 4 | Single-nuclear MPAS incorporating single-nucleus RNA sequencing supports the existence of dorsally derived cortical inhibitory neurons in humans. a, Frontal and temporal small punches of the right hemisphere of ID05 NEUN$^+$ 191 single-nuclei DNA were subjected to primary template-directed amplification (PTA) followed by single-nuclear MPAS (snMPAS) genotyping with concurrent single-nucleus RNA sequencing (snRNA-seq), termed ResolveOME, for analysis. **b**, A double-ranked plot divided in half based on brain region. Sixty-eight MVs identified in anywhere between 2 and 14 cells (no. of detection). A total of 118 (85 excitatory and 33 inhibitory) neurons were plotted. MVs were further classified according to whether they were detected in both frontal and temporal lobes (F–T shared) versus a single-lobe F only (purple) or T only (blue). Dark purple or blue represent detection in three or more nuclei, and light purple or blue represent detection in only two nuclei. ExN and InN shared indicates MVs shared between ExNs and InNs exclusively within F-only or T-only MVs. The *x* axis represents cell type, that is, ExN (green); InN1 (orange), which is all inhibitory neurons except for InN2; and InN2 (red), which is inhibitory neurons carrying at least one ExN and InN-shared MV. For example, the MV 13-69308268-A-G (arrow) was detected in two nuclei (indicated by asterisks): an ExN and an InN. **c**, Distribution of the MV 13-69308268-A-G across cortical areas in ID05. Lolliplot (left) and geoclone showing sqrt-t (AFs) of each sorted nuclear pool in different cortical locations, reproducing single-nucleus data (right). The dashed line indicates detection threshold. The grey boxes indicate data are not available. **d**, Pseudo-bulk analysis after aggregation based on individual MVs (*x* axis) and cell types and regions (*y* axis). **e**, Model for the shared origin of local ExN and InN cortical neurons. Dorsal clones (triangle and square) can produce both TBR1$^+$ excitatory neurons and DLX1$^+$ inhibitory neurons. Ventrally derived DLX1$^+$ inhibitory neurons (circle) tangentially migrate and are more likely to disperse across the cortex. dSVZ, dorsal subventricular zone; LV, lateral ventricle.

a sequence-based phylogenic tree was reconstructed to deconvolve the clonal relationship between single cells (Extended Data Fig. 9; Methods). Although the overall structure was not very stable from the bootstrap, for the closest branch that has the highest bootstrap values between two samples, which in general has higher bootstrap values, we found 3.5 times more ExN–InN pairs in the same lobe (14 pairs) than those in different lobes (4 pairs). This result supports evidence that a substantial proportion of human cortical excitatory and inhibitory neurons derive from related cortical progenitor cells.

We also calculated the probability of observing cortical inhibitory neurons carrying seemingly locally enriched mosaic variants by chance using mosaic variants shared in more than two cells in one lobe but none in the other lobe. Among the 33 inhibitory neurons analysed, we identified 15 with distinct local mosaic variants. These mosaic variants were found exclusively in one lobe and were shared with at least two other local cells, including one excitatory neuron within the same lobe. This observation was highly unlikely to occur by chance (one-tailed permutation test $P < 0.0001$; Extended Data Fig. 8g; Methods), suggesting the

existence of locally derived inhibitory neurons from progenitor cells that also produce local excitatory neurons.

Inhibitory neurons carrying locally enriched mosaic variants (InN2; Fig. 4b) showed comparable inhibitory neuronal marker expression (Extended Data Fig. 8h) but a decreased tendency for RNA expression of CGE markers (Extended Data Fig. 8i) or a RELN$^+$ neuronal marker (Extended Data Fig. 8j) compared with those not carrying locally enriched mosaic variants (InN1). Instead, InN2 displayed an increased tendency for a parvalbumin$^+$ inhibitory neuronal marker (Extended Data Fig. 8k) and some unique transcription patterns (Extended Data Fig. 8l,m). Thus, a substantial portion of dorsally derived inhibitory neurons may contribute to parvalbumin$^+$ neurons, whereas COUPTFII$^+$ neurons contribute to ventrally derived inhibitory neurons of other classes.

We further observed 26 mosaic variants specific to only local excitatory neurons but not inhibitory neurons in the same lobe (Fig. 4b), implying that a subset of fate-restricted dorsal telencephalic neural progenitors generate mostly excitatory neurons. This was also supported by pseudo-bulk analysis, by aggregating cells based on each cell type and region, demonstrating local excitatory neuron-specific mosaic variants (Fig. 4d). We could not find evidence of mosaic variants specific to cortical inhibitory neurons within anatomically defined regions, which may be due to their sparse population from total cortical cells, or a limited number of inhibitory neuron-specific mosaic variants. Together, single-cell-level genotyping incorporating transcriptomics supports the concept that dorsal telencephalic neural progenitor cells may have the potential to generate both excitatory and inhibitory neurons, even among the progenitor pool predominantly generating excitatory neurons (Fig. 4e).

Anterior–posterior restriction in a lobe

We used this same approach to study clonal dynamics within a single human cerebral lobe. Previous data have suggested that clonality between lobes is more restricted than within a lobe[16]. Our previous data have indicated that a restriction of clonal spread (RCS) along the anterior–posterior axis follows the establishment of the midline RCS, but did not consider cell-type-specific effects[17]. We thus assessed clonal dynamics of DLX1$^+$ and TBR1$^+$ cells, selecting the parietal lobe for analysis. We performed high-density biopsies from a total of 17 small punches offset by 1-cm distances followed by MFNS and MPAS genotyping, capable of distinguishing between anterior–posterior and dorsal–ventral RCSs with samples clustered based on allele fractions of informative mosaic variants (Fig. 5a,b and Supplementary Data 8). Of note, the main clusters (C1 and C2) were formed along the anterior–posterior rather than the dorsal–ventral axis in both cell types in hierarchical clustering (Fig. 5b,c) and UMAP plots (Extended Data Fig. 10). The RCS dominated along the anterior–posterior over the dorsal–ventral axis in both cell types, represented by the mosaic variants 18-33999883-C-T, 4-10646818-G-A and 3-172635725-G-A (Fig. 5d). Absolute values of normalized allele fraction difference through the anterior–posterior axis were larger than those through the dorsal–ventral axis (Fig. 5e,f). These results suggest that the anterior–posterior axis RCS is established before the dorsal–ventral axis RCS in both cell types.

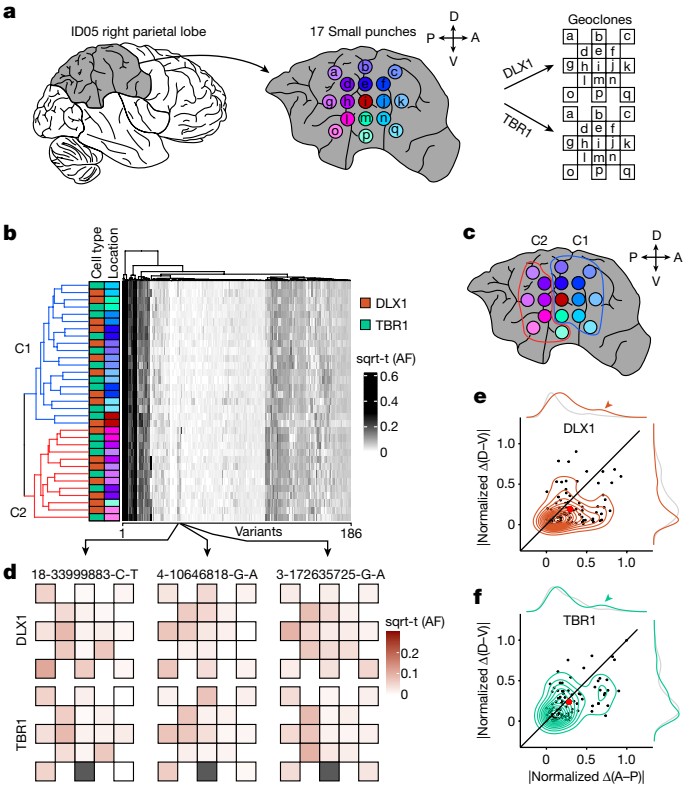


Fig. 5 | Earlier establishment of the anterior–posterior axis compared with the dorsal–ventral RCS within a cortical lobe. a, Workflow for the observation of clonal dynamics in a lobe. A total of 17 punches were radially sampled and subjected to MFNS to assess MVs. The AF of MVs in different sites were mapped onto the geoclones (checkerboard). A, anterior; D, dorsal; P, posterior; V, ventral. **b**, Heatmap and dendrogram hierarchical cluster of sorted nuclear pools based on sqrt-t (AF)s of 186 informative MVs from ID05 detected in the right parietal lobe. The sidebars on the left of the heatmap indicate cell type and sample location information, according to the convention in Fig. 3. The colours for lobar location are taken from panel **a**. **c**, The dendrogram highlights two main clusters (C1 in blue; C2 in red) that when mapped back onto the sampled spatial coordinates are separated in the anterior–posterior dimension (red and blue circles). **d**, Geoclones of three individual MVs from **b** (box, arrows), showing that shades of pink are more different in the anterior–posterior axis than in the dorsal–ventral axis. The grey boxes indicate data are not available. **e**,**f**, Cell-type-resolved contour plots of 94 shared MVs from **b** (centre) with coloured kernel density estimation plots (in the periphery) for DLX1 (**e**) or TBR1 (**f**), showing MVs from both cell types with a greater normalized difference of sqrt-t (AFs) in the anterior–posterior than in dorsal–ventral axis. Grey indicates a kernel density estimation plot of the orthogonal axis. The arrowheads indicate the local peak of the density estimation plot due to MVs with greater difference of AFs in the anterior–posterior than in the dorsal–ventral axis. The black line indicates the identity line. The dots represent individual MVs. The large red dot indicates the average across all MVs.

Discussion

Here, using clonal dynamics, we interrogated cellular origins across the neurotypical mature human forebrain. Previous studies have relied on the dynamics of single-cell transcriptomic or epigenomic profiles and were not able to reconstruct cellular lineages among human forebrain cell types based on clonality. We found that mosaic variants originating in locally born cellular populations demonstrate stronger lineage restriction within the hippocampus than restriction to either the neocortex or basal ganglia. This is consistent with previous viral barcode tracing in mice showing hippocampal lineage restriction before neural tube closure at E9.5, at a time even before *Prox1* expression[14,43]. We hypothesize that hippocampal lineage restriction may be complete by the time anterior neural plate boundary expresses WNT3A that defines the cortical hem anlage at post-conception week 6 in humans[44,45]. It is also possible that the pallial and subpallial lineage restriction to future subcortical and cortical structures occurs after anterior neural plate boundary definition at the time of dorsal and ventral axis differentiation under the control of BMP signalling[46,47]. Future studies might assess if the early hippocampal lineage restriction that we observed here occurs in other vertebrate species, where

WNT and BMP-mediated neurulation is well conserved[48–50]. From the clinical perspective, this restriction may explain why some hemimegalencephaly cases show structural abnormalities in both the cortex and the basal ganglia[51]. More investigations on the distribution of dysplastic cells across the hippocampus, cortex and basal ganglia can provide insights into brain dysplasias.

Although previous studies in mammals have suggested most or all cortical inhibitory neurons derive from the ventral telencephalon in mammals[3,5,13,22–24], subsequent studies in non-human primates, the human fetal brain or stem cell culture have suggested a potential primate-specific dorsal pallial contribution[6,7]. Our cMVBA data in the mature human brain provide direct evidence for such a dorsal source of cortical inhibitory neurons. Furthermore, both cortical excitatory and inhibitory neurons showed similar clonal mosaic variant dynamics within the same lobe, whereas some excitatory neurons had additional local mosaic variants not detected in inhibitory neurons at the single-cell level. This implies that early dorsal progenitors can produce both cell types. However, further fate restriction may occur in some cell populations or at certain times in development that produce exclusively excitatory or inhibitory neurons within the same location. Future studies with additional donors could better generalize our main findings and define the timing of this lineage restriction.

The cell-type-resolved spatial clonal relationship within single cortical lobes was not previously studied in detail due to spatial resolution limits of previous methods. Here, using MFNS, we describe the RCS of excitatory and inhibitory neurons along the anterior–posterior before the dorsal–ventral axis within a lobe. This result implies that most cortical neural progenitors maintain the potential to produce excitatory and inhibitory neurons even after a lobe is clonally restricted along the anterior–posterior axis. From a clinical perspective, examining whether dysplastic cells within a lobe are more likely to spread along the dorsal–ventral axis than the anterior–posterior axis in paediatric focal epilepsy may provide valuable insights for refining surgical strategies.

This study also incorporates several technical advances. We substantially improved experimental and computational mosaic variant detection pipelines, which enabled high-quality libraries to be generated from low-abundance cell types. We used MFNS, which used non-formaldehyde fixative to preserve DNA quality, and modified MPAS, which now incorporates an on-bead first amplification step. The incorporation of mosaic variant calling software based on deep learning identified mosaic variants that conventional mosaic variant callers may have overlooked. These protocols increased mosaic variant discovery efficiency by an estimated 250-fold, requiring substantially less materials, and providing remarkably improved spatial resolution with ultra-high-depth (10,000×) measured allele fractions across 17 different locations in a single lobe. In addition, we incorporated the ResolveOME approach to simultaneously profile single-cell RNA and DNA from archived tissue. Although we estimate a false-positive rate of our MPAS validation method of 5% (Methods), and may be locus specific, we cannot fully exclude that some discovered mosaic variants could be false. Despite this caution, these approaches could support future investigations into clonal relationships in human tissues at the single-cell level.

Despite our findings, there remain notable disparities when compared with in vitro human inhibitory neuron studies in which cortical inhibitory neurons, derived from human dorsal forebrain organoid outer radial glia, exhibited elevated *NR2F2* expression[6,7]. By contrast, our study suggests that a substantial number of COUPTFII$^+$ cortical clones displayed dispersed clonal dynamics, arguing against a purely radial distribution. This discrepancy could reflect differences between in vitro and in vivo experimental conditions and underscores the need for further investigation of cortical inhibitory neuron diversity.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07292-5.

1. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes. *Science* **278**, 474–476 (1997).
2. Letinic, K., Zoncu, R. & Rakic, P. Origin of GABAergic neurons in the human neocortex. *Nature* **417**, 645–649 (2002).
3. Wonders, C. P. & Anderson, S. A. The origin and specification of cortical interneurons. *Nat. Rev. Neurosci.* **7**, 687–696 (2006).
4. Petanjek, Z., Berger, B. & Esclapez, M. Origins of cortical GABAergic neurons in the cynomolgus monkey. *Cereb. Cortex* **19**, 249–262 (2009).
5. Hansen, D. V. et al. Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat. Neurosci.* **16**, 1576–1587 (2013).
6. Delgado, R. N. et al. Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* **601**, 397–403 (2022).
7. Andrews, M. G. et al. LIF signaling regulates outer radial glial to interneuron fate during human cortical development. *Cell Stem Cell* **30**, 1382–1391.e5 (2023).
8. Bulfone, A. et al. Spatially restricted expression of Dlx-1, Dlx-2 (Tes-1), Gbx-2, and Wnt-3 in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries. *J. Neurosci.* **13**, 3155–3172 (1993).
9. Puelles, L. & Rubenstein, J. L. Forebrain gene expression domains and the evolving prosomeric model. *Trends Neurosci.* **26**, 469–476 (2003).
10. Furuta, Y., Piston, D. W. & Hogan, B. L. Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. *Development* **124**, 2203–2212 (1997).
11. Grove, E. A., Tole, S., Limon, J., Yip, L. & Ragsdale, C. W. The hem of the embryonic cerebral cortex is defined by the expression of multiple Wnt genes and is compromised in Gli3-deficient mice. *Development* **125**, 2315–2325 (1998).
12. Monuki, E. S., Porter, F. D. & Walsh, C. A. Patterning of the dorsal telencephalon and cerebral cortex by a roof plate-Lhx2 pathway. *Neuron* **32**, 591–604 (2001).
13. Bandler, R. C. et al. Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* **601**, 404–409 (2022).
14. Ratz, M. et al. Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. *Nat. Neurosci.* **25**, 285–294 (2022).
15. Dang, H. et al. Monoclonal antibody specific to acid phosphatase isoenzyme 4. *Prostate* **9**, 47–55 (1986).
16. Bizzotto, S. et al. Landmarks of human embryonic development inscribed in somatic mutations. *Science* **371**, 1249–1253 (2021).
17. Breuss, M. W. et al. Somatic mosaicism reveals clonal distributions of neocortical development. *Nature* **604**, 689–696 (2022).
18. Park, S. et al. Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature* **597**, 393–397 (2021).
19. Rakic, P. Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.* **145**, 61–83 (1972).
20. Kriegstein, A. R. & Noctor, S. C. Patterns of neuronal migration in the embryonic cortex. *Trends Neurosci.* **27**, 392–399 (2004).
21. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
22. Wichterle, H., Turnbull, D. H., Nery, S., Fishell, G. & Alvarez-Buylla, A. In utero fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain. *Development* **128**, 3759–3771 (2001).
23. Ma, T. et al. Subcortical origins of human and monkey neocortical interneurons. *Nat. Neurosci.* **16**, 1588–1597 (2013).
24. Arshad, A. et al. Extended production of cortical interneurons into the third trimester of human gestation. *Cereb. Cortex* **26**, 2242–2256 (2016).
25. Alzu'bi, A. et al. The transcription factors COUP-TFI and COUP-TFII have distinct roles in arealisation and GABAergic interneuron specification in the early human fetal relencephalon. *Cereb. Cortex* **27**, 4971–4987 (2017).
26. Alzu'bi, A. et al. Distinct cortical and sub-cortical neurogenic domains for GABAergic interneuron precursor transcription factors NKX2.1, OLIG2 and COUP-TFII in early fetal human telencephalon. *Brain Struct. Funct.* **222**, 2309–2328 (2017).
27. Marks, J. R. et al. Unifying comprehensive genomics and transcriptomics in individual cells to illuminate oncogenic and drug resistance mechanisms. Preprint at *bioRxiv* https://doi.org/10.1101/2022.04.29.489440 (2023).
28. Lodato, M. A. et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).
29. Rodin, R. E. et al. The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. *Nat. Neurosci.* **24**, 176–185 (2021).
30. Ernst, A. et al. Neurogenesis in the striatum of the adult human brain. *Cell* **156**, 1072–1083 (2014).
31. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
32. Ju, Y. S. et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* **543**, 714–718 (2017).
33. Fasching, L. et al. Early developmental asymmetries in cell lineage trees in living individuals. *Science* **371**, 1245–1248 (2021).
34. Ginhoux, F. & Garel, S. The mysterious origins of microglia. *Nat. Neurosci.* **21**, 897–899 (2018).

35. Prinz, M., Jung, S. & Priller, J. Microglia biology: one century of evolving concepts. *Cell* **179**, 292–311 (2019).
36. Gilbert, E., Shanmugam, A. & Cavalleri, G. L. Revealing the recent demographic history of Europe via haplotype sharing in the UK Biobank. *Proc. Natl Acad. Sci. USA* **119**, e2119281119 (2022).
37. Diaz-Papkovich, A., Anderson-Trocme, L. & Gravel, S. A review of UMAP in population genetics. *J. Hum. Genet.* **66**, 85–91 (2021).
38. Xu, Q., Cobos, I., De La Cruz, E., Rubenstein, J. L. & Anderson, S. A. Origins of cortical interneuron subtypes. *J. Neurosci.* **24**, 2612–2622 (2004).
39. Miyoshi, G. et al. Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons. *J. Neurosci.* **30**, 1582–1594 (2010).
40. Mayer, C. et al. Clonally related forebrain interneurons disperse broadly across both functional areas and structural boundaries. *Neuron* **87**, 989–998 (2015).
41. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
42. Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
43. Oliver, G. et al. Prox 1, a prospero-related homeobox gene expressed during mouse development. *Mech. Dev.* **44**, 3–16 (1993).
44. LaBonne, C. & Bronner-Fraser, M. Neural crest induction in *Xenopus*: evidence for a two-signal model. *Development* **125**, 2403–2414 (1998).
45. Saint-Jeannet, J. P., He, X., Varmus, H. E. & Dawid, I. B. Regulation of dorsal fate in the neuraxis by Wnt-1 and Wnt-3a. *Proc. Natl Acad. Sci. USA* **94**, 13713–13718 (1997).
46. Faure, S., de Santa Barbara, P., Roberts, D. J. & Whitman, M. Endogenous patterns of BMP signaling during early chick development. *Dev. Biol.* **244**, 44–65 (2002).
47. Stuhlmiller, T. J. & Garcia-Castro, M. I. Current perspectives of the signaling pathways directing neural crest induction. *Cell. Mol. Life Sci.* **69**, 3715–3737 (2012).
48. Bingman, V. P., Salas, C. & Rodriguez, F. in *Encyclopedia of Neuroscience* (eds Binder, M. D., Hirokawa, N. & Windhorst, U.) 1356–1360 (Springer Berlin Heidelberg, 2009).
49. Grillner, S., Robertson, B. & Stephenson-Jones, M. The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *J. Physiol.* **591**, 5425–5431 (2013).
50. Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B. & Grillner, S. Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Curr. Biol.* **21**, 1081–1091 (2011).
51. Sepulveda, W., Sepulveda, F., Schonstedt, V., Stern, J. & Diaz-Serani, R. Neuroimaging findings in fetal hemimegalencephaly: case study and review. *Fetal Diagn. Ther.* https://doi.org/10.1159/000535406 (2023).

Methods

Donor recruitment

Organs of ID05 were collected from the UC San Diego Anatomical Material Program (UCSD-21-160). Organs of ID01 were donated from a female individual 70 years of age, whose cause of death was 'global geriatric decline' with a contributing cause of 'post-surgical malabsorption' as documented[17]. Organs of ID05 were donated from a female individual 73 years of age, whose medical history indicated 'knee replacement, fractured pelvis, hernia, fractured fibula, hypothyroidism, empyema, pulmonary arterial hypertension and scleroderma'. Both donors were documented to be of European ancestry. Organs were collected within a 26-hour postmortem interval for both donors (ID01: 24 hours, ID05: 26 hours). Previous medical history showed no signs of neurological, psychiatric or cancer diseases for either, and tested negative for infection with HIV, hepatitis B or COVID-19.

According to 45 CFR 46.102(e)(1), the use of human anatomical cadaver specimens of ID01 and ID05 is exempt from oversight of the University of California, San Diego Human Research Protections Program (IRB) but are subjected to oversight by the University of California, San Diego Anatomical Materials Review Committee (AMRC). This study was overseen and approved by the AMRC. The approval number is 106135. Donors met AMRC qualifications: (1) obtain information or biospecimens through intervention or interaction with the individual, and uses, studies or analyses the information or biospecimens; or (2) obtains, uses, studies, analyses or generates identifiable private information or identifiable biospecimens.

Tissue dissection

All dissection was performed by an anatomical pathologist or neuropathologist. For dissection to capture mosaic variants from ventral telencephalic progenitors of ID01, archived frozen limbic system parts of both hemispheres were sliced at 1-cm thickness, and tissues from the caudate, putamen, amygdala, globus pallidus, hippocampus (CA1–CA3, and dentate gyrus wherever distinguishable) and thalamus were obtained. For ID05, after the removal of the meninges, approximately 500 mg of tissues from cerebral cortical regions, caudate, putamen, globus pallidus, hippocampus, cerebellum, heart, liver, adrenals, kidneys and leg skin samples were collected before freezing. Extensive 10–17 sublobar punch biopsies were collected from the right occipital lobe and the right parietal lobe from ID05 with an 8-mm skin punch. Sample information is summarized in Extended Data Fig. 1 and Supplementary Data 1. The dissection procedure was conducted on ice at room temperature for ID01 and in a cold room for ID05. During dissection, subsamples and the remnants of the large pieces were immediately labelled and snap-frozen on dry ice, and stored at −80 °C.

DNA extraction of bulk tissue

Small cortical biopsies were first cut in half on dry ice. Half of the biopsy was stored as backup and partly used for single nuclei fluorescence-activated nuclei sorting (FANS) for ID01 and ID05. The other half of the cortical biopsy was homogenized with a pellet pestle motor (749540-0000, Kimble) and resuspended with 450 µl RLT buffer (40724, Qiagen) in a 1.5-ml microcentrifuge tube (1615-5500, USA Scientific). The same experimental procedure was carried out on both punches from the cerebellum, heart, liver and both kidneys. Nuclear preparations were pelleted at 1,000*g* for 5 min and resuspended with 450 µl RLT buffer in a 1.5-ml microcentrifuge tube. Both homogenates and nuclear preparations were then treated with the same protocol: following vortexing for 1 min, samples were incubated at 70 °C for 30 min. Of the Bond-Breaker TCEP solution (77720, Thermo Scientific), 50 µl and 120 mg stainless steel beads (0.2-mm diameter; SSB02, Next Advance) were added, and cellular/nuclear disruption was performed for 5 min on a DisruptorGenie (Scientific Industries), supernatant was transferred to a DNA Mini Column from an AllPrep DNA/RNA Mini Kit (80204, Qiagen) and centrifuged at 8,500*g* for 30 s, washed with Buffer AW1 (80204, Qiagen), centrifuged at 8,500*g* for 30 s and washed again with Buffer AW2 (80204, Qiagen), and then centrifuged at full speed for 2 min. DNA was eluted two times with 50 µl of pre-heated (70 °C) EB (80204, Qiagen) through centrifugation at 8,500*g* for 1 min as previously documented[17].

Whole-genome library preparation and deep sequencing

A total of 1.0 µg of extracted DNA was used for PCR-free library construction using the KAPA HyperPrep PCR-Free Library Prep kit (KK8505, Roche). Mechanical shearing using the Covaris microtube system (SKU 520053, Covaris) was performed to generate fragments with a peak size of approximately 400 bp. Each fragmented DNA sample went through multiple enzymatic reactions to generate a library in which an Illumina dual index adapter would be ligated to the DNA fragments. Bead-based double size selection was performed to ensure that the fragment size of each sample was between 300 and 600 bp as measured by an Agilent DNA High Sensitivity NGS Fragment Analysis kit (DNF-474-0500, Agilent). The concentration of ligated fragments in each library was quantified with the KAPA Library Quantification Kits for Illumina platforms (KK4824, Roche/KAPA Biosystems) on a Roche LightCycler 480 Instrument (Roche). Libraries with concentrations of more than 3 nM and fragments with a peak size of 400 bp were sequenced on an Illumina NovaSeq 6000 S4 and/or S2 flow cell. Each library was sequenced in six to eight independent pools. For each sequencing run, 24 whole-genome sequencing (WGS) libraries were normalized to obtain a final concentration of 2 nM using 10 mM Tris-HCl (pH 8 or 8.5; 50-190-8153, Fisher Scientific). Of the PhiX library, 0.5–1% was spiked into the library pool as a positive control. The normalized libraries in a pool with a total of 311 µl libraries were incubated with 77 µl of 0.2 N sodium hydroxyl (NaOH) (82023-092, VWR) at room temperature for 8 min to denature double-stranded DNA. Of 400 mM Tris-HCl, 78 µl was used to terminate the denaturing process. The denatured library with a final loading concentration of 400 pM in a pool was loaded on the S4 flow cell using Illumina SBS kits (20012866, Illumina) with the following setting on the NovaSeq 6000: PE150:S4 flow cell, dual index, read 1:151, index_read2:8; index_read3:8; read 4:151. The target for WGS with high-quality sequencing raw data was 120 GB or greater with a Q30 > 90% per library per sequencing run. In case the first sequencing run generated less than that, additional sequencing was performed by sequencing the same library on a NovaSeq 6000 S2 flow cell with a 2 × 101 read length for ID01 as documented[17], and all data were generated at 2 × 151 read length for ID05. FASTQ files generated with Picard's (v2.20.7) SamToFastq command from the DRAGEN platform were used as input for the bioinformatic pipeline for ID01, and bcl2fastq2 (v2.20) generated FASTQ files from raw sequence files were used for ID05.

WGS data processing

FASTQ files were then aligned to the human_g1k_v37_decoy genome by BWA's (v0.7.17) mem with -K 100000000 -Y parameters. SAM files were compressed to BAM files via view command in SAMtools (v1.7). BAM files were subsequently sorted by SAMBAMBA's (v0.7.0) sort command and duplicated reads marked by its markdup command. Reads aligned to the indel regions were realigned with GATK's (v3.8-1) RealignerTargetCreator and IndelRealigner following the best practice guideline. Base quality scores were recalibrated using GATK's (v3.8.1) BaseRecalibrator and PrintReads. Germline heterozygous variants were called by GATK's (v3.8.1) HaplotypeCaller. The distribution of library DNA insertion sizes for each sample was summarized by Picard's (v2.20.7) CollectInsertSizeMetrics. The depth of coverage of each sample was calculated by BEDTools's (v2.27.1) coverage command. The code and Snakemake wrapper of the pipeline are freely accessible on GitHub (https://github.com/shishenyxx/Human_Inhibitory_Neurons)[52].

Mosaic SNV/indel detection in WGS data

Mosaic single-nucleotide variants (SNVs)/mosaic small (typically below 20 bp) indels were called by using a combination of four different computational methods: (1) MosaicHunter (single-mode, v1.0)[53] with a posterior mosaic probability > 0.05 (refs. 17,54); (2) single-mode of GATK's (v4.0.4) Mutect2 (ref. 55) with 'PASS' followed by DeepMosaic (v1.0.1)[56]; (3) single-mode of GATK's (v4.0.4) Mutect2 with PASS followed by MosaicForecast (v8-13-2019)[57] were implemented for sample-specific or tissue-shared variants; and (4) the intersection of variants from the paired-mode of Mutect2 and Strelka2 (v2.9.2) (set on PASS for all variant filter criteria)[58] were collected for sample-specific variants. For the panel of normal samples required for the pipeline of DeepMosaic and MosaicForecast, we used an in-house panel of similarly (300×) sequenced normal tissues (n = 15 sperm and 11 blood samples from 11 individuals)[17]. For 'tumour'–'normal' comparisons, required by the Mutect2 and Strelka2 pipelines, we used left–right combined heart tissues as 'normal'. Variants were excluded if: (1) residing in segmental duplication regions as annotated in the UCSC genome browser (UCSC SegDup) or RepeatMasker regions, (2) residing within a homopolymer or dinucleotide repeat with more than 3 units, or (3) overlapped with annotated germline indels. We further removed any variants with a population allele frequency higher than 0.001 in gnomAD (v2.1.1)[59]. Finally, variants with a lower confidence interval of allele fraction < 0.001 were considered noises from reference homozygous and removed. Fractions of mutant alleles (that is, allele fraction) for variants called in one sample were calculated in all the other samples together with the exact binomial confidence intervals using scripts described below for MPAS analysis. This bioinformatic pipeline yielded a total of 898 candidate mosaic variants for ID01 and 2,195 candidate mosaic variants for ID05 (for skin samples, because of the clonal nature, only 10% of the total calls were randomly selected) that were interrogated with MPAS. Scripts for variant filtering are provided on GitHub (https://github.com/shishenyxx/Human_Inhibitory_Neurons)[52].

Formaldehyde-fixed nuclear preparation for sorting

For DARPP32/NEUN dual staining with basal ganglia nuclei, frozen caudate and putamen of ID01 were homogenized in 1% formaldehyde in Dulbecco's phosphate-buffered saline (DPBS; Corning) using a motorized homogenizer (Fisherbrand PowerGen 125), then incubated on a rocker at room temperature for 10 min, quenched with 0.125 M glycine at room temperature on a rocker for 5 min, then centrifuged at 1,100g in a swinging bucket centrifuge. The following steps were all performed on ice except where indicated. Homogenates were washed twice with NF1 buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 5 mM $MgCl_2$, 0.1 M sucrose and 0.5% Triton X-100 in UltraPure water) and centrifuged at 1,100g for 5 min at 4 °C in a swinging bucket centrifuge. Next, pellets were resuspended in 5 ml NF1 buffer and Dounce homogenized 5× in a 7 ml Wheaton Dounce Tissue Grinder (DWK Life Sciences) using a 'loose' pestle. After 30 min of incubation on ice, homogenates were Dounce homogenized 20× with a 'tight' pestle and filtered through a 70-µm strainer. To remove myelin debris, homogenates were overlaid on a sucrose cushion (1.2 M sucrose, 1 M Tris-HCl pH 8.0, 1 mM $MgCl_2$ and 0.1 M dithiothreitol) and centrifuged at 3,200g for 30 min with acceleration and brakes on 'low'. Pellets of nuclei were washed with NF1 buffer and centrifuged at 1,600g for 5 min and stored at −80 °C, same as previously documented[17,60].

Nuclear preparation for MFNS or unfixed nuclei sorting

For the MFNS protocol, approximately 200 mg of freshly frozen tissue stored at −80 °C was prepared, and subsequent procedures were conducted using solutions maintained at 4 °C. The prepared tissue was homogenized in 300 µl of lysis buffer (composed of 10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40 and 1 mM dithiothreitol in nuclease-free water) while kept on ice. Next, an additional 9.7 ml of lysis buffer was added to the homogenate, and the mixture was incubated on ice for 5 min. The homogenate was then passed through a 70-µm cell strainer (352350, Falcon) and centrifuged at 1,100g for 5 min at 4 °C. The supernatant was discarded, and the remaining pellet was gently resuspended and washed with 10 ml of sorting buffer (containing 1% BSA, 1 mM EDTA and 10 mM HEPES in 1× HBSS solution). For the density gradient centrifugation step, the resuspended pellet in 25% Iodixanol solution (OptiPrep, D1556, Millipore) was layered onto a 29% Iodixanol cushion. The centrifugation was carried out at 10,000g with a swinging bucket rotor, using low acceleration and braking, at 4 °C for 40 min, the pellet resuspended in 80% methanol prechilled and stored at −20 °C for at least 30 min before further use. For samples to perform single-nucleus transcriptome and primary template-directed amplification with ResolveOME, brain tissue from participant ID05 was homogenized in 1 ml of ice-cold NIB (composed of 0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 10 mM Tris (pH 7.5), 100 mM dithiothreitol and 0.1% Triton X-100) and subjected to homogenization, incubated on a rocker for 5 min at 4 °C and then centrifuged at 1,000g, using low acceleration and braking in a swinging-bucket centrifuge, and the pellet was reconstituted in 0.5 ml of sorting buffer and filtered through a 70-um strainer. Nuclei in the flow-through were immediately subjected to staining.

FANS

Pellets of defrosted and homogenized brain nuclei were washed twice in sorting buffer and then resuspended in 0.2 ml sorting buffer and incubated overnight at 4 °C. The following antibodies were used: NEUN Alexa Fluor 488 (1:2,500; MAB377, Millipore Sigma), TBR1 unconjugated (1:1,000; ab31940, Abcam), DLX1 (1:200; HPA045884, Atlas Antibodies), COUPTFII (1:400; PP-H7147-00, Novus Biologicals) and DARPP32 (1:400; ab40801, Abcam). The following day, nuclei were washed with staining buffer and, in case an unconjugated antibody was used, nuclei were stained subsequently for 30 min with goat anti-rabbit Alexa 647 (1:4,000; A21244, Thermo Fisher Scientific) for TBR1, DLX1 and DARPP32, or donkey anti-mouse Alexa 647 (1:4,000; A32787, Thermo Fisher Scientific) for COUPTFII. Stained nuclei were washed one more time with staining buffer and passed through a 70-µm strainer. Immediately before the sort, nuclei were stained with 0.5 µg ml^{-1} DAPI. For single-nucleus transcriptome and whole-genome amplification with ResolveOME, nuclei were stained with propidium iodide (1:20; BMS500PI, Invitrogen). Nuclei for the cell type of origin were sorted either on a MoFlo Astrio EQ sorter (Beckman Coulter), BD FACSAria II (Becton-Dickinson) or BD InFlux Cytometer (Becton-Dickinson) similar to previous documentation[17]. At least 1,000 methanol-fixed or more than 50,000 formaldehyde-fixed sorted nuclei were pooled in each tube. Sorted nuclei were pelleted in staining buffer at 1,600g for 10 min. Nuclei for DNA extraction and bisulfite sequencing were stored at −80 °C. FANS data were visualized using FlowJo v10 software. Following MPAS (see below) sorted populations were deemed to be of sufficient overall quality if at least 95% variants were sequenced above more than 1,000×.

Low-input DNA extraction from sorted nuclei

For low-input DNA extraction from sorted nuclei, we further developed an on-bead DNA extraction method: sorted nuclei were centrifuged down for 1 min at 1,000g (4 °C) in a 200-µl PCR tube and resuspended in 20 µl lysis buffer that consists of 30 mM Tris-HCl (pH 8.0), 0.5% (v/v) Tween-20 (P1379, Sigma-Aldrich), 0.5% (v/v) IGEPAL CA-630 (I8896, Sigma), 1.25 µg ml^{-1} protease (19155, Qiagen) as final concentration in nuclease-free water (AM9937, Ambion). The mixture was lightly vortexed for 10 s and centrifuged at 1,500g for 1 min (4 °C). The tube was then subjected to 50 °C for 12 min and 75 °C for 30 min on a thermocycler. To each lysate, 20 µl of Agencourt AMPure XP beads (A63881, Beckman Coulter) was added. The final AMPure beads to sample ratio was 1:1. After pipetting to achieve mixing, the mixture was left at room

temperature for 5 min, placed on a magnet for 5 min and the supernatant removed. The remaining material was washed twice with 150 µl of 80% (v/v) ethanol for 1 min each. The DNA was then suspended in low-TE solution and kept at −20 °C until bisulfite sequencing or MPAS.

Bisulfite sequencing of sorted nuclei for cell type of origin

The low-TE resuspension including DNA and beads were processed by the Pico Methyl-Seq Library Prep Kit (D5455, Zymo Research) to generate bisulfite sequencing libraries. Samples were sequenced at PE150 on the Illumina NovaSeq 6000 platform.

Bisulfite sequencing data processing and data visualization

FASTQ files were analysed with the Bismark bisulfite mapper and methylation marker (v0.23.1)[61], and read pairs were treated as singletons according to the suggestions of the developers. The pipeline was run with Snakemake (v6.12.3), and bedgraph was generated with BEDTools (v2.30.0). Python (v3.10) with argparse, textwrap and numpy packages and R (4.1.3) with lsa, pheatmap and ggfortify packages were used for visualization. Human single-cell methylation data from published literature[31] were downloaded, and reads from excitatory neurons and inhibitory neurons were pooled separately and used as positive controls according to the cell-type labels of the original authors. Codes for plotting the cosine similarity and hierarchical clustering of methylation patterns of gold-standard inhibitory neurons, sorted DLX1$^+$ cells, excitatory neurons and sorted TBR1$^+$ cells, and control samples from the heart are available on GitHub (https://github.com/shishenyxx/Human_Inhibitory_Neurons)[52].

Single-nucleus transcriptome and primary template-directed amplification

After sorting, a total of 192 single nuclei from each of 2 samples (right temporal and frontal cortex of ID05) were snap-frozen on dry ice. Nuclei underwent the ResolveOME workflow (BioSkryb Genomics Inc.). In brief, Biotin-dT-primed first strand cDNA was generated. After termination of the reaction and nuclear lysis, whole-genome amplification with primary template-directed amplification was performed[62]. The mRNA-derived cDNA was affinity purified with streptavidin beads from the combined pool of cDNA and amplified genome. Remaining cDNA were pre-amplified on beads. Independently, amplified cDNA and single-cell genomic DNA from each cell underwent SPRI (B23319, Beckman Coulter) cleanup before library preparation. Illumina libraries were prepared using the ResolveOME library preparation kit (BioSkryb Genomics Inc.) with NEXTFLEX Unique Dual Index Barcodes (NOVA-534100, PerkinElmer Applied Genomics). Libraries were sequenced at low pass (2× 50-bp paired end) targeting 2 million reads on a NextSeq (Illumina) instrument. Libraries of interest were identified based on quality control sequencing, and were subsequently sequenced at paired-end 150 bp (DNA libraries) and paired-end 100 bp (RNA-derived libraries) on a NovaSeq X Plus (Illumina) platform.

MPAS and snMPAS

Two customized AmpliSeq Custom DNA Panel for Illumina (20020495, Illumina) were used for MPAS for ID01 (#203019) and MPAS for ID05 (#201745), respectively. Designed genomic regions are provided in Supplementary Data 9. A list of 259 mosaic variants used in our previous study and 639 candidate mosaic variants from the mosaic variant detection pipeline in ID01 described above was subjected to the AmpliSeq design system. For the first panel, we randomly selected 120 high-confidence heterozygous variants as positive controls. These heterozygous variants presented with estimated allele fractions between 48% and 52% for all the 25 sequenced bulk tissues, and with read depths between 270 and 330×. Of the 120 variants, 45 were private variants and 75 were present in gnomAD at different population allele frequencies. We also randomly selected 30 reference homozygous variants as negative controls. These reference homozygous variants presented with approximately 0% allele fraction across all sequenced samples, with an average depth of 270–330×, and gnomAD (v2.1.1) allele frequency of more than 0.5 to exclude any potential contamination or amplification bias. For the second panel, 2,195 candidate mosaic variants detected from ID05, and 152 randomly chosen variants detected as heterozygous and 59 as alternative homozygous in ID05 were subjected to the AmpliSeq design system. DNA from extracted tissue, amplified single nuclei and a duplicate unrelated control sample was diluted to 5 ng µl^{-1} in low-TE provided in AmpliSeq Library PLUS (384 reactions) kit (20019103, Illumina). For sorted nuclei, the low-TE resuspension including DNA and beads stored after the 'Low-input DNA extraction from sorted nuclei' step was prepared. AmpliSeq was carried out following the manufacturer's protocol (document 1000000036408v07). For amplification in bulk samples, 14 cycles each for 8 min were used; for amplification in low-input sorted nuclei on beads, additional cycles and time were added accordingly based on the table from the manufacturer. After amplification and fragmentase universal primer assay (FUPA) treatment, libraries were barcoded with AmpliSeq CD Indexes (20031676, Illumina) and pooled with similar molecular numbers based on measurements made with a Qubit dsDNA High Sensitivity kit (Q32854, Thermo Fisher Scientific) and a plate reader (PlateReader AF2200, Eppendorf). To avoid index hopping, the two MPAS library pools for ID01 and ID05 and the snMPAS library pool for ID05 were sequenced on separate lanes on different NovaSeq 6000 runs. Of FASTQ data, 192 GB was obtained from the ID01 MPAS libraries, 339 GB of FASTQ data were obtained from the ID05 MPAS libraries and 193 GB of FASTQ data were obtained from the ID05 snMPAS libraries, all aiming for an average of 10,000× for each variant.

Data analysis for MPAS and snMPAS

Raw reads from MPAS and snMPAS were mapped to the human_g1k_v37_decoy genome with BWA's (v0.7.17) mem command. BAM files were processed without removing PCR duplicates. Reads near indels were realigned with GATK's (v3.8.1) IndelRealigner and base quality scores were recalibrated with GATK's (v3.8.1) BaseRecalibrator. The final BAM files were parsed by SAMtools's (v1.7) mpileup and the 95% confidence intervals of the measured allele fraction of all the candidate mosaic variants, together with the homozygous (negative control) and heterozygous (positive control) variants, were estimated based on an exact binomial estimation (https://github.com/shishenyxx/Human_Inhibitory_Neurons)[52]. Following depth calculation, regions of 639 mosaic candidates, 120 heterozygous variants (positive controls) and 30 homozygous variants (negative controls) were detected and subjected to the next genotyping steps with 259 previously validated mosaic variants[17] for ID01, and 2,195 mosaic candidates, 152 heterozygous variants (positive controls) and 59 homozygous variants for ID05. The genotypes of candidate mosaic variants from MPAS were determined by comparing them to the allele fraction distribution of the reference homozygous and heterozygous variants. The exact binomial lower bounds of all reference homozygous variants with more than 30 read depth were estimated and the 95% single-tail confidence threshold for the lower bound was calculated to be 0.001677998 (ID01) and 0.002360687 (ID05). The distribution of the exact binomial upper bound of all heterozygous variants was calculated, and 0.4706769 (ID01) and 0.4779163 (ID05) were the threshold for the upper bond based on an approximately 5% false discovery rate based on the built-in heterozygous and alternative homozygous genomic positions. Mosaic candidates from WGS were considered positive if variants met the following criteria at the same DNA samples: (1) the 95% exact binomial lower bound was more than 0.001677998 (ID01) or more than 0.002360687 (ID05), (2) the lower confidence interval of the unrelated control sample was less than 0.001677998 (ID01) or less than 0.002360687 (ID05), (3) the 95% exact binomial upper bound was less than 0.4706769 (ID01) or less than 0.4779163 (ID05),

(4) the sequencing depth was more than 30, (5) the assessed alternative allele was supported by 3 reads or more, and (6) for ID01, variants were detected in the heart, globus pallidus, caudate, putamen and thalamus in WGS. These criteria ensured that the false discovery rate for each variant was under 5%. After the MPAS quantification, 328 mosaic variants (28 newly positively validated mosaic variants from brains, 41 from the heart and the original 259 variants[17]) from ID01 and 780 mosaic variants from ID05 were considered positively validated in the sample where the variant was originally detected. The validation rate was 36.5% (328 of 898) and 35.5% (780 of 2,195) in ID01 and ID05, respectively. Of 328 mosaic variants excluding 41 heart mosaic variants from ID01 and all 780 mosaic variants from ID05, 287 mosaic variants were used for all the analysis presented throughout the paper. In snMPAS, mosaic candidates from WGS were considered positive if the lower confidence interval of allele fraction was larger than the upper confidence interval of allele fraction in the normal control sample. To assess precision of snMPAS data, we first calculated the chance of the failure of detection of true-positive variant calls (dropout), using heterozygous variants. Among allele fraction values of the variants known as heterozygous in ID05 bulk and homozygous in control bulk samples, 94.137% (5,379 of 5,714) of those were recovered as heterozygous in snMPAS, resulting in an estimated false-negative rate of 5.863%. Next, we calculated the chance of false-positive discovery using homozygous genomic sites. The known homozygous genomic positions in bulk samples recovered as homozygous were 98.948% (7,149 of 7,225), resulting in an estimated false-positive rate of 1.052%.

Mutational signature analysis

Mutational signature analysis was performed using a web-based somatic mutation analysis toolkit (Mutalisk[63]). The PCAWG SigProfiler full screening model was used.

Computational deconvolution for DLX1$^+$ populations

Given the consistency of the FANS sorting strategy and the consistent contribution in each cell type carrying the same mutation, we assumed that the contamination rate α for TBR1$^+$ in DLX1$^+$ nuclei is the same across different samples. For variant number n, the vector of observed variant allele fraction (AF) in DLX1$^+$ nuclear pools is:

$$\begin{aligned} AF_{DLX1-observed} &= \begin{Bmatrix} AF_{DLX1-observed_1} \\ AF_{DLX1-observed_2} \\ \cdots \\ AF_{DLX1-observed_n} \end{Bmatrix} \\ &= (1-\alpha) \times AF_{DLX1-theoretical} + \alpha \times AF_{TBR1-observed} \\ &= (1-\alpha) \times \begin{Bmatrix} AF_{DLX1-theoretical_1} \\ AF_{DLX1-theoretical_2} \\ \cdots \\ AF_{DLX1-theoretical_n} \end{Bmatrix} + \alpha \times \begin{Bmatrix} AF_{TBR1-observed_1} \\ AF_{TBR1-observed_2} \\ \cdots \\ AF_{TBR1-observed_n} \end{Bmatrix} \end{aligned}$$

Thus, the theoretical allele fraction rates for each DLX1$^+$ nuclear pool in the same lobe is calculated as:

$$AF_{DLX1-theoretical} = \begin{Bmatrix} \frac{AF_{DLX1-observed_1} - \alpha \times AF_{TBR1-observed_1}}{1-\alpha} \\ \frac{AF_{DLX1-observed_2} - \alpha \times AF_{TBR1-observed_2}}{1-\alpha} \\ \cdots \\ \frac{AF_{DLX1-observed_n} - \alpha \times AF_{TBR1-observed_n}}{1-\alpha} \end{Bmatrix}$$

The contamination rate α was set as 0.25 for Extended Data Fig. 7c,d, based on the fraction of excitatory neuronal fractions in DLX1$^+$ nuclei described in Fig. 1d.

Simulated TBR1$^+$ nuclei contamination for COUPTFII$^+$ populations

Given that n is the number of variants and α is the fraction of excitatory neurons in sorted DLX1$^+$ nuclear pools, the allele fractions of artificial contaminated COUPTII$^+$ nuclear pools contaminated by TBR1$^+$ nuclear pools ($AF_{COUPTF2-mixture}$) were generated by mixing α fraction of TBR1$^+$ nuclear pool allele fractions ($AF_{COUPTF2-observed}$) and $1-\alpha$ fraction of COUPTFII$^+$ nuclear pool allele fractions ($AF_{TBR1-observed}$):

$$AF_{COUPTF2-mixture} = \begin{Bmatrix} (1-\alpha) \times AF_{COUPTF2-observed_1} + \alpha \times AF_{TBR1-observed_1} \\ (1-\alpha) \times AF_{COUPTF2-observed_2} + \alpha \times AF_{TBR1-observed_2} \\ \cdots \\ (1-\alpha) \times AF_{COUPTF2-observed_n} + \alpha \times AF_{TBR1-observed_n} \end{Bmatrix}$$

The fraction α was set as 0.25 for Extended Data Fig. 7e,f based on the portion of excitatory neurons in DLX1$^+$ nuclear pools described in Fig. 1d.

Estimating the contribution of dorsal and ventral origin for DLX1$^+$ inhibitory neurons. After computational decontamination for DLX1$^+$ nuclear pools, we obtained $AF_{DLX1-theoretical}$, to further estimate the contribution from dorsally (AF_{Dorsal}) and ventrally ($AF_{Ventral}$) derived clones assuming there is no third origin, we introduced β as the proportion of dorsally derived clones to have:

$$\begin{aligned} AF_{DLX1-theortical} &= \begin{Bmatrix} AF_{DLX1-theoretical_1} \\ AF_{DLX1-theoretical_2} \\ \cdots \\ AF_{DLX1-theoretical_n} \end{Bmatrix} \\ &= \beta \times AF_{Dorsal} + (1-\beta) \times AF_{Ventral} \\ &= \beta \times \begin{Bmatrix} AF_{Dorsal_1} \\ AF_{Dorsal_2} \\ \cdots \\ AF_{Dorsal_n} \end{Bmatrix} + (1-\beta) \times \begin{Bmatrix} AF_{Ventral_1} \\ AF_{Ventral_2} \\ \cdots \\ AF_{Ventral_n} \end{Bmatrix} \end{aligned}$$

The loss function $L(\beta)$ for the estimation is defined as the residual sum of squares:

$$\begin{aligned} L(\beta) &= \sum_{i=1}^{n} (AF_{estimated_i} - AF_{DLX1-theoretical_i})^2 \\ &= \sum_{i=1}^{n} (\beta \times AF_{Dorsal_i} + (1-\beta) \times AF_{Ventral_i} - AF_{DLX1-theoretical_i})^2 \end{aligned}$$

For each lobe, the optimal contribution of β (ranging from 0 to 1, step 0.01) is obtained by minimizing $L(\beta)$. AF_{Dorsal} is represented by $AF_{TBR1-observed}$ given the known dorsal origin, and $AF_{Ventral}$ is represented by $AF_{COUPTF2-observed}$ given the high potential of ventral origin demonstrated in Fig. 3.

Permutation test for the significance of the snMPAS result

Mosaic variants that presented more than 2 nuclei and a maximum of 14 nuclei from Fig. 4b, detection events of each mosaic variant were randomly reassigned within a total of 118 nuclei (16 and 17 InNs in F and T respectively; 40 and 45 ExNs in F and T, respectively), maintaining the original detection frequency. The number of inhibitory neurons sharing mosaic variants exclusively in one lobe and shared with at least two other local cells, including one excitatory neuron, were used as the outcome and a null distribution was generated from the 10,000 permutations. The probability of having more than or equal to 15 inhibitory neurons sharing mosaic variants exclusively in one lobe and shared with at least two other local cells, including one excitatory neuron was calculated and used as the one-tailed permutation P value.

snRNA-seq with Chromium platform

Nuclei that underwent MFNS (16,000 nuclei) were resuspended in a sorting buffer to make the desired concentration (800–1,000 nuclei per µl) targeting 10,000 nuclei per reaction. Gel beads emulsion generation, cDNA and sequencing library constructions were performed in accordance with instructions in the Chromium Single Cell 3′ Reagent Kits user guide (v3.1). Each library pool was sequenced with 200 million read pairs using NovaSeq 6000.

snRNA-seq bioinformatics pipeline

For the snRNA-seq data made with Chromium platform, FASTQ files from single-nucleus libraries were processed through the Cell Ranger (v6.0.2) analysis pipeline with -include-introns option and hg19 reference genome. Seurat (v4.0.5) package was used to handle single-nucleus data objects. Nuclei passed a control filter (nCount > 400, nFeature_RNA < 2,000, percentage of mitochondrial gene < 10%) was used for downstream analysis. A total of 15,896 protein-coding genes were used for further downstream analysis. To balance the nucleus number for each group, a total of 500 random nuclei for each $DAPI^+$, $DLX1^+$, $TBR1^+$ and $COUPTFII^+$ group were selected. Data were normalized and scaled with the most variable 1,000 features using the 'ScaleData' functions. Dimensionality reduction by principal component analysis and UMAP embedding was performed using runPCA and runUMAP functions. Clustering was performed by FindNeighbors and FindClusters functions. For the full-length transcript snRNA-seq through ResolveOME platform, quality control were carried out for raw FASTQ from snRNA-seq from ResolveOME files using fastqc (v0.11.8). Preprocessing was carried out with cutadapt (v1.16). Cleaned FASTQ files were aligned to the GRCh38 human genome and gencode (v27) gtf annotation using STAR (v2.6.0c). Aligned BAM files were indexed with SAMtools (v1.7). PCR duplicates were marked with Picard (v2.20.7) MarkDuplicates. Post-alignment quality control was carried out with Picard's (v2.20.7) CollectRnaSeqMetrics, CollectInsertSizeMetrics, CollectGcBiasMetrics and qualimap (v2.2.2-dev). Raw read counts were collected with featureCounts (v2.0.0). Transcripts were collected with rsem (v1.3.1) with seed 12345 using the same gtf file. We also used trimmed-means data from the Human Multiple Cortical Areas SMART-seq dataset (https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-smart-seq) for the reference group. The Seurat (v4.0.5) package was used for further analysis. Nuclei that passed a control filter (nCount > 1,000, nFeature_RNA > 500 and percentage of mitochondrial gene < 30%) were used for downstream analysis (Supplementary Data 10). The SCTransform function was used for normalizing and scaling data. For both snRNA-seq data made with the Chromium platform and ResolveOME, cell-type identification was performed using known cell-type markers expressed in the brain including excitatory (*RORB*, *CUX2* and *SATB2*), inhibitory neuron (*GAD1* and *GAD2*), astrocyte (*SLC1A2* and *SLC1A3*), oligodendrocyte (*MOBP* and *PLP1*), oligodendrocyte precursor cell (*PDGFRA*), microglia (*PTPRC*) and endothelial cell markers (*CLDN5* and *ID1*), as well as using positive markers found by the FindAllMarkers function with the 1,000 most variable features in scaled data. The final visualization of various snRNA-seq data was performed by r-ggplot2 (v3.3.5).

Phylogenetic tree analysis

From the 68 mosaic variants in 118 cell-type-resolved single nuclei in Fig. 4b, the alleles at each genomic position were combined into one 'pseudo-sequence' for each sample, and a sequenced-based phylogenic tree was reconstructed to deconvolve the clonal relationship between single cells. Multiple sequence alignments were carried out with MUSCLE[64], and a maximum likelihood phylogenetic tree was constructed with Molecular Evolutionary Genetics Analysis (v11.0.13)[65]. Maximum likelihood fit-based model selection was carried out for 24 different nucleotide substitution models, and the Kimura 2-parameter model was elected as the best-performing substitution model. After 1,000 bootstrap replications and using the nearest-neighbour interchange heuristic method, the bootstrap consensus phylogenetic tree was generated, as shown in Extended Data Fig. 9. Cell types were labelled based on transcriptomic information from the same cell.

Statistical tests and packages for customized plots

Hierarchical clustering with *P* value via bootstrap resampling was performed using r-pvclust (v2.2.0) package. Pearson's product-moment correlation is calculated using cor.test() in R. One-way ANOVA with Tukey multiple comparisons of means was performed with aov() and TukeyHSD() function. Various heatmaps with dendrograms and sidebars were generated by the ComplexHeatmap (v2.16.0) package. Various plots including violin plots, scatter plots, contour plots, bar plots, UpSet plots and lolliplots were generated using r-ggplot2 (v3.4.3). The oncoplot was generated using maftools (v2.16.0) in R. UMAP analysis was performed with the r-umap (v0.2.10.0) package. snRNA-seq data were analysed and plotted using the Seurat4 (4.2.0) package in R.

Reporting summary

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

Data availability

Raw WGS and MPAS and snMPAS are available through the Sequence Read Archive (accession number PRJNA799597) and NDA (accession number study 919) for ID01 and ID05. The 300× WGS panel of normal is available on the Sequence Read Archive (accession number PRJNA660493). human_g1k_v37 reference: http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/. gnomAD: https://gnomad.broadinstitute.org/. Human multiple cortical areas SMART-seq dataset: https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-smart-seq.

Code availability

Details and codes for the data processing and annotation are provided on GitHub (https://github.com/shishenyxx/Human_Inhibitory_Neurons)[52].

52. Chung, C. et al. shishenyxx/Human_Inhibitory_Neurons: 1.0.1. Zenodo https://doi.org/10.5281/ZENODO.10772159 (2024).
53. Huang, A. Y. et al. MosaicHunter: accurate detection of postzygotic single-nucleotide mosaicism through next-generation sequencing of unpaired, trio, and paired samples. *Nucleic Acids Res.* **45**, e76 (2017).
54. Breuss, M. W. et al. Autism risk in offspring can be assessed through quantification of male sperm mosaicism. *Nat. Med.* **26**, 143–150 (2020).
55. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
56. Yang, X. et al. Control-independent mosaic single nucleotide variant detection with DeepMosaic. *Nat. Biotechnol.* **41**, 870–877 (2023).
57. Dou, Y. et al. Accurate detection of mosaic variants in sequencing data without matched controls. *Nat. Biotechnol.* **38**, 314–319 (2020).
58. Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
59. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
60. Nott, A. et al. Brain cell type-specific enhancer–promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
61. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
62. Gonzalez-Pena, V. et al. Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.2024176118 (2021).
63. Lee, J. et al. Mutalisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and epigenomic signatures. *Nucleic Acids Res.* **46**, W102–W108 (2018).
64. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
65. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).

Acknowledgements We thank the individuals who donate their bodies and tissues for the advancement of research; T. Komiyama for feedback; and the San Diego Supercomputer Center (grant no. TG-IBN190021 to X.Y. and J.G.G.) for computational help. This work was supported by the National Institute of Mental Health (NIMH) (grants U01MH108898 and R01MH124890 to J.G.G. and R21MH134401 to X.Y. and J.C.M.S.), the Larry L. Hillblom Foundation Grant (to J.G.G.), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) (grant K99HD111686 to X.Y.), a 2021 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (30598 to C.C.) and the Rady Children's Institute for Genomic Medicine. This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a US National Institutes of Health SIG grant (no. S10OD026929 to C.C., X.Y. and J.G.G.). We are grateful to C. Fine, M. Espinoza and M. Banihassan (UCSD) for technical assistance with flow cytometry experiments, supported by the UCSD Stem Cell Program and a CIRM Major Facilities grant (FA1-00607) to the Sanford Consortium for Regenerative Medicine. This publication includes data generated at the UCSD Human Embryonic Stem Cell Core Facility, using the BD Biosciences Influx, FACS Aria Fusion and FACS Aria II Flow Cytometry Sorters. Images in Fig. 1 and Extended Data Fig. 1 were created and modified using BioRender (https://biorender.com).

Author contributions C.C., X.Y. and J.G.G. designed the study. C.C., X.Y., R.F.H., K.I.V., C.B., V.S., S.M., M.W.B., J.C.M.S., S.T.B., G.N. and S.F.K. organized, handled and sequenced human samples. C.C., X.Y. and K.K. performed the ResolveOME experiment. C.C. and Y.L. performed the MFNS experiment. C.C., X.Y., A.P. and R.N. performed the bioinformatics and data analyses. C.C., X.Y. and J.G.G. wrote the manuscript. All authors reviewed the manuscript. C.C. and X.Y. contributed equally to this work.

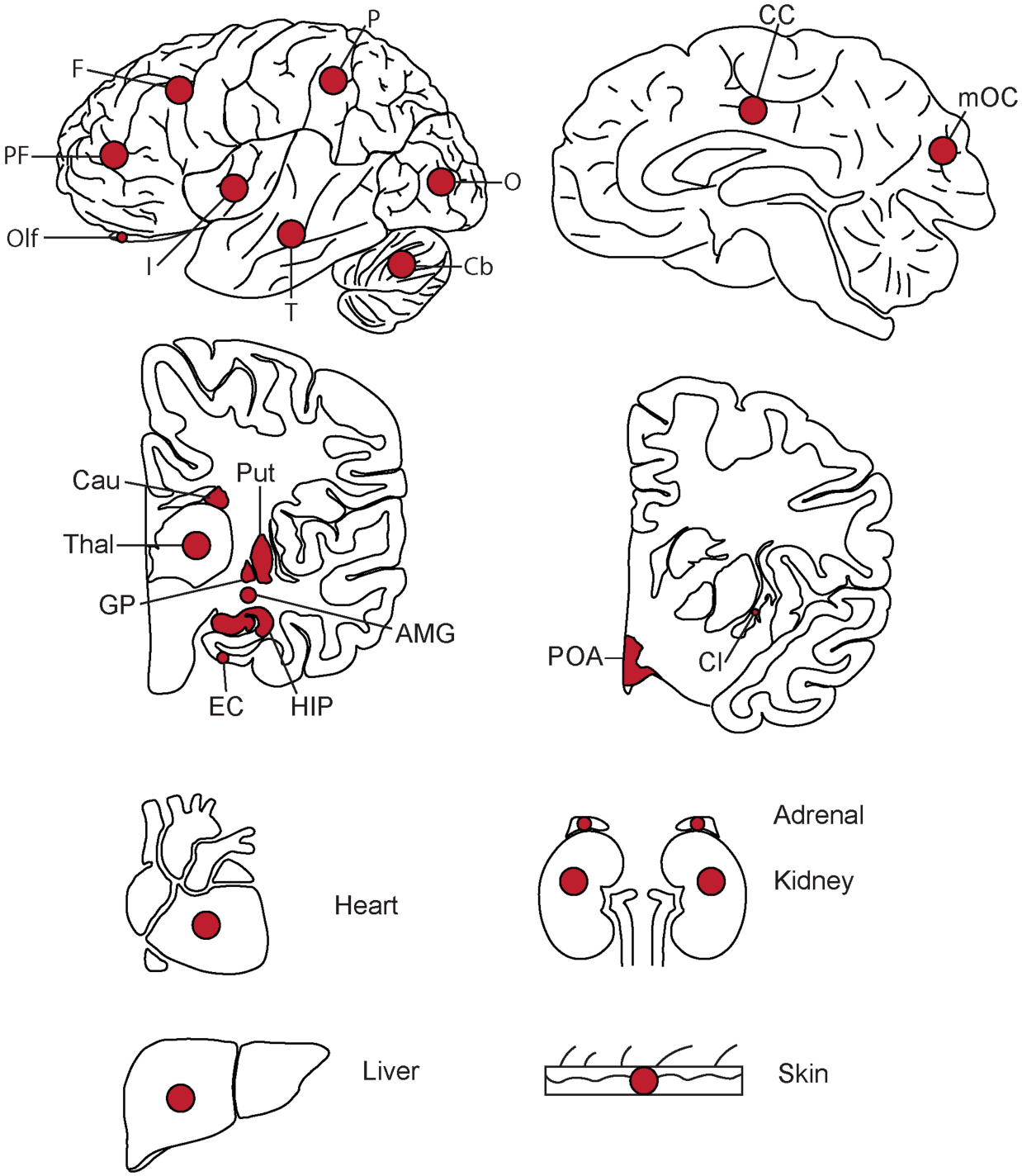
Competing interests K.K. is a senior scientist at Bioskryb Genomics Inc. All other authors declare no competing interests.

Additional information

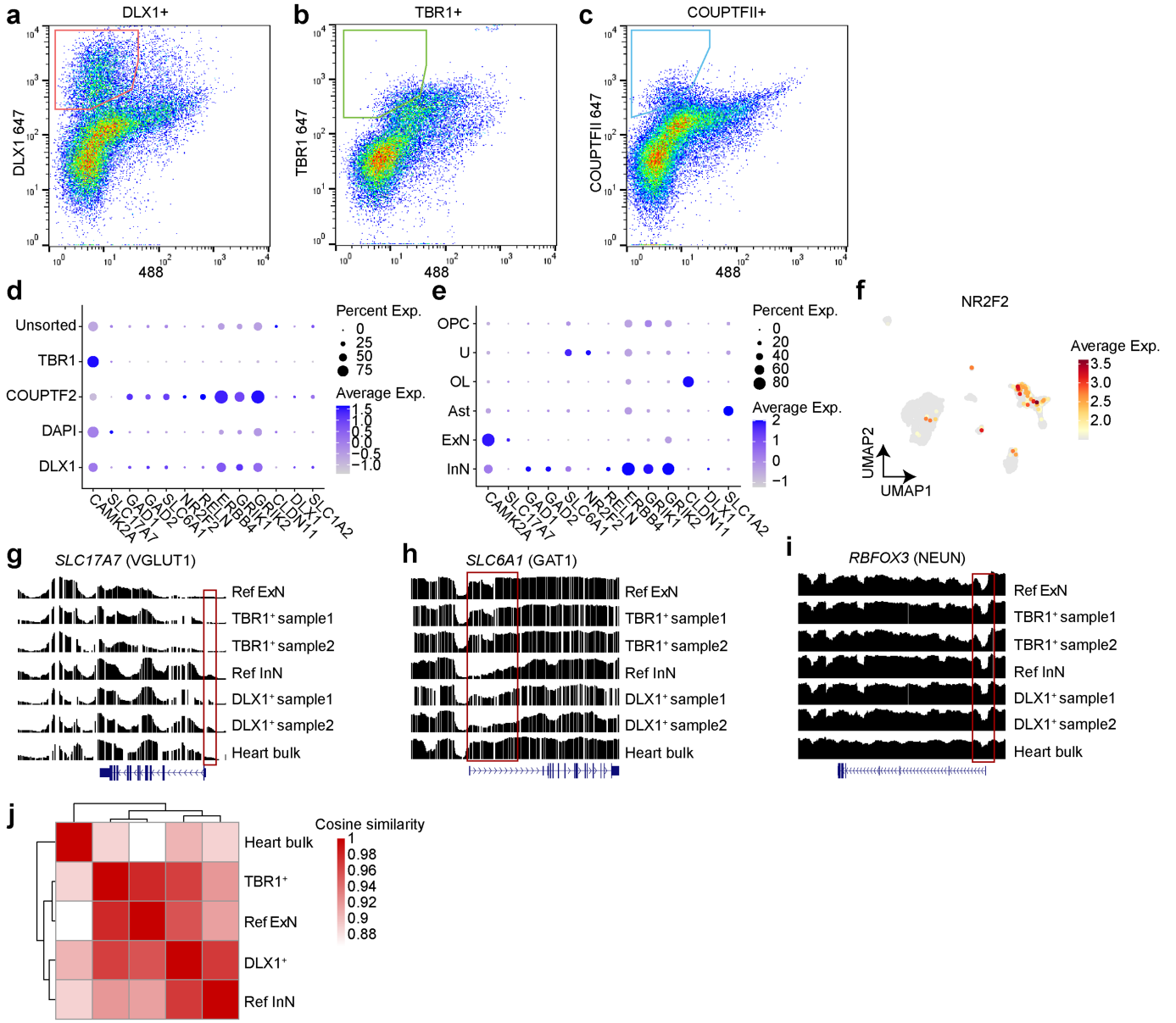
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07292-5.

Correspondence and requests for materials should be addressed to Joseph G. Gleeson.

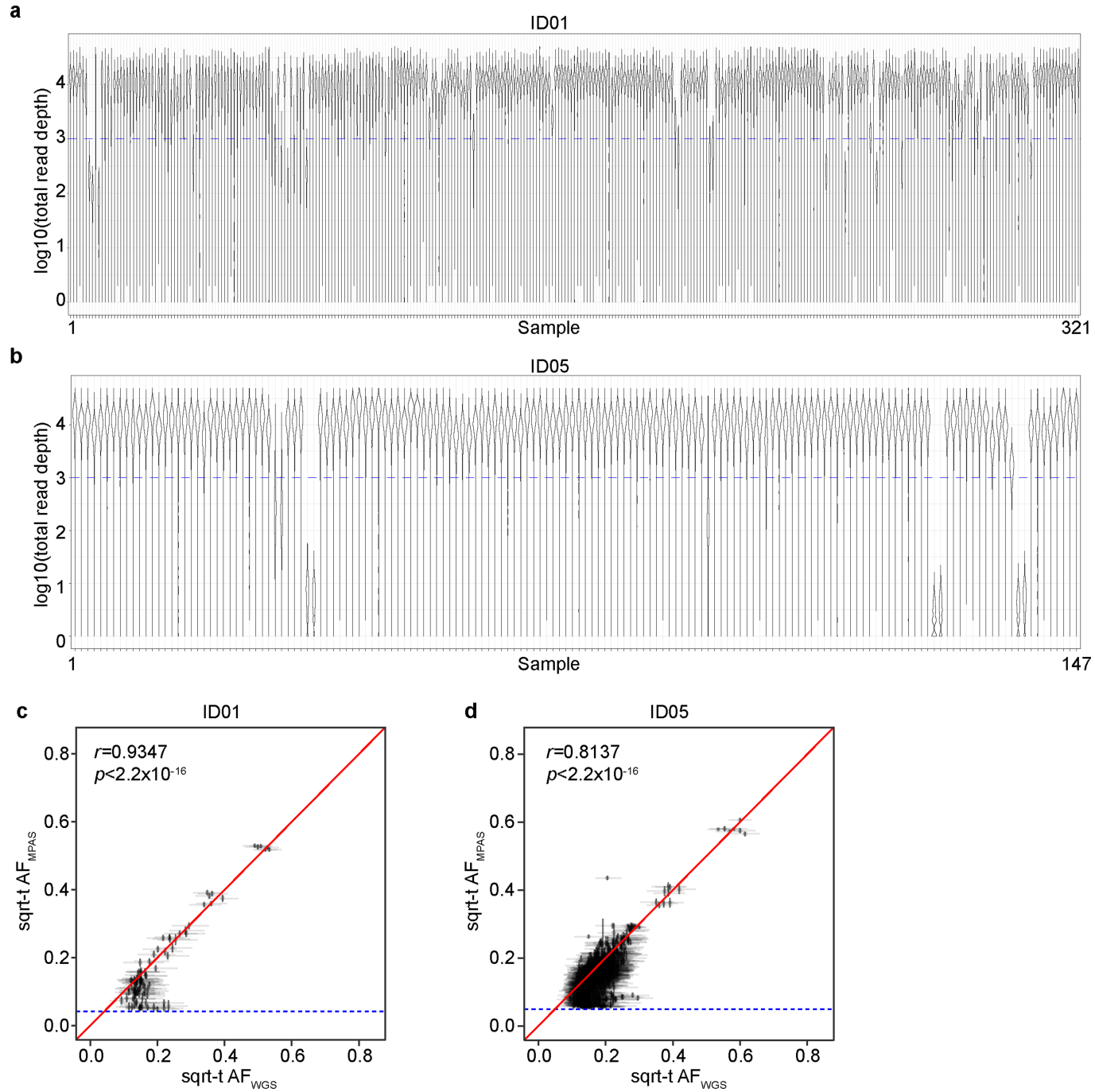
Peer review information *Nature* thanks Young Seok Ju and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



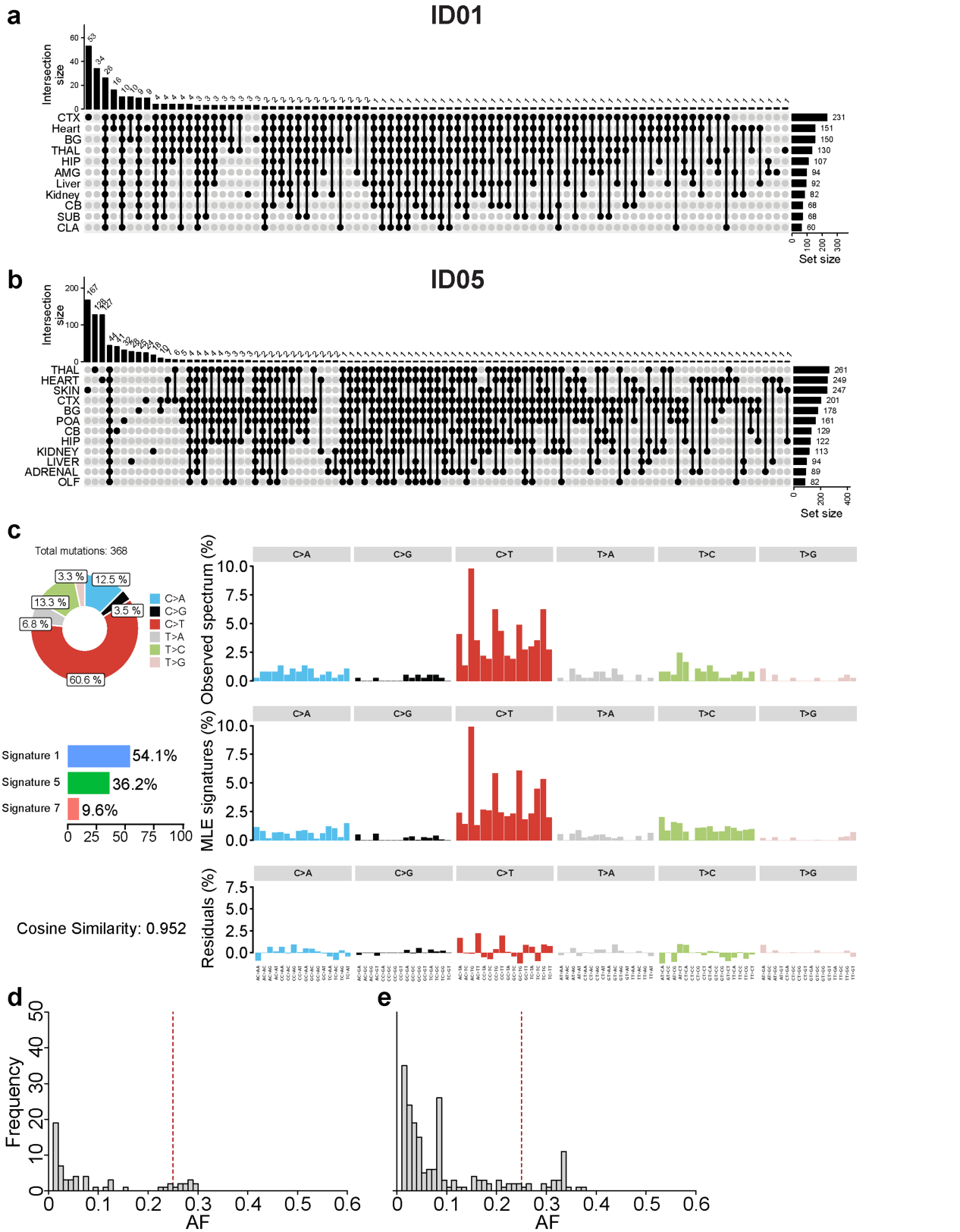
Extended Data Fig. 1 | Tissues collected from ID01 and ID05. Red dots indicate approximate sites of punch biopsies. Abbreviations: PF, prefrontal cortex; F, frontal cortex; P, parietal cortex; O, occipital cortex; T, temporal cortex; I, insular cortex; Cb, Cerebellum; CC, cingulate cortex; mOC, medial occipital cortex; Cau, Caudate; Put, Putamen; Thal, Thalamus; GP, globus pallidus; EC, entorhinal cortex; HIP, hippocampus; AMG, amygdala; POA, preoptic area; Cl, Claustrum. Graphics were created with BioRender (https://biorender.com).



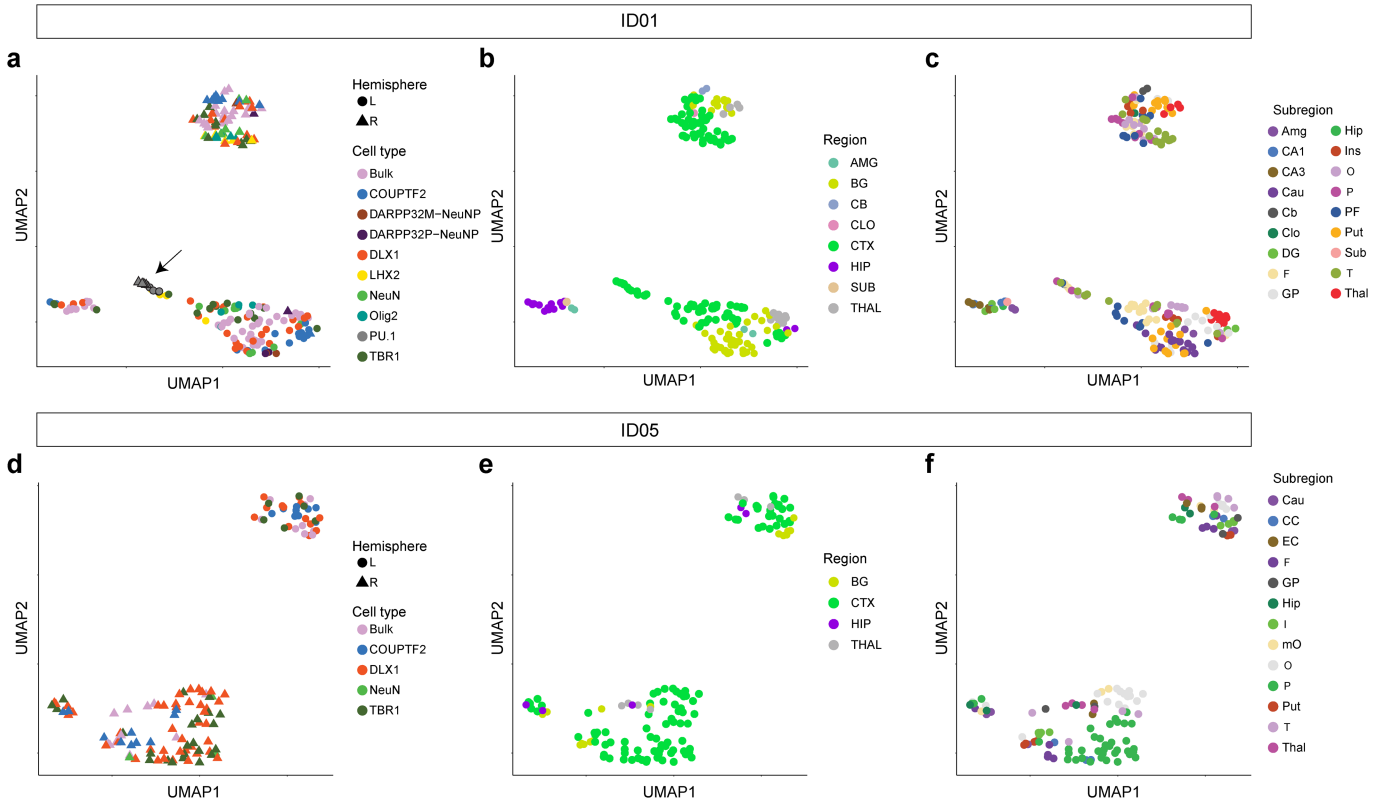
Extended Data Fig. 2 | Bisulfite sequencing in sorted TBR1$^+$ and DLX1$^+$ nuclear pools correlate with excitatory and inhibitory neuron methylome signatures. (a-c) MFNS gating strategy on 30,000 single brain nuclei using DLX1, TBR1, and COUPTFII antibodies. X-axis: 488 channel intensity for monitoring autofluorescence signals. Y axis: Fluorescence intensity from antigen-bound antibodies. (d) snRNA-seq of post-MFNS nuclei confirming enrichment of targeted nuclear types. (e) Marker expression in assigned nuclear types correlating with targeted nuclear types. (f) UMAP plot *NR2F2* expression pattern (encoding COUPTFII) highlighting a subpopulation of inhibitory neurons (compare with Fig. 1c). (g-i) Reference excitatory and inhibitory neuronal methylome signatures (aggregated from an available public single-nuclei methylome dataset) compared to methylome signatures of sorted nuclei and a bulk heart sample. Normalized relative methylation levels (y-axes) and genomic positions (x-axes) of genes listed at top. (g) Methylome signature of *SLC17A7* encoding VGLUT1, an excitatory neuronal marker in the brain, showing reduced methylation (i.e. representing activation) across the gene body and especially near the transcription start site (TSS, red box) in TBR1$^+$ excitatory neuron samples. (h) Methylome signature of *SLC6A1* encoding VGAT1, an inhibitory neuronal marker in the brain, showing reduced methylation across the gene body and especially near the TSS (red box). (i) Methylome signature of *RBFOX3* encoding NEUN, a mature neuronal marker in the brain, showing reduced methylation at the TSS in neurons compared with bulk heart. Ref ExN, reference excitatory neurons; Ref InN, reference inhibitory neurons. (j) Heatmap and dendrograms based on cosine similarities of global methylation patterns between groups. Two different TBR1$^+$ or DLX1$^+$ nuclear pools were aggregated. The TBR1$^+$ nuclear pool was clustered with Ref ExN while the DLX1$^+$ clustered near the pool with Ref InN. The control heart bulk sample was distant from either group.



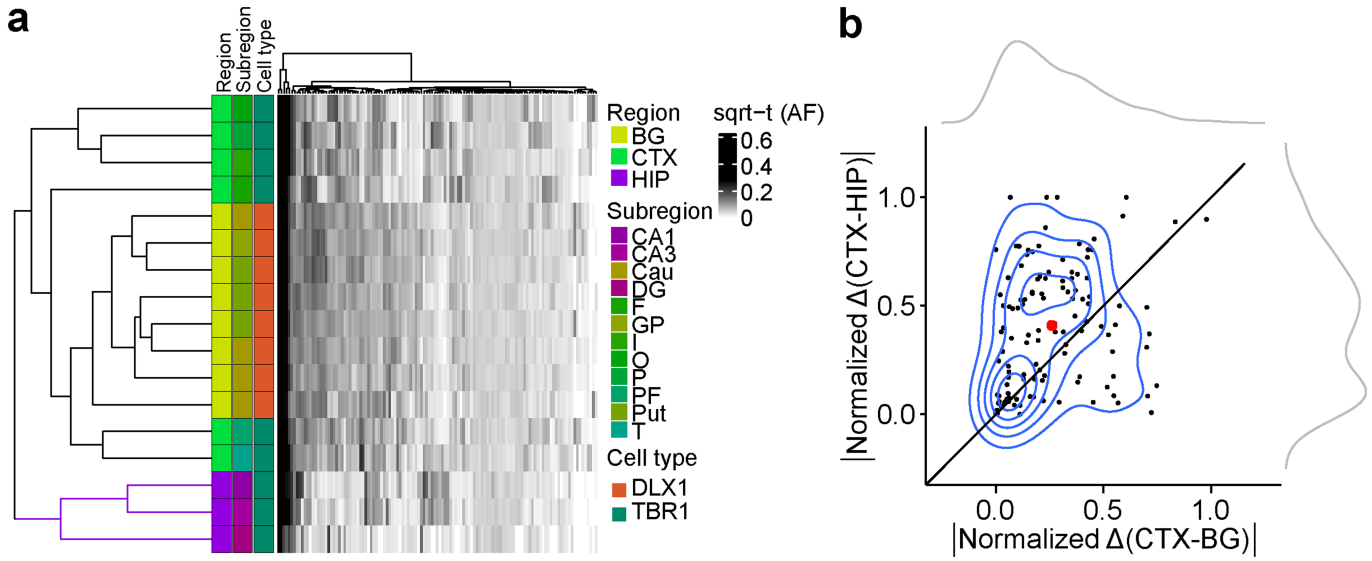
Extended Data Fig. 3 | Quality controls of MPAS results. (a-b) Violin plot distribution of log-transformed total read depths (y-axes) of individual variant positions in 321 or 147 samples from ID01 or ID05 (x-axes), respectively. The blue dashed lines indicate 1000× read depth. (c-d) Correlation between sqrt-t AF of individual variants from WGS and MPAS. Error bars of individual points: square-root-transformed lower and upper bounds for binomial distribution of individual AFs. Blue horizontal dashed lines: Lower bound for binomial distribution detection threshold. *r* and *p*-values (two-tailed) from Pearson's Product-Moment correlation. Identity lines (red).



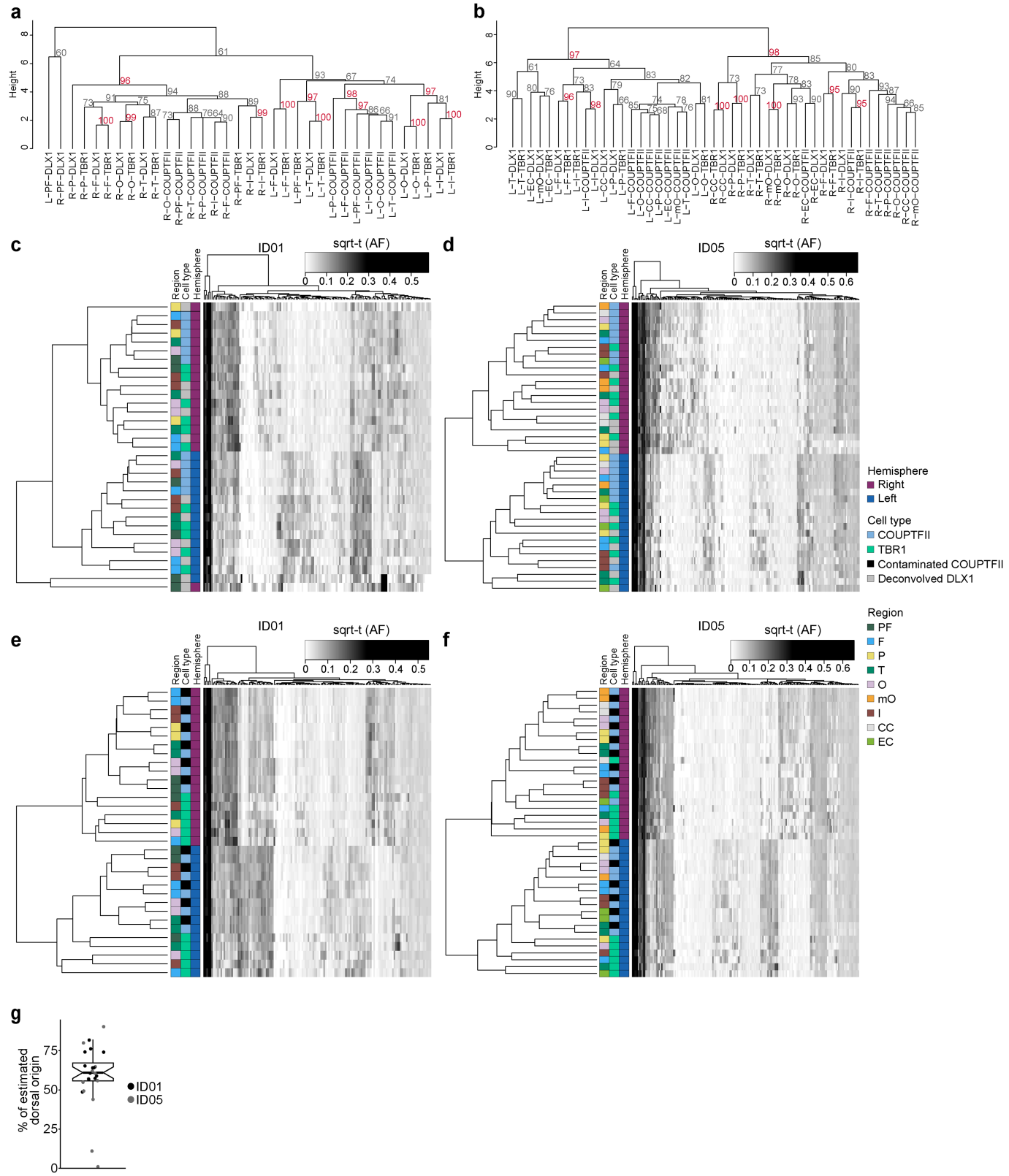
Extended Data Fig. 4 | Basic characteristics of positively validated MVs from the cMVBA pipeline. (a) ID01. (b) ID05. CTX, cortex; BG, Basal ganglia; THAL, thalamus; HIP, hippocampus; AMG, amygdala; CB, cerebellum; SUB, subiculum; CLA, Claustrum; POA, preoptic area; OLF, olfactory bulb. (c) Mutational signature analysis using 368 brain-specific somatic single nucleotide variants (sSNVs) from ID01 and ID05 using Mutalisk. Clonal sSNVs show clock-like signatures such as SBS 1 and 5, reflecting embryonic developmental origins. (d-e) AF distributions of organ-shared early embryonic MVs in ID01(d) and ID05 (e) reflect the asymmetric clonal division in early human embryos. Vertical dashed lines (red): expected peaks (AF = 25%) from the first symmetric cell division, absent in observed distribution, suggesting asymmetric divisions.



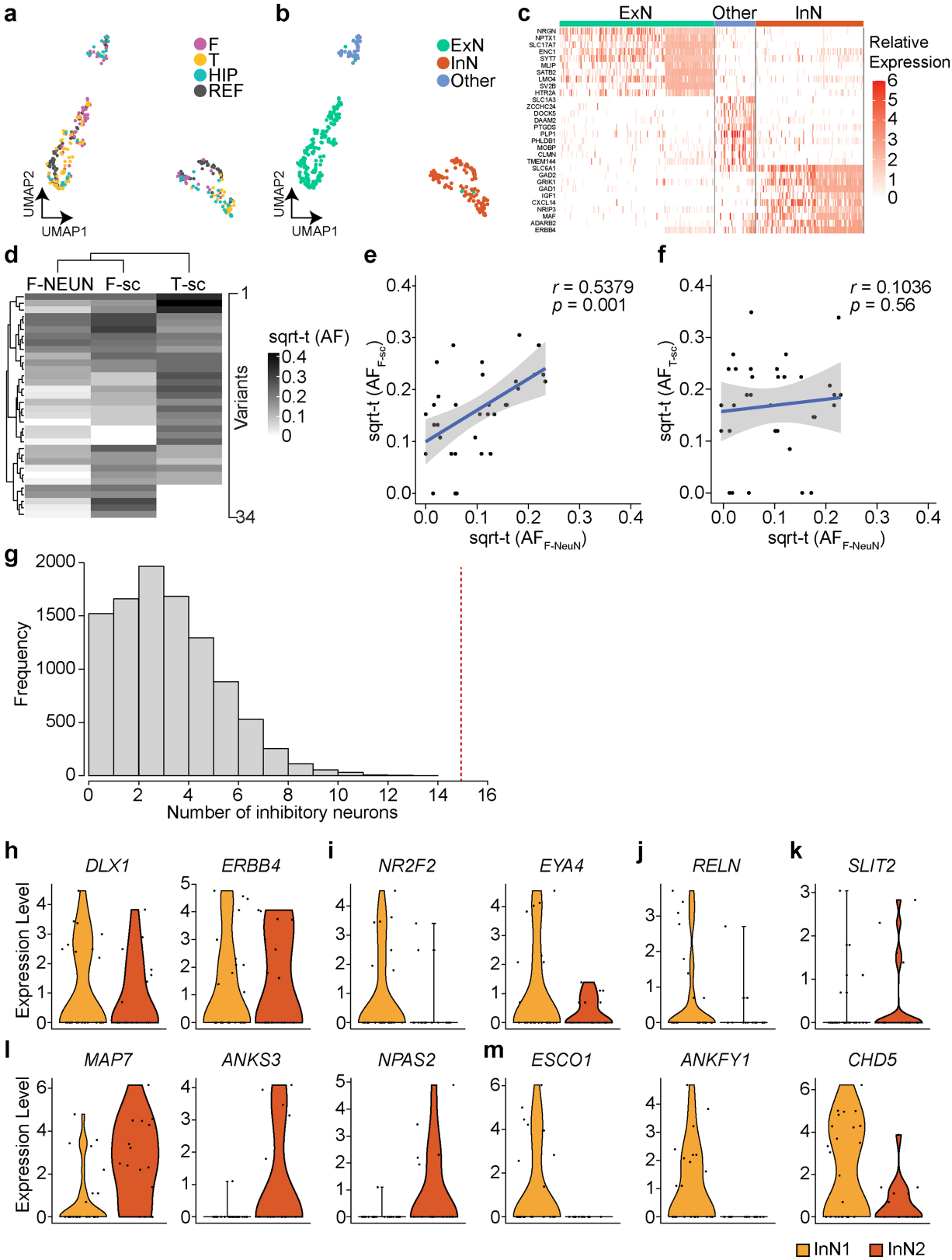
Extended Data Fig. 5 | UMAP relationships between samples from the brain based on AFs of validated MVs. Clustering by the same hemisphere validates lateralization of brain-derived cell clones except for the independent origin of microglia (marked by PU.1, arrow). (a-c) UMAP clustering in ID01 samples labeled by (a) cell type, (b) gross region, or (c) subregion, respectively. Clustered samples tend to show similar AF patterns. (d-f) UMAP clustering in ID05 samples labeled by (d) cell type, (e) region, or (f) subregion, respectively. Although PU.1 cells were not sorted in ID05, other findings are similar between ID01 and ID05.



Extended Data Fig. 6 | Evidence for HIP lineage restriction occurring prior to CTX or BG in ID01 sorted nuclear pools. (a) Heatmap with 17 sorted nuclear samples based on sqrt-t AFs of 121 informative MVs from ID01, similar to Fig. 2c, showing greater HIP lineage separation compared with CTX or BG (purple compared with green or yellow). (b) Contour plot (at center) with 121 informative MVs derived from (a) and two kernel density estimation plots (at periphery). Axes show the absolute normalized difference value for each MV between the average AF of CTX and BG (CTX-BG) or CTX and HIP regions (CTX-HIP). Solid line: identity. Red dot: averaged x and y values of individual data points. sqrt-t AF, square-root transformed allele fraction; CTX, cortex; BG, basal ganglia; HIP, hippocampus; Cau, caudate; DG, dentate gyrus; HIP and Hip, hippocampal tissue; I, insular cortex; O, occipital cortex; P, parietal cortex; PF, prefrontal cortex; Put, putamen; T, temporal cortex; GP, globus pallidus.



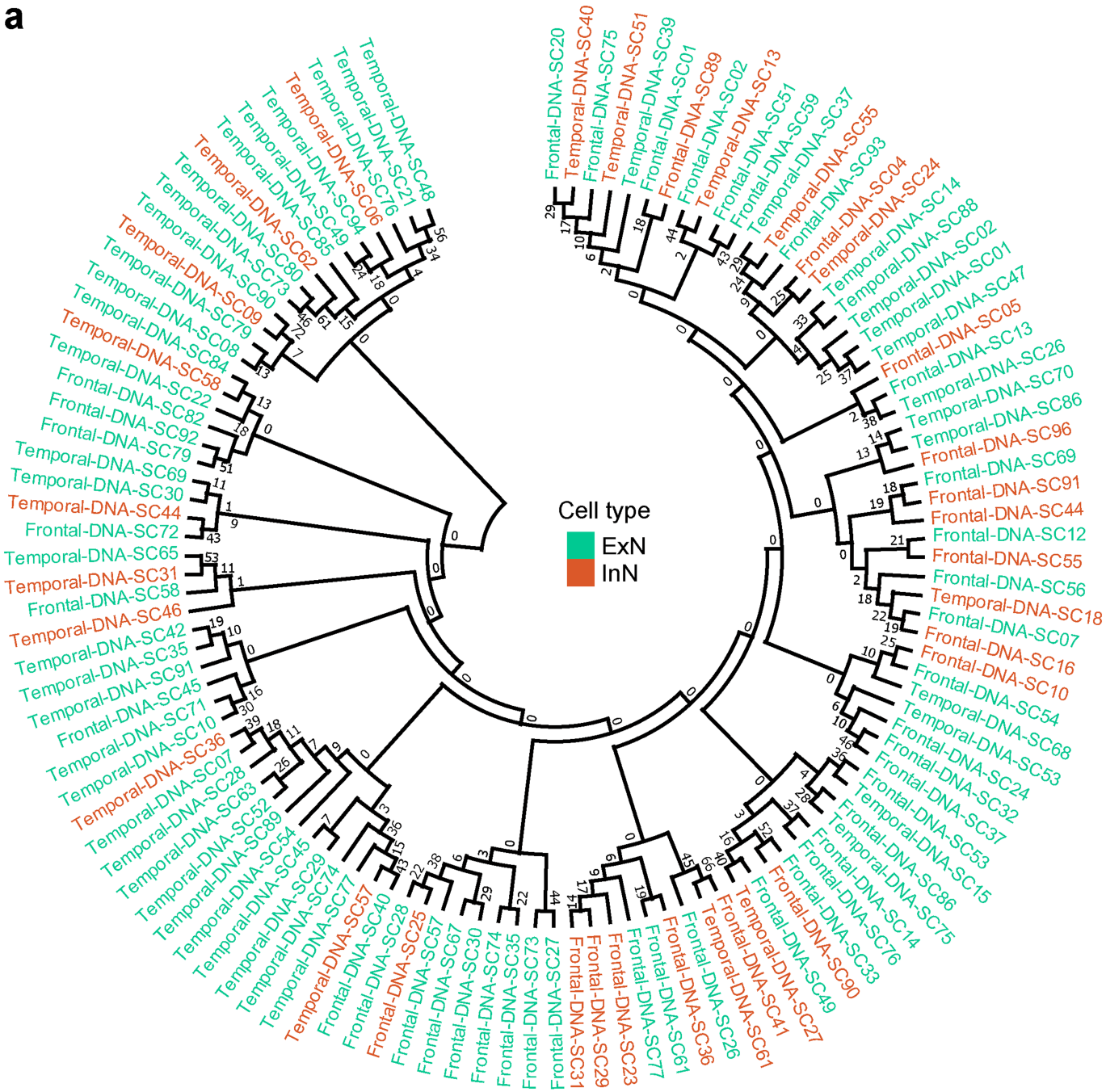
Extended Data Fig. 7 | Confidence for dendrograms with sorted nuclei from cortical areas. (a) Bootstrapping results of ID01. (b) Bootstrapping results of ID05. The percentage of 10,000 replicates showing relationships between sqrt-t AFs for TBR1$^+$ and DLX1$^+$ nuclei in the same geographic region were more similar than TBR1$^+$ nuclei from two different geographic regions. COUPTFII$^+$ nuclei clustered among themselves, outside of the DLX1 and TBR1 clusters. Approximated unbiased *p*-value > 95% (red): the hypothesis "the cluster does not exist" rejected with a significance level (<5%). (c-d) Heatmaps and hierarchical clustering results after computational deconvolution of DLX1$^+$ nuclei (grey) from Fig. 3b (c) and Fig. 3c (d). (e-f) Heatmaps and hierarchical clustering results after the simulated TBR1$^+$ nuclei contamination for COUPTFII$^+$ nuclear pools (black) from Fig. 3b (e) and Fig. 3c (f). (g) The estimated proportion of dorsally derived cortical inhibitory neurons within deconvolved DLX1$^+$ nuclei of each lobe. The least square method is used (Methods). 11, 13 cortical lobes for ID01 and ID05, respectively. Median, thick horizontal line at the center; 95% confidence intervals, the notch of the box plot; 75 and 25% quantiles of data, upper and lower bounds of the box; Whiskers, maxima and minima excluding outliers.



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

Extended Data Fig. 8 | Quality controls of the ResolveOME dataset in ID05. (a) A UMAP plot of snRNA-seq using 225 NEUN$^+$ nuclei and 121 aggregated reference cell types. F, frontal; T, temporal; HIP, hippocampus; REF, reference dataset. (b) UMAP labeled by cell types. Note that UMAP clusters separate by cell type (ExN, InN or Other) more than by location. (c) Relative expression of cell type markers within clusters, confirming cell identity. (d) Hierarchical clustering based on sqrt-t AFs of 34 informative MVs shared in 5 to 29 cells in single-nuclear data. F- NEUN, sorted frontal NEUN$^+$ nuclei pool; F-sc, pseudo-bulk snMPAS data from a frontal lobe punch; T-sc, snMPAS data from a frontal (F) lobe punch. (e) Correlation between sqrt-t AFs of MVs between F- NEUN and F-sc. (f) Correlation between sqrt-t AFs of MVs between F- NEUN and T-sc. In e and f, linear regression with upper and lower 95% prediction intervals displayed by blue solid lines and gray surrounding area; sqrt-t (AF), sqrt-t AF. Pearson's Product-Moment correlation with *r* and *p*-values (two-tailed) in e and f. (g) Null distribution of the frequency of the number of inhibitory neurons carrying MVs exclusively detected in one lobe and shared with at least two other local cells, including one excitatory neuron within the same lobe. 10,000 permutations. The portion to the right of the red dashed line, compared to the entire distribution, represents the probability ($p < 0.0001$, one-tailed permutation test) of having 15 or more InNs. (h-m) RNA expression levels of informative genes between InN1 ($n = 17$) and InN2 ($n = 16$) (Fig. 4b) in snRNA-seq. (h) Comparable expression levels of inhibitory neuronal markers between both groups. (i) Decreased tendency for the expression of CGE-derived cell markers in InN2 compared to InN1, implying COUPTFII+ inhibitory neurons are unlikely InN1, consistent with previous observations in sorted nuclear populations. (j) RELN$^+$ inhibitory neuronal marker showed decreased expression tendency in InN2 compared to InN1. (k) Increased expression tendency for parvalbumin-positive (PV$^+$) inhibitory neuronal marker in InN2 compared to InN1, implying dorsally derived inhibitory neurons include PV$^+$ neurons. (l, m) top 3 genes increased (l) or decreased (m) in InN2 compared to InN1 among the most variable 3000 protein-coding genes.

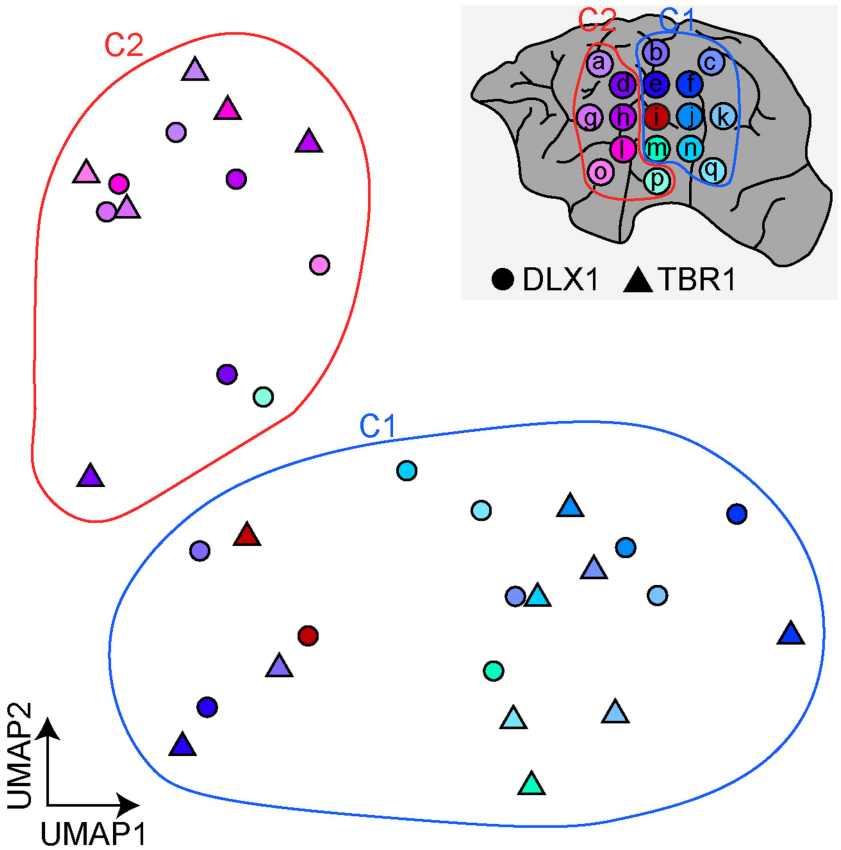
a



b

Same lobe	ExN-ExN	20
	ExN-InN	14
	InN-InN	1
Different lobe	ExN-ExN	2
	ExN-InN	4
	InN-InN	2

Extended Data Fig. 9 | Phylogenic tree analysis. (a) Phylogenic tree generated after 1000 bootstrap replications based on the 68 MVs in 118 single nuclei in Fig. 4b. Bootstrap values supporting each edge are labeled beside branches of the tree. (b) The number of pairs diverging from the latest branch that has the local highest-confident edge is shown based on the lobe and cell type. For example, the number of excitatory-excitatory neuron pairs within the same lobe clustered with the local highest-confident edge was 20.



Extended Data Fig. 10 | UMAP plots with sorted nuclear pools based on sqrt-t AFs of 186 informative MVs from Fig. 5. Colors of data points correspond to the spatial information in the grey box.

nature portfolio

Corresponding author(s): Joseph G. Gleeson

Last updated by author(s): Mar 4, 2024

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	WGS and MPAS/snMPAS were performed on Illumina instruments using their proprietary platform and DRAGEN platform (WGS only).
Data analysis	Mutect2 (v4.0.4.0 from GATK), MosaicHunter (single-mode, v 1.0), GATK (v 4.0.4), DeepMosaic (v 1.0.1), MosaicForecast (v 8-13-2019), Strelka2 (v 2.9.2), gnomAD (v 2.1.1), FlowJo (Ashland, Oregon), Bismark (v 0.23.1), Snakemake (v 6.12.3), BEDTools (v 2.30.0), Python (v 3.10), R (4.1.3), MEGA (v11.0.13), MUSCLE (v1), Cell Ranger (v6.0.2), Seurat (v4.0.5), fastqc (v0.11.8), cutadapt (v1.16), GRCh38 human genome and gencode (v27) gtf anntotation using STAR (v2.6.0c), SAMtools (v1.7), Picard (v2.20.7), qualimap (v2.2.2-dev), featureCounts (v2.0.0), rsem (v1.3.1), r-pvclust (v2.2.0), ComplexHeatmap (v2.16.0), r-ggplot2 (v3.4.3), maftools (v2.16.0), r-umap (v0.2.10.0), SAMBAMBA (v0.7.0). Scripts for variant filtering are provided on GitHub (https://github.com/shishenyxx/Human_Inhibitory_Neurons).

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

Raw whole genome sequencing and massive parallel amplicon sequencing data (MPAS) and single nucleus MPAS (snMPAS) are available through SRA (accession number: PRJNA799597) and NDA (accession number: study 919) for ID01 and ID05. The 300× WGS panel of normal is available on SRA (accession number: PRJNA660493).
human_g1k_v37 reference: http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/
gnomAD: https://gnomad.broadinstitute.org/
Human Multiple Cortical Areas SMART-seq dataset: https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-smart-seq
Details and codes for the data processing and annotation are provided on GitHub (https://github.com/shishenyxx/Human_Inhibitory_Neurons).

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	We use already expired postmortem organs from donated cadavers, which are not considered as human subjects.
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	ID01 were donated from a 70-year-old female, cause of death was 'global geriatric decline' with a contributing cause of 'post-surgical malabsorption'. ID05 were donated from a 73-year-old female, medical history indicated 'knee replacement, fractured pelvis, hernia, fractured fibula, hypothyroidism, empyema, pulmonary arterial hypertension, and scleroderma'. Both donors were documented to be of European ancestry. Organs were collected within a 26-hour postmortem interval for both donors (ID01: 24 hrs, ID05: 26 hrs). Prior medical history showed no signs of neurological, psychiatric or cancer diseases for either, and tested negative for infection with HIV, Hepatitis B, or COVID-19.
Recruitment	no participants were recruited.
Ethics oversight	According to 45 CFR 46.102(e)(1), The use of human anatomical cadaver specimens of ID01 and ID05 are exempt from oversight of the University of California, San Diego Human Research Protections Program (IRB) but are subject to oversight by the University of California, San Diego Anatomical Materials Review Committee (AMRC). This study was overseen and approved by the AMRC. The approval number is 106135. Donors met AMRC qualifications: (i) Obtain information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (ii) Obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens.

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	This is a pioneer descriptive study with limited expectations. The number of individuals included was limited by the availability of biological sample (due to postmortem interval, history of disease, etc). Samples from both hemispheres and all neocortical lobes as well as non-brain organs were included for ID01 and ID05.
Data exclusions	We did not exclude any generated sequencing data, but filtered detected variants as described in the methods.

Replication	This study consists two cadavers ID01 and ID05.
Randomization	This is a descriptive study, and no ramdomization was performed. Cadavers with short PMI are very rare, and it is very difficult to collect enough number of cadavers for randomization. We collected any of cadavers meeting our criteria.
Blinding	No groups were allocated by the scientists.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	NEUN Alexa Fluor 488 (Clone A60, 1:2,500; Millipore Sigma, MAB377, RRID:AB_2298772), TBR1 (1:1,000; abcam, ab31940, RRID:AB_2200219), DLX1 (1:200; Atlas Antibodies, HPA045884, RRID:AB_10960361), COUPTFII (1:400; Novus Biologicals, PP-H7147-00, RRID:AB_1964214), DARPP32 (1:400; abcam, ab40801, RRID:AB_731843), goat anti-rabbit Alexa 647 (1:4,000; ThermoFisher Scientific, A21244, RRID:AB_2535812) and donkey anti-mouse Alexa 647 (1:4000; ThermoFisher Scientific, A32787, RRID:AB_2762830).
Validation	The validation results of all antibody specificity to formaldehyde-fixed nuclei was published in PMID:35444276 or 24561062. We further validated the performance of DLX1, COUPTFII and TBR1 antibodies under the MFNS condition through single-cell RNAseq or bisulfite sequencing in this study.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	<Formaldehyde-fixed nuclear preparation for sorting> For DARPP32/NEUN dual staining with BG nuclei, frozen Cau and Put of ID01 were homogenized in 1% formaldehyde in Dulbecco's phosphate-buffered saline (DPBS, Corning) using a motorized homogenizer (Fisherbrand PowerGen 125), then incubated on a rocker at room temperature for 10 min, quenched with 0.125 M glycine at room temperature on a rocker for 5 min, then centrifuged at 1,100×g in a swinging bucket centrifuge. The following steps were all performed on ice except where indicated. Homogenates were washed twice with NF1 buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 5mM MgCl2, 0.1M sucrose, 0.5% Triton X-100 in UltraPure water) and centrifuged at 1,100×g for 5 min at 4°C in a swinging bucket centrifuge. Next, pellets were resuspended in 5 ml NF1 buffer and Dounce homogenized 5x in a 7 ml Wheaton Dounce Tissue Grinder (DWK Life Sciences) using a 'loose' pestle. After 30 minutes of incubation on ice, homogenates were Dounce homogenized 20x with a 'tight' pestle and filtered through a 70 µm strainer. To remove myelin debris, homogenates were overlaid on a sucrose cushion (1.2M sucrose, 1 M Tris-HCl pH 8.0, 1 mM MgCl2, 0.1 M DTT) and centrifuged at 3,200×g for 30 min with acceleration and brakes on 'low'. Pellets of nuclei were washed with NF1 buffer and centrifuged at 1,600×g for 5 min and stored at -80°C, same as documented. <Nuclear preparation for MFNS or unfixed nuclei sorting> For the MFNS protocol, approximately 200 mg of freshly frozen tissue stored at -80°C was prepared, and subsequent

procedures were conducted using solutions maintained at 4°C. The prepared tissue was homogenized in 300 µl of lysis buffer (composed of 10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40, and 1 mM Dithiothreitol in nuclease-free water) while kept on ice. Next, additional 9.7 ml of lysis buffer was added to the homogenate, and the mixture was incubated on ice for 5 minutes. The homogenate was then passed through a 70 µm cell strainer (FALCON, 352350) and centrifuged at 1100g for 5 minutes at 4°C. The supernatant was discarded, and the remaining pellet was gently resuspended and washed with 10 ml of sorting buffer (containing 1% bovine serum albumin, 1 mM EDTA, and 10 mM HEPES in 1x HBSS solution). For the density gradient centrifugation step, the resuspended pellet in 25% Iodixanol solution (OptiPrep™, Millipore D1556) was layered onto a 29% Iodixanol cushion. The centrifugation was carried out at 10,000g with a swinging bucket rotor, employing low acceleration and braking, at 4°C for 40 min, the pellet resuspended in 80% methanol prechilled and stored at -20°C for at least 30 minutes before further use. For samples to perform single-nucleus transcriptome and PTA with ResolveOME, brain tissue from participant ID05 were homogenized in 1 ml of ice-cold NIB (composed of 0.25 M sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM Tris pH 7.5, 100 mM DTT, and 0.1% Triton X-100) and subjected to homogenization, incubated on a rocker for 5 min at 4 °C and then centrifuged at 1,000g employing low acceleration and braking in a swinging-bucket centrifuge, and the pellet was reconstituted in 0.5 ml of sorting buffer and filtered through a 70-um strainer. Nuclei in the flow-through were immediately subjected to staining.

For pellets of defrosted and homogenized brain nuclei were washed twice in sorting buffer and then re-suspended in 0.2 ml sorting buffer and incubated overnight at 4°C. The following antibodies were used: NEUN Alexa Fluor 488 (1:2,500; Millipore Sigma, MAB377), TBR1 unconjugated (1:1,000; Abcam, ab31940), DLX1 (1:200; Atlas Antibodies, HPA045884), COUPTFII (1:400; Novus Biologicals, PP-H7147-00), and DARPP32 (1:400; Abcam, ab40801). The following day, nuclei were washed with staining buffer and in case an unconjugated antibody was used, nuclei were stained subsequently for 30 minutes with goat anti-rabbit Alexa 647 (1:4,000; ThermoFisher Scientific, A21244) for TBR1, DLX1 and DARPP32, or donkey anti-mouse Alexa 647 (1:4000; ThermoFisher Scientific, A32787) for COUPTFII. Stained nuclei were washed one more time with staining buffer and passed through a 70 µm strainer. Immediately before the sort, nuclei were stained with 0.5 µg/ml DAPI. For single-nucleus transcriptome and whole-genome amplification with ResolveOME, nuclei were stained with Propidium Iodide (1:20; Invitrogen, BMS500PI). Nuclei for the cell type of origin were sorted either on a MoFlo Astrio EQ sorter (Beckman Coulter), BD FACSAria II (Becton-Dickinson), or BD InFlux Cytometer (Becton-Dickinson) similar to previous documentation17. At least 1000 methanol-fixed or >50,000 formaldehyde-fixed sorted nuclei were pooled in each tube. Sorted nuclei were pelleted in staining buffer at 1,600×g for 10 minutes. Nuclei for DNA extraction and bisulfite sequencing were stored at -80℃. FANS data was visualized using FlowJo v10 software (Ashland, Oregon). Following MPAS (see below) sorted populations were deemed to be of sufficient overall quality if at least 95% variants were sequenced above >1,000×.

Instrument

Nuclei for the cell type of origin were sorted either on a MoFlo Astrio EQ sorter (Beckman Coulter), BD FACSAria II (Becton-Dickinson), or BD InFlux Cytometer (Becton-Dickinson) similar to previous documentation (PMID 35444276).

Software

FANS data was visualized using FlowJo v10 software (Ashland, Oregon).

Cell population abundance

Single-cell RNAseq or bisulfite sequencing was performed to validate cell type identity.

Gating strategy

For the population level sorting, we selected of DAPI positive nuclei, gating on singlets. Excitatory and inhibitory neurons were sorted by using TBR1 and DLX1 or COUPTFII antibodies. For single nuclei sorting, doublets were excluded and nuclei positive for PI were further gated into NeuN+ populations.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.