Perinatal Elimination of Genetically Aberrant Neurons from Human Cerebral Cortex

Diane D. Shao[1,2*], Yifan Zhao[3,4*], Joseph Brew[3], Urmi Ghosh[2,5], Lovelace J. Luquette[3], Sijing Zhao[2,6], Andrea J. Kriz[2], Xuyu Qian[2,5], Vinayak V. Viswanadham[3], Takumi Taketomi[7], Fuminori Tsuruta[7,8], Peter J. Park[3†], Christopher A. Walsh[1,2,5†]

[1]Department of Neurology, Boston Children's Hospital, Boston, MA, USA.
[2]Department of Pediatrics, Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA.
[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.
[4]Program in Health Sciences & Technology, Harvard Medical School & Massachusetts Institute of Technology, Boston, MA, USA.
[5]Howard Hughes Medical Institute, Chevy Chase, MD, USA.
[6]Biological and Biomedical Sciences Graduate Program, Harvard Medical School, Boston, MA, USA.
[7]Ph.D. Program in Human Biology, University of Tsukuba, Tsukuba, Ibaraki, Japan.
[8]Institute of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan.

[†]Corresponding authors: peter_park@hms.harvard.edu; christopher.walsh@childrens.harvard.edu;
[*]These authors contributed equally to this work.

Abstract

Human neurons are postmitotic and long-lived, requiring precise genomic regulation to maintain function over a lifetime. Normal neuronal function is highly dependent on gene dosage, with copy number variants (CNVs) and heterozygous point mutations associated with a host of neurodevelopmental and neuropsychiatric conditions [1–3]. Here, we investigated the landscape of somatic CNVs arising in fetal human brains, and how they change over development, to understand the processes that generate neurons with low rates of aneuploidy. We identified 7,725 CNVs in >1,200 single neurons from human postmortem brain of 16 neurotypical individuals, ranging in age from gestational week 14 to 90 years old using Tn5-transposase-based single-cell whole-genome amplification. We surveyed CNVs in another 44,861 nuclei with 10X Multiome analysis. Up to 45% of postmitotic neurons in the prenatal cortex showed aberrant genomes, characterized by widespread CNVs of multiple chromosomes, but this reduces sharply after birth ($p<0.01$). We identified micronuclei in the developing cortex in situ, reflecting chromosomal material missegregated during neurodevelopment [4–6]. The size of CNV appeared to define the trajectory of neuronal elimination, since cells with widespread CNVs were eliminated earlier and faster than cells with smaller CNVs. CNVs in surviving neurons were depleted for genes that are dosage-sensitive or involved in neurodevelopmental disorders ($p<0.05$), suggesting selective elimination of neurons with CNVs involving these critical genes. Neurons with high CNV burdens also showed abnormal expression of synaptic gene sets, indicating that abnormal synaptic gene regulation may contribute to neuronal elimination. Elimination of defective neuronal genomes during during synaptogenesis may represent a critical process of genome quality control and a vulnerable target of factors that contribute to neurodevelopmental disease.

Introduction

Genome-scale copy number variants (CNV) are common in early post-zygotic development–with up to 73% of embryos exhibiting large chromosomal aneuploidy [5, 7–9]. Similarly, human neural progenitor cells (NPCs) also display extensive mosaic aneuploidy, involving up to 30% of progenitors [10, 11]. Rates of aneuploidy in adult neurons have been controversial, with some reports of aneuploidy [12], while other reports suggest frequent, predominantly sub-chromosomal CNVs [13–15]. These discrepancies arise from differences in methodology and the sensitivity of techniques, making it unclear whether the high aneuploidy rates observed early in development are maintained through to adulthood in the brain.

The control of copy number is especially crucial in brain, since genetic variation in neurons can have profound functional impact. Autism, schizophrenia, epilepsy, intellectual disability and other neuropsychiatric conditions are all commonly caused by chromosomal or sub-chromosomal CNV, as well as by heterozygous loss of function point mutations in more than one thousand genes with essential brain functions [16–18]. Many of these genes encode structural proteins of the synapse, or regulators of synaptic plasticity, suggesting that key aspects of neuronal plasticity underlying these conditions are intensely sensitive to gene dosage [19, 20].

While a variety of cell cycle checkpoints regulate progenitors to prevent or eliminate cells with aberrant genomes [12, 21–23], it is not clear whether any such pathways might exist to control gene dosage in noncycling cells like neurons. One hypothesis suggests that genetically aberrant neurons may eliminated through "programmed cell death" [22, 24], a form of widespread apoptosis which in human cerebral cortex is estimated to remove 30-60% of neurons during development [21, 25]. In all examined mammalian species, the most dramatic wave of apoptosis eliminates ∼30% of neurons in the perinatal period [26], followed by ongoing elimination of postnatal neurons [27, 28] coincident with the establishment of neuronal circuits and connectivity mediated by synapse formation and pruning. While roles for synapse formation in the regulation of neuronal survival have been described for more than a century [29, 30], the mechanisms and significance of widespread neuronal loss have not been clear, since dying neurons tend to be dispersed across the cortex, without clear relationship to neuronal cell type, laminar location, or topographic patterns of connectivity [22, 31, 32].

Using single-cell whole genome sequencing (scWGS), we compared copy number variants (CNVs) in postmitotic neurons during neurodevelopment and aging. Our analysis shows widespread CNVs across chromosomes in fetal neurons, followed by removal of most neurons with highly defective genomes around the time of birth. We also confirmed prior findings of a slower removal of defective cells occurring postnatally into aging [13], suggesting that potential mechanisms of genome quality control in postmitotic neurons preferentially eliminate cells with aneuploid genomes. Our data also show that this genome quality control mechanism is linked to gene pathways essential for synapse development, including disease-relevant genes. These findings suggest that synaptogenesis plays a key role in genome quality control in neurons that acts most dramatically near birth, but extends over time. This process is potentially vulnerable to environmental or genetic defects that disrupt normal brain development.

Results

Accurate copy number detection in single neurons identifies aberrant genomes in prenatal cortical neurons

To profile CNVs in the human prenatal brain, we generated low-coverage scWGS data from 1,298 neurons, collected from non-diseased postmortem brains from (1) seven prenatal donors ranging from gestational weeks 14-25 and (2) nine postnatal brains from infancy to 90 years (Fig. 1a, Table 1). To obtain a pure population of postmitotic prenatal cerebral cortical neurons, we cryo-dissected the human prenatal cortical plate (Fig. 1b) followed by flow cytometry to isolate single neuronal nuclei. Dissection of cortical plate neurons provides an exclusively postmitotic, noncycling fetal neuronal population, because NPCs and glial nuclei are located in undissected deeper cortical regions. Neurons migrate into the cortical plate after withdrawing from the cell cycle and becoming permanently postmitotic. Single-nucleus RNA sequencing (snRNA-seq) confirmed that our cortical plate dissection yielded 98% purity for neurons (Fig. 1b, Extended Data Fig. 1a). Postnatal nuclei suspensions were stained with anti-NeuN antibody (neuronal marker) and DAPI prior to flow cytometry to isolate neuronal nuclei as previously described [33] (Extended Data Fig. 1b). scWGS libraries were generated using Tn5-based amplification (TbA) and sequenced at ∼0.05X depth. We performed stringent quality control to distinguish high-quality vs poor-quality libraries (Extended Data Fig. 1c-j; Methods), resulting in 1,084 high-quality single-cell neuronal genomes.

To identify CNVs, we utilized HiScanner [34], an algorithm we previously developed specifically for CNV detection in scWGS data. For robustness against amplification noise and sparse sequencing coverage, HiScanner models technology-specific allelic dropout patterns from scWGS data and leverages probabilistic segmentation to maintain high detection precision. [35, 36]. In addition, we designed another CNV validation approach: Fragment Overlap Density (FOD) metric. For Tn5-transposase-based libraries, we expect all amplified fragments to retain the start and end position of the original tagmented molecule; thus, in a copy-loss region where only one allele is present, no fragments should overlap, unlike in copy-neutral or copy-gain regions (Fig. 1c). This heuristic is represented by the FOD metric (Methods), which we use to further increase our CNV detection accuracy (Supplemental Results; Extended Data Fig. 2).

Surprisingly, we found that neurons in the human cerebral cortex with aberrant genomes were strikingly abundant in prenatal neurons (Fig. 1d), but constitute a smaller proportion of postnatal neurons (Fig. 1e). Many genomes with alterations had multiple affected regions across more than one chromosomes, often involving large copy number changes. In prenatal samples, copy losses were more frequent than copy gains, with 66% (3293/4981) of CNVs being copy losses, and 34% (1688/4981) copy gains. In postnatal neurons, copy losses (74%, 2034/2744) were again more frequent than copy gains (26%, 710/2744) (Fig. 1d-e, insets).

Copy number signatures classify aberrant genomes with widespread CNVs

Using mutational signature analysis [37, 38] based on the COSMIC copy number signature catalog [38], we classified the cells into four distinct "signature groups": *Diploid Predominant* (COSMIC SigCN1 signature), *Gain Predominant* (combination of SigCN1 and SigCN2), *Focal Loss Predominant* (SigCN1 and SigCN9), and *Broad Loss Predominant* (SigCN1 and SigCN13) (Fig. 2a, Extended Data Figs. 3 and 4). The *Diploid Predominant* group consisted primarily of diploid cells, with a small subset harboring sub-chromosomal CNVs (hereafter called the SigCN1-specific CNV neurons). The other three groups showed alterations throughout the genome, involving multiple chromosomes (hereafter called the widespread CNV neurons). The neuronal copy number signatures showed high resemblance (cosine similarities 0.92-0.98; Extended Data Fig. 3d) to those defined in COSMIC from

cancers, suggesting that neuronal CNVs may have arisen in dividing cells, presumably NPCs, in the developing human brain.

The proportion of neurons with widespread CNVs (last 3 of the 4 groups defined above) in prenatal samples was on average 29.1%; this proportion decreased significantly to ∼10.9% postnatally (p=0.0082; Student's t-test, two tailed) (Fig. 2b). The mean during infancy (0-2 years) was 12.4% and slowly declined to 7.2% for elderly samples (>60 years) (Fig. 2c). The postnatal slow decline of the widespread CNV proportion is significantly different from the faster postnatal decline in the proportion of neurons with smaller, SigCN1-specific CNVs (p=0.04; Extended Data Fig. 5a-b). The SigCN1-specific CNV proportion declines with age over a lifetime, from ∼35% in infants to <10% in elderly individuals (Extended Data Fig. 5b), similar to prior reports of sub-chromosomal CNVs [13, 14]. In neurons with SigCN1-specific CNVs, the sizes of copy losses were significantly smaller in postnatal compared to prenatal samples (p<0.001, Student t-test two-tailed, Extended Data Fig. 5c-d). Thus, it appears that larger CNV size is associated with higher likelihood of cellular elimination and more rapid elimination, whereas neurons with smaller CNVs continue to be eliminated more gradually postnatally.

To ensure that the observed reduction in the proportions of neurons with aberrant genomes are not affected by poor quality cells, we checked that there is no significant difference in the percentage of cells passing QC filters (Extended Data Fig. 6a-b). Linear regression on various factors showed that age is the only significant factor associated with the decrease in proportion of neurons with widespread CNVs after birth (p=0.023). Repeating the same analysis using a high confidence set, i.e. those cells with highest genome coverage (>100M reads), recapitulated the results showing a higher proportion of neurons with widespread CNVs prenatally (Extended Data Fig. 6c-d).

Another potential confounder we analyzed was whether disruption of nuclear integrity may falsely elevate CNV rates, given that our tissues were acquired postmortem. When we disrupted nuclear integrity of cultured NPC cells by applying 24 hours of staurosporin treatment (Extended Data Fig. 7a), we found that cells with disrupted nuclear integrity had a widespread CNV rate of only 9%, after exclusion of cells that fail QC. Thus, we concluded that nuclear disruption would not falsely account for the large proportion of widespread CNV that we observed in postmortem brains. Further, the same four signature groups were seen in the NPC mutational signature decomposition as in the human brain (Extended Data Fig. 7b), indicating a consistent biological etiology for the CNVs observed in cultured NPCs.

Neuronal CNVs result from genomic stress in progenitor cells

To determine whether CNVs arise in replicating progenitor cells or postmitotic neurons, we analyzed their associations with DNA replication timing [39, 40], and physiological double strand breaks (DSBs) [41, 42]. We found that CNV breakpoint density was highly correlated with replication timing (r^2=0.88, p=0.007), such that CNV breakpoints were more frequent in early-replicating regions (Fig. 3a-b, Extended Data Fig. 8a). This result suggests that some CNVs are generated in mitotic progenitor cells during DNA replication. We next evaluated hotspots of DSBs in NPCs determined previously by in-suspension Breaks Labeling In Situ and Sequencing (BLISS) [41]. These DSBs are known to arise in neurons at sites related to transcription, topologically associating domains, and chromatin loop anchors, and have been associated with CNV generation in neurological disease [42]. In our data, BLISS hotspots (top 1% of BLISS regions) were significantly associated with CNV breakpoints (p¡0.0001; Extended Data Fig. 8b). While BLISS hotspots and replication timing have a minor correlation (r^2=0.27), the significantly strong association of CNV breakpoints with both features suggests two potential mechanisms of CNV generation: (1) errors during DNA replication in dividing progenitor cells, and (2) DSB-mediated CNV formation in progenitors or early postmitotic neurons.

DNA replication stress has been associated with the formation of micronuclei, which encapsulate missegregated chromosomes or fragments that lie outside of the nucleus, as observed in both early

postzygotic cell divisions and cancer [5, 43, 44]. Thus, we hypothesized that micronuclei would be visible in some human prenatal neurons, as they have also recently been observed in mice [45]. High-resolution microscopy on fixed slices of human fetal brain at mid-gestation, stained with DAPI, revealed co-occurrence of micronuclei in cells with neuronal identity as assayed by immunostaining with the neuronal marker, SATB2 (Fig. 3c). We observed that micronuclei were associated with ∼6% of primary nuclei in the cortical plate (Table 2) providing direct, *in situ* demonstration of relatively common, missegregated chromosomal material at midgestation.

We sought to orthogonally confirm our observations that the proportion of cells with widespread CNVs are eliminated over time by evaluating whether micronuclei are also cleared over time (Fig. 3d-f). We isolated micronuclei from human brain tissue from fetal, infant, and elderly individuals by serial gradient sucrose fractionation with visualization to identify micronuclei-enriched genetic content as previously described [46] (Fig. 3d). We sequenced the micronuclear and nuclear fractions separately and compared the genetic content of the two fractions from each sample (Fig. 3e-f, Table 3). We found that micronuclei in the prenatal samples harbored aberrant genetic material, observed as a high deviation from the read depth compared to that of the nuclear fraction (Fig. 3e). The magnitude of this deviation decreased in infancy and aged brains, indicating clearance of micronuclei genetic material over time (Fig. 3f). Similar to prior reports of micronuclei in other systems [5, 44, 47], micronuclei in the human brain preferentially contained genetic content from larger chromosomes (Extended Data Fig. 9), with the exception of overrepresentation of chromosomes 13, 18, 21, X and Y in micronuclei. This trend likely reflects a relative tolerance to aneuploidy of these specific chromosomes in humans, since these chromosomes represent the most common aneuploidies viable in human germline [48]. The clearance of micronuclear material with age thus provides orthogonal evidence for the reduction of neurons with widespread CNVs with age.

Neuronal CNVs disrupt gene modules relevant to synaptic signaling and maintenance

In order to identify pathways regulating the elimination process by associating CNV changes in single neurons with gene expression, we performed 10X multiomic snATAC/RNA-seq of 44,861 single nuclei (Fig. 4a, Extended Data Fig. 10). We generated 10X multiome libraries from three prenatal and five postnatal brains, followed by dimensionality reduction and cell type annotation using Seurat [49] and CNV calling using epiAneufinder [50]. We used snRNAseq to determine cell types and generate gene module scores for each gene set, based on the Gene Ontology database. Unlike CNV calling from scWGS, CNV calling from snATAC lacks sufficient resolution for signature decomposition of CNVs, so we segregated nuclei broadly into "CNV neurons" vs "non-CNV neurons." To identify CNV neurons, we inspected the distribution of the proportion of genomic regions affected by CNVs and, using Tukey's rule, classified outlier cells as CNV neurons (Extended Data Fig. 10a).

Our findings from snATAC confirmed a significant reduction of CNV neurons between prenatal to postnatal ages (Fig. 4b), suggesting that these CNV neurons identified by snATAC are most similar to the widespread CNV observed in scWGS. We confirmed that CNV neurons have high cell quality metrics consistent with the rest of the dataset (Extended Data Fig. 10b). We found CNV neurons for excitatory, inhibitory, and maturing neuron subsets, as well as CNV progenitor cells, as expected (Extended Data Fig. 10c). CNV analysis among glial cells did not detect differences in CNV rate prenatally versus postnatally, suggesting that the perinatal removal of CNV neurons identifies a process specific to postmitotic neurons and not shared by glial cells (Extended Data Fig. 10d).

In order to determine pathways that are specifically dysregulated in CNV neurons that might drive cellular elimination, we inferred gene modules that were differentially expressed ("Differential Gene Modules", DGMs) between CNV neurons and non-CNV neurons. The significant DGMs associated with postnatal CNV neurons (Supplemental Table 1) frequently represented components of synaptic function and connectivity (Fig. 4c). To ensure that this enrichment did not simply reflect

that our analysis focused on neurons, we developed a null background by permuting CNV status within cell types. We determined that these synaptic DGMs distinguish a subset of excitatory neurons across all layers of the cortex, and a representative DGM is shown (Fig. 4d). In progenitor cells, "CNV progenitor cells" are significantly enriched for known gene modules related to cell cycle checkpoints, e.g., CDKN1A- and TP53-related pathways (Extended Data Fig. 10e). These results highlight differential regulation of dividing CNV progenitor cells versus postmitotic CNV neurons, with CNV neurons demonstrating significant differences in synapse function gene expression compared to non-CNV neurons.

Our differential gene module analysis revealed that synaptic gene modules are significantly downregulated in CNV neurons compared to non-CNV neurons during the postnatal period, coinciding with the established timelines for neuronal elimination via programmed cell death. This process is characterized by extensive neuronal elimination around birth, followed by ongoing cell death postnatally through processes thought to foster appropriate synaptic connectivity during brain development [22, 26, 51–53]. Hierarchical clustering of CNV neurons based on synaptic and cell death gene modules revealed two distinct subpopulations: one subset (hereafter, "Cluster 1") exhibited reduced synaptic module expression and up-regulated cell death pathways (apoptosis and autophagy) (Fig. 4d-e), while the remaining neurons ("Cluster 2") maintained normal to high synaptic module expression with low cell death pathway activation. The proportion of CNV neurons in Cluster 1 was highest in infant and youth samples and declined with age (Fig. 4f), consistent with the timeline of synaptic connectivity and pruning during early life [27, 28]. Notably, Cluster 1 CNV neurons were absent in our midgestation samples, and limitations in third-trimester sample availability prevented assessment of late gestation. These findings suggest that Cluster 1 gene expression patterns emerge sometime during the perinatal period, and that the processes of synaptic function and cell death contribute to the decline in CNVs observed in the postnatal period.

Genetic selection at key neurodevelopmental loci

Since the elimination of neurons across perinatal and postnatal ages favors survival of euploid neurons, we sought to identify genetic loci (particularly coding genes contained therein) that are preferentially preserved or disrupted by neuronal CNVs, in order to identify signals of genetic selection. Using our scWGS data from 1,298 neurons, we identified regions of the genome that were "coldspots," regions rarely or never disrupted in surviving neurons–indicating negative selection as disruption of these regions are poorly tolerated–or "hotspots," regions recurrently disrupted in surviving neurons that imply CNVs at these loci are tolerated and thus not negatively selected. To maximize analytical power, we included breakpoints from CNVs of all sizes. "Coldspots" frequently overlapped regions with higher gene density (Fig. 5a), a pattern not observed for "hotspots."

Given this observation, we hypothesized that "coldspots" may harbor key neurodevelopmental genes, whose loss neurons or progenitors may not tolerate. We assessed "coldspots" for enrichment of autism (SFARI [16]), epilepsy (OMIM [54]), neurodevelopmental disorders [55], and schizophrenia [55] gene sets (Fig. 5b). Autism genes, which frequently regulate synapse formation, maintenance, and pruning [26, 51–53], were significantly associated with "coldspots." Further analysis using a proteomic dataset of high-confidence postsynaptic genes revealed that these synaptic proteins were also significantly enriched within "coldspots" (Fig. 5b), reinforcing the critical role of synaptic genes in neuronal genome selection. "Coldspots" were also enriched for loss-of-function intolerant genes (Gnomad [56] pLI > 0.9) and for housekeeping genes critical to intrinsic cell function and showed no enrichment of cancer genes (COSMIC Consensus [57]). Collectively, these findings suggest that negative selection against copy number loss in autism-related and synapse-associated genes shapes the neuronal genetic landscape. Coupled with our DGM analyses, these data imply that regulation or disruption of synaptic gene modules—whether via large-scale genomic alterations or through gene-specific loss of synaptic proteins—critically influences neuronal survival during neurodevelopment.

Discussion

Our data suggest that widespread CNVs are a prominent feature of postmitotic neurons during human neurodevelopment and are strikingly dynamic over time. These genetic abnormalities share characteristics with those seen in early post-fertilization zygotes, including signatures of genetic stress and the formation of micronuclei [4, 5, 44]. We describe prenatal neurons with widespread CNVs which undergo two distinct phases of elimination, a dramatic reduction of widespread CNV neurons during the perinatal period, and a slower ongoing removal of both widespread CNVs and SigCN1-specific CNVs over the course of the lifespan (Fig. 6). This pattern parallels two waves of neuronal cell death described in cerebral cortex development and maturation [22].

Our data indicate that previously described neurobiological processes of cellular apoptosis and programmed cell death [22, 27, 28] may represent a core mechanism of genetic control for postmitotic neurons. We show that the dysregulation of synaptic gene modules in a subset of neurons with high CNV burden contributes to cellular elimination, along with selective pressure against dosage changes in critical neurodevelopmental genes related to synaptogenesis. Correct synaptic connectivity is critical in the human cerebral cortex for cell survival [58], and competition among neurons for synaptic space is known to strongly modulate neuronal cell death [26, 51–53]. We show that neurons with high CNV burdenshow dysregulated expression of synaptic pathway genes and negative selection against neurons with disruption of autism-related genes and other dosage-sensitive genes. The force of selection is likely related to the extent of CNVs and their impact on synaptic function, with widespread CNVs and gross aneuploidy predicted to impact fitness more severely than smaller CNVs. The perinatal timing of the reduction of prenatal neurons with widespread CNVs proportionally mirrors the time course of a large wave of neuronal loss of ∼30% occurring during late gestation [59–62], a period of critical neuronal circuit formation, though patterns of cell death in humans have not been characterized as fully as in model mammalian species due to the limited availability of appropriate tissue.

Our findings of widespread CNV prenatally in noncycling neurons of the human brain was unexpected. Prior studies of aging human cortex focused on postnatal neurons, and described elimination of subchromosomal CNVs [13–15] with age—similar to our *Diploid Predominant* SigCN1-specific CNV (Fig. 6). Isolated observations of abnormal karyotypes in postnatal neurons similar to our widespread CNV neuron population have been made [14, 15, 63] but not systematically investigated. Recently widespread aneuploidy was described in murine cortical neurons during murine midgestation [64], suggesting mechanisms of neuronal genetic variability and selection during neurodevelopment that may be conserved across species.

Given that persistence of aneuploid neurons would disrupt neuronal circuit function, the perinatal period through adolescence emerges as a critical developmental window ensuring that neurons surviving into adulthood are predominantly euploid. Further refinement of the cellular genetic landscape then occurs throughout life, likely with some interplay with the functional importance of individual genes. The relative absence of CNV that encompass dosage-sensitive and synapse-related genes in surviving neurons suggests intrinsic negative selective mechanisms during central nervous system development. We propose a core selective mechanism governing genomic integrity in postmitotic neurons. Future studies will explore how disruptions to the processes of perinatal neuronal genomic selection, through environmental or genetic influences, may permit genomically unfit neurons to persist, with potentially profound implications for brain function. Finally, age-related or neurodegeneration-related accumulation of mutations [33, 65–68] in neurons also disrupt essential neuronal genes [33], so extension of neuronal genome quality-control mechanisms into advanced age may have broad impact for long-term brain health and aging.

Supplementary information. Supplemental materials accompanying this paper include Supplemental_Information.pdf and Table_S1.xlsx.

Acknowledgements. We acknowledge the sacrifices of the donors who provided tissue for this study to advance our understanding of the brain. Tissue sources were provided by the University of Maryland Brain and Tissue Bank. We acknowledge the Boston Children's Hospital Flow Cytometry Core for developing and carrying out protocols to isolate single neurons. We acknowledge David S. Pellman and Greg Brunette for thoughtful discussion about identifying evidence for genetic stress. We acknowledge Peter Ly for sharing of micronuclei isolation protocols and helpful discussion. We acknowledge Takeshi Nagata for supporting micronuclear analysis.

Declarations

Funding. D.D.S. is supported by National Institutes of Health (NIH) National Institutes of Neurological Disorders and Stroke (NINDS) K08-NS128074. D.D.S., U.G., and C.A.W. are supported by the Templeton Foundation. C.A.W. is an investigator of the Howard Hughes Medical Institute (HHMI) and supported by NINDS R01NS032457. Y.Z., J.B., and P.J.P. are supported by R01-HG012573 and R01-CA269805. S.Z. is supported by the Program in Genetics and Genomics PhD Training Grant T32GM141745. X.Q. is supported by K99 NS135123 from the NINDS. A.J.K. is supported by HHMI Jane Coffin Childs Fellowship.

Competing interests. P.J.P. is a member of the scientific advisory board (SAB) for Bioskryb Genomics, Inc. C.A.W. is a member of the SAB of Bioskryb Genomics, Inc, (cash, equity), Mosaica Therapeutics (cash, equity), and an advisor to Maze Therapeutics (equity).

Data and materials availability. The Tn5-based single cell whole-genome sequencing data will be deposited on dbGap. BLISS BED files were downloaded from https://doi.org/10.6084/m9.figshare.18530531.v2. The ENCODE replication timing track of BG02ES was downloaded from https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=wgEncodeUwRepliSeq. The Gene Ontology Cellular Component terms (C5) and curated gene sets (C2) were downloaded from Molecular Signatures Database [69]. The house keeping gene list is part of C2 (named "HSIAO_HOUSEKEEPING_GENES"). The SFARI Autism gene list was downloaded from https://gene.sfari.org/. The epilepsy gene list was downloaded from https://https://omim.org/. The neurodevelopmental disorder gene list and schizophrenia gene list were downloaded from http://www.brainvar.org/. The cancer driver gene list was downloaded from https://cancer.sanger.ac.uk/census. The pLI gene list was downloaded from https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=gnomadPLI. The human postsynaptic gene list [70] was downloaded from https://static-content.springer.com/esm/art%3A10.1038%2Fnn.2719/MediaObjects/41593_2011_BFnn2719_MOESM28_ESM.xls. The postnatal reference for scRNA-seq [71] was downloaded from https://console.cloud.google.com/storage/browser/neuro-dev/Processed_data. The COSMIC pan-cancer signature panel (version 3.4) was downloaded from SigProfilerAssignment GitHub.

Code availability. We provide the analysis code as well as detailed instructions about environment set up and software installation are available as a zip file with our submission. We intend to release the analysis code as a open-source GitHub repository in the near future.

Author contribution. D.D.S, Y.Z., P.J.P, C.A.W. conceptualized, supervised, and designed this project. C.A.W. and P.J.P. provided resources for this project. D.D.S., J.B., T.T., F.T., and Y.Z. performed analysis. U.G., S.Z., A.J.K., X.Q., and D.D.S. performed experiments. D.D.S. and Y.Z. wrote the manuscript. All authors provided edits to the manuscript.

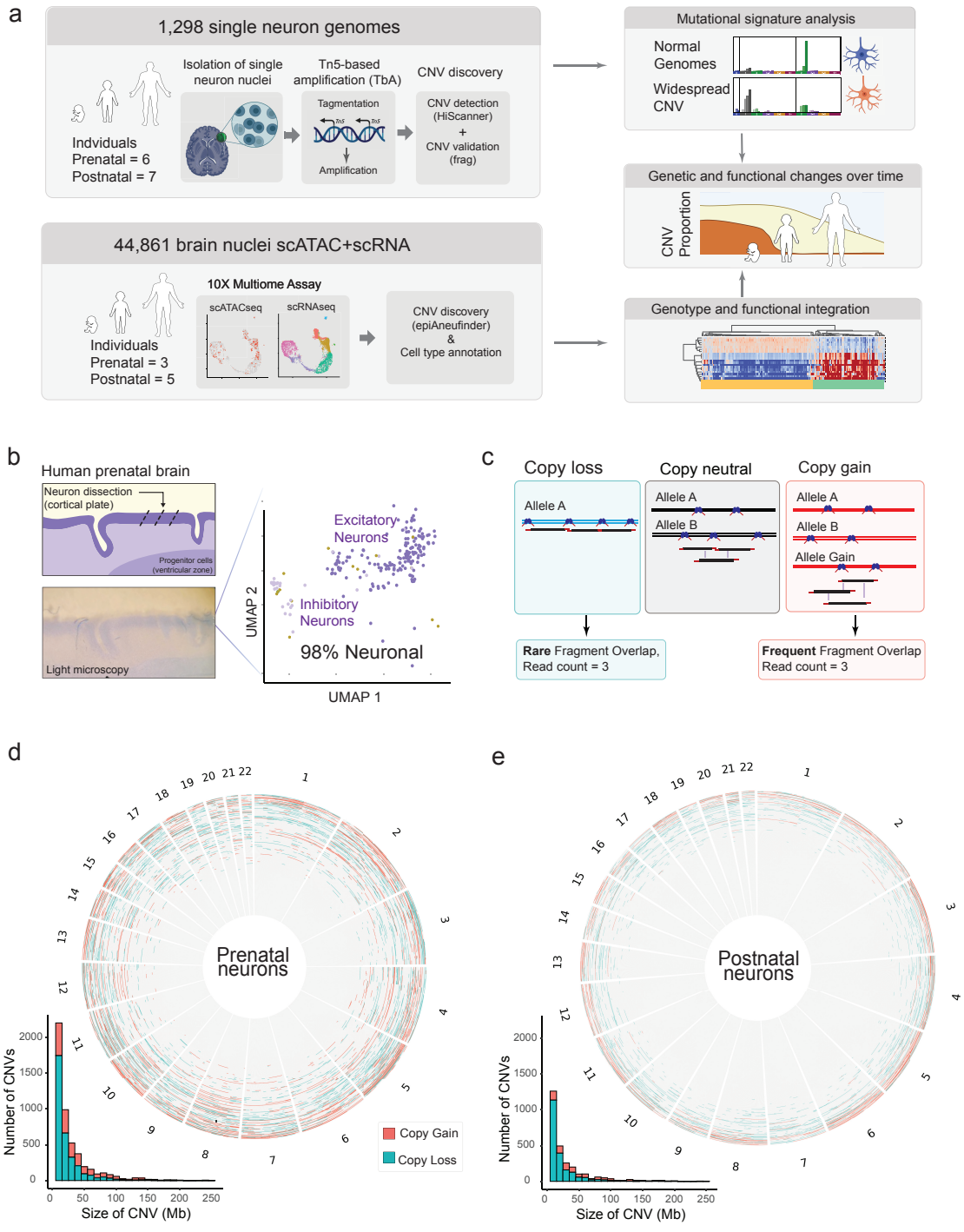


Fig. 1: A subpopulation of human postmitotic neurons harbors aberrant copy number patterns genome-wide. (Continued on next page)

Fig. 1: **a.** Overview of study design for scWGS, copy number calling, and downstream analyses. Created with BioRender.com. **b.** Cortical plate dissection of normal prenatal human brain isolates neuronal nuclei, confirmed by single-nucleus RNA sequencing. Neuronal nuclei (dark and light purple); non-neuronal nuclei (yellow). **c.** Schematic illustration of Tn5-based amplification libraries in which single-genomes create distinct patterns of molecular overlap that distinguishes copy number states. **d.** Circos plot of CNV profiles and gain/loss distribution in 459 high quality prenatal human cortical neurons. Chromosomes are arranged circumferentially. Neurons are ordered based on total CNV sizes. Bar plot indicates size and frequency of gains and losses. **e.** Circos plot of CNV profiles and gain/loss distribution in 625 high quality postnatal human cortical neurons. Chromosomes are arranged circumferentially. Neurons are ordered based on total CNV sizes. Bar plot indicates size and frequency of gains and losses.

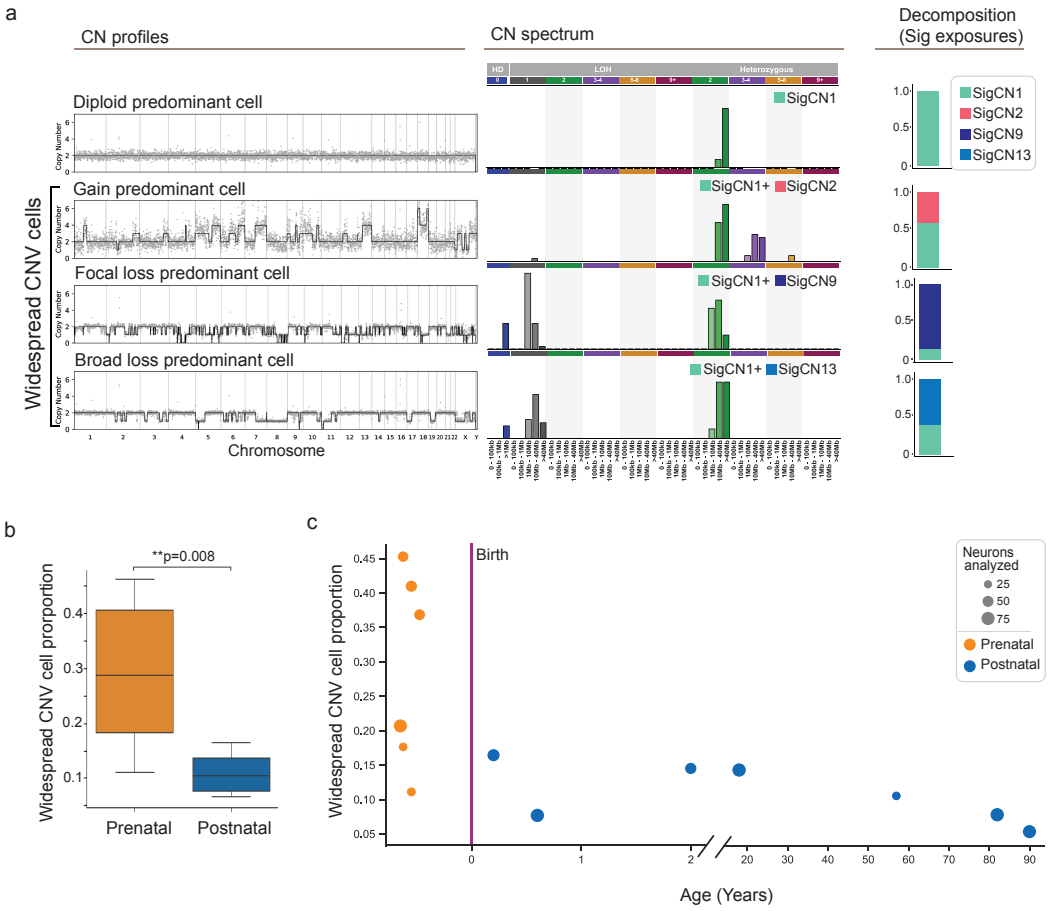


Fig. 2: **Widespread CNVs that predominate prenatally are associated with copy number signatures. a.** Copy number signatures represented by the four primary signature clusters identified across all human neurons. The left panel shows representative single-neuron copy number profiles for each of the signature groups, the middle panel shows the composite copy number spectrum characteristic of each group, and the bar plots in the right panel shows the signature exposure decomposition. **b.** Proportion of neurons with widespread CNVs observed in human prenatal versus postnatal neurons (n=6 prenatal and n=7 postnatal brains). Box plots indicate the median, first and third quartiles (hinges) and the most extreme data points no farther than 1.5X IQR from the hinge (whiskers). IQR, interquartile range. p, p-value, two-tailed Student's test. **c.** Frequency of neurons with widespread CNVs rapidly decreases during the perinatal period and remains low throughout life. Each point represents neurons analyzed for a single individual.

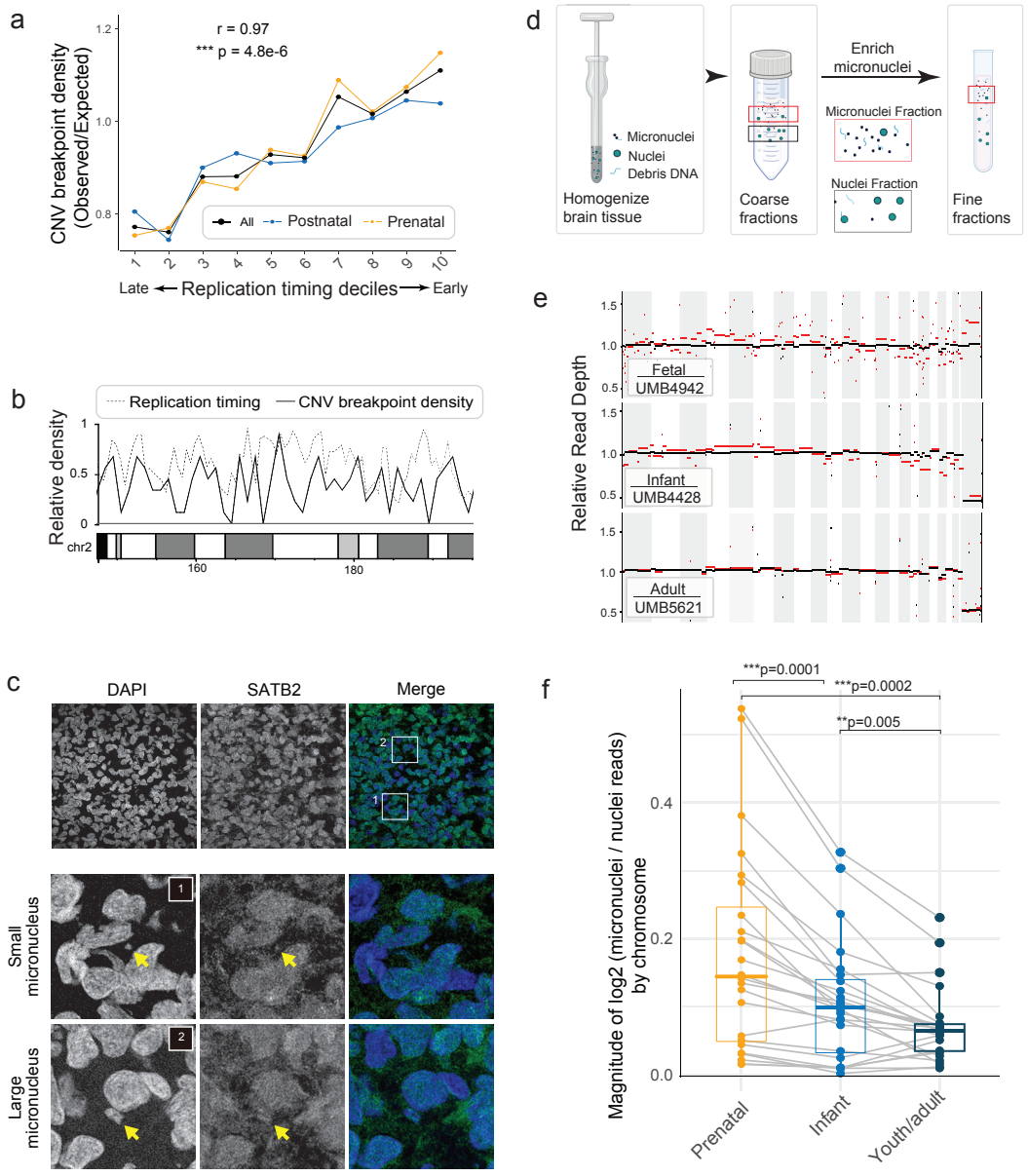


Fig. 3: **Etiology of CNV breakpoints. a.** Association between CNV breakpoints and replication timing deciles. **b.** Representative region on chromosome 2 showing local association between frequent regions of CNV breakpoints and replication timing. **c.** Micronuclei in the cortical plate are associated with neuronal nuclei, marked by SATB2 staining. **d.** Schematic of serial sucrose gradient fractionation to obtain micronuclear and nuclear fractions for sequencing. **e.** Relative depth across each chromosome after genome sequencing of micronuclear fraction (red) compared to nuclear fraction (black), showing that the DNA content of postnatal micronuclei increasingly resembles that of the nuclear fraction with increasing age. **f.** Micronuclear fraction of genomic content relative to paired nuclear fractions shows reduction between prenatal, infant, and youth/elderly cases. Each point represents one chromosome in a donor brain. Box plots indicate the median, first and third quartiles (hinges) and the most extreme data points no farther than 1.5X IQR from the hinge (whiskers).

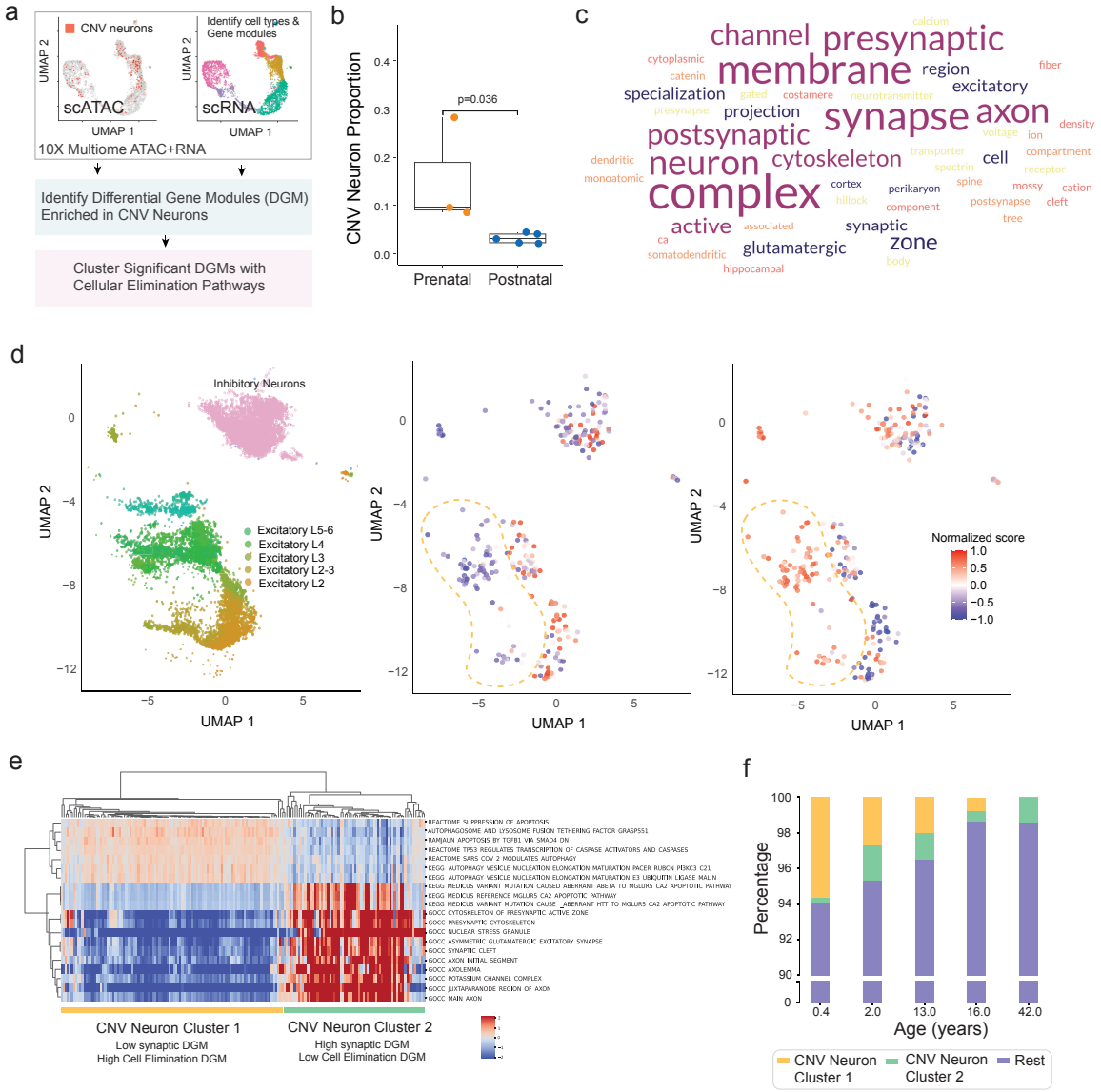


Fig. 4: Differential gene modules (DGMs) regulating the synapse are associated with CNV neurons. (Continued on next page)

Fig. 4: **a.** Schematic of snATAC/RNA-seq workflow to identify differential gene modules (DGMs) associated with CNV states. **b.** Proportion of CNV neurons based on snATAC data in prenatal (n=3) and postnatal (n=5) brain samples show sharp decrease after birth. Box plots indicate the median, first and third quartiles (hinges) and the most extreme data points no farther than 1.5X IQR from the hinge (whiskers). **c.** Word cloud of statistically significant gene ontology DGMs enriched in CNV neurons. **d.** UMAP visualization of snRNA-seq data from all postnatal brain cells. Left panel: excitatory and inhibitory neuron clusters. Right panels: expression of representative DGMs for synaptic function and cell death. Yellow dashed line highlights an extended cluster of excitatory neurons that show both decreased synapse gene module scores and increased cell death gene module scores (Cluster 1 neurons). **e.** Hierarchical clustering of CNV neurons based on significant DGM and cell elimination pathway expression patterns reveals two distinct clusters. Cluster 1 neurons (left) represents cells with decreased synaptic gene expression and high expression of cell death pathways, while Cluster 2 neurons (right) show the opposite pattern. **f.** Cluster 1 CNV neurons are more prevalent in younger individuals.

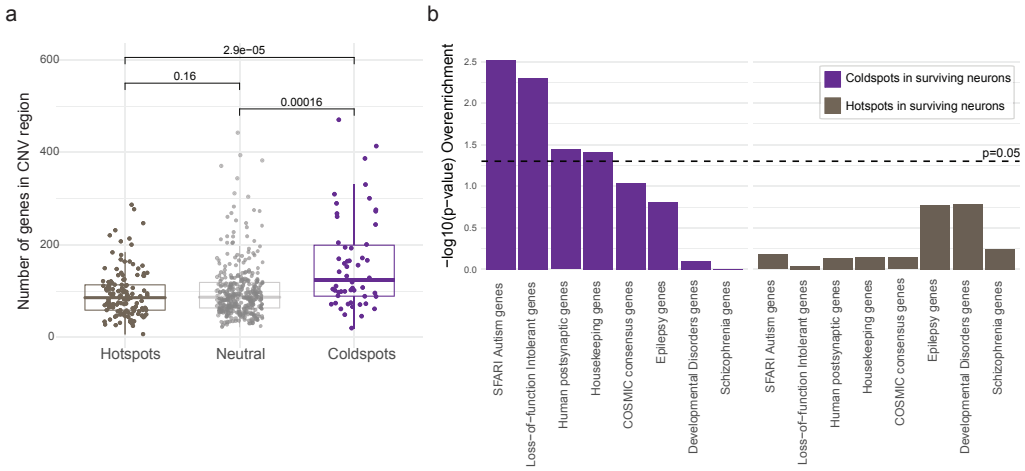


Fig. 5: **Gene-level selection of CNV. a.** Box plot showing the number of genes of surviving neurons in CNV hotspots, neutral, and coldspots across the genome. Each dot represents an individual genomic bin, with box plots showing median, interquartile range, and whiskers extending to 1.5X the interquartile range. **b.** Bar graph illustrating the enrichment p-value of specific gene categories in CNV coldspots in surviving neurons, suggesting that loss of specific gene categories may not be tolerated by the individual cell. The dashed line indicates a significance threshold at p=0.05.

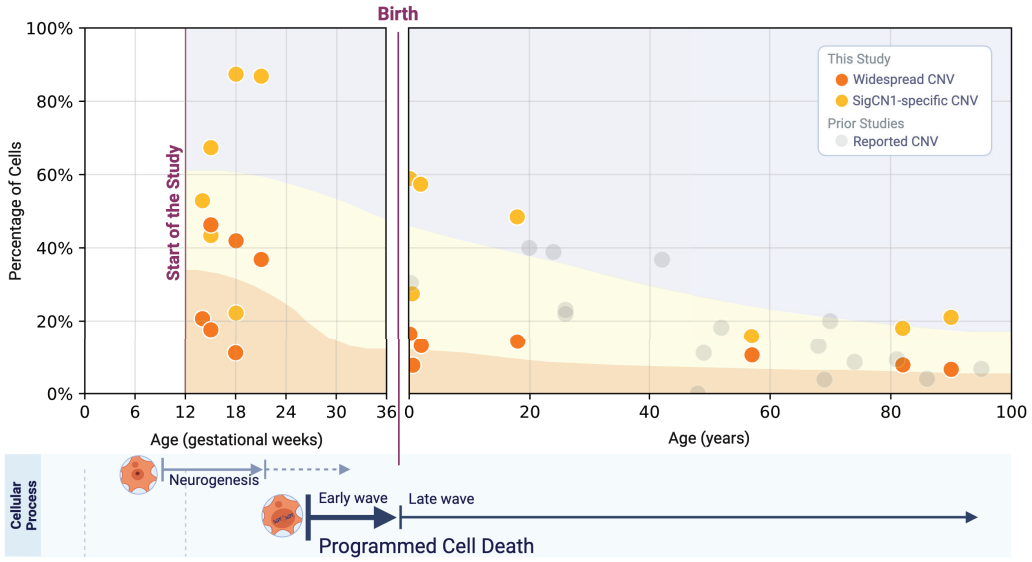
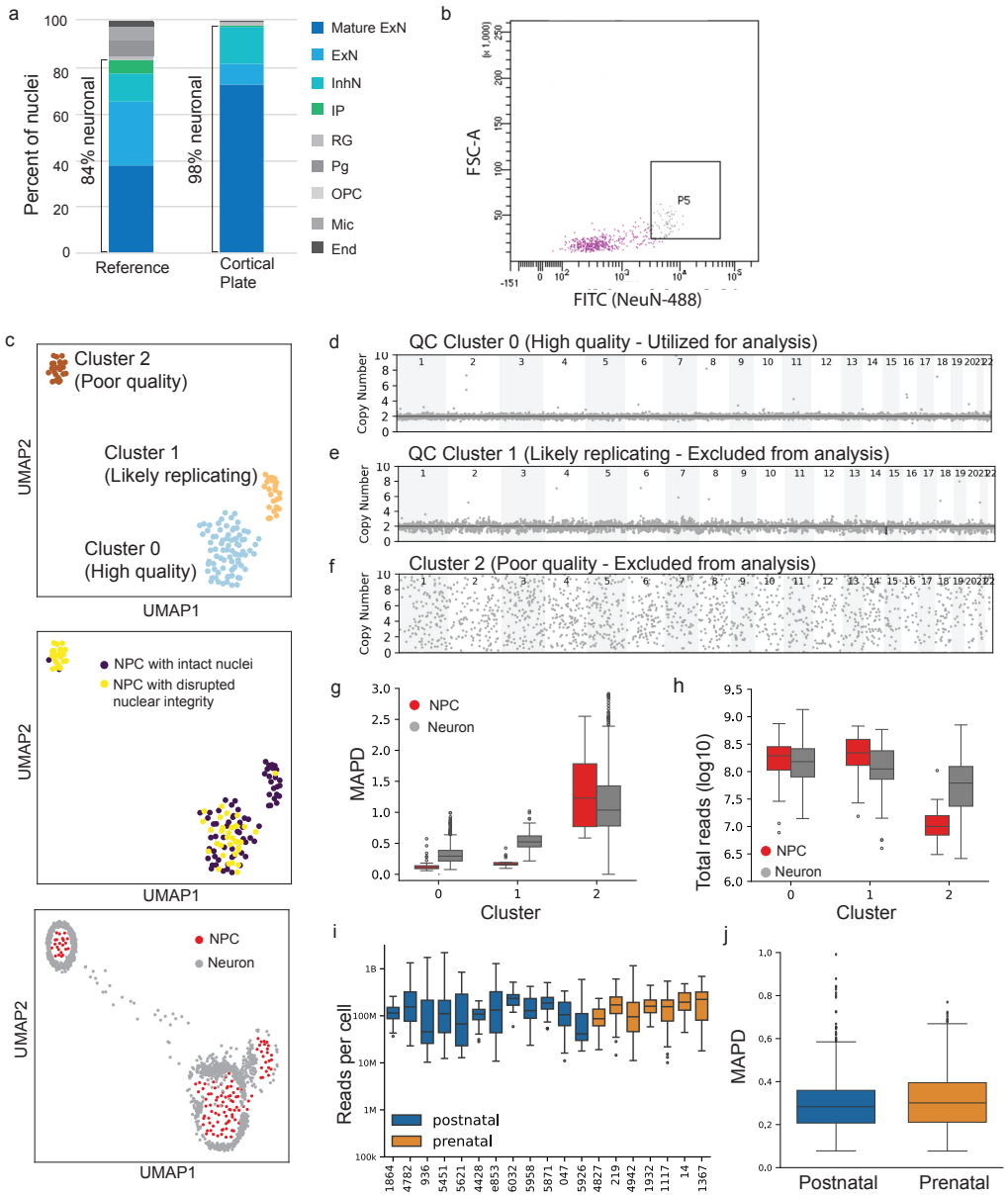
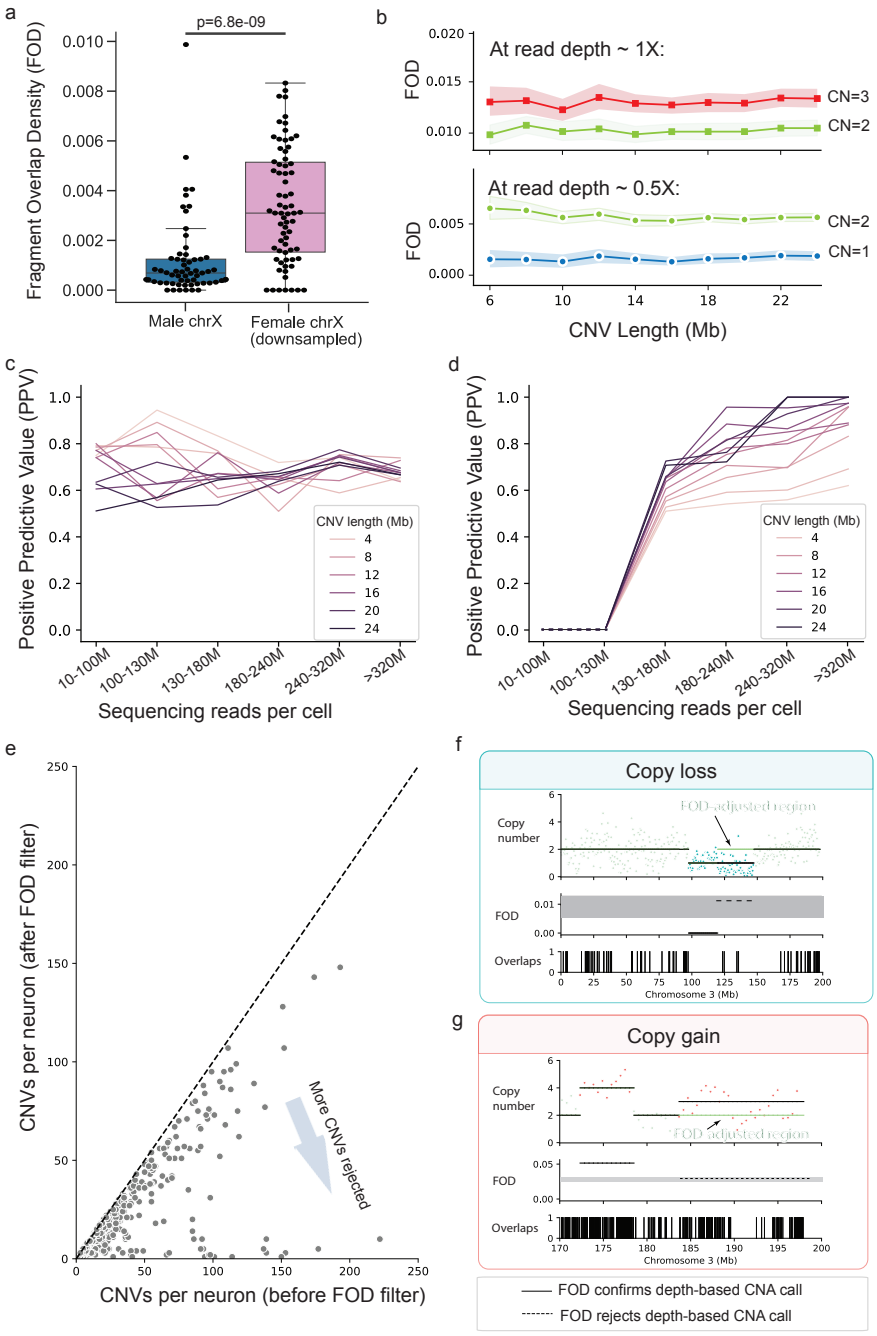


Fig. 6: **Schematic of findings from this study.** Summary of CNV abnormalities across all ages for our study and meta-analysis of prior studies [13]. Neurogenesis in the cerebral cortex begins roughly at 10-12 gestational weeks, continuing to about 24 gestational weeks for excitatory neurons of the cerebral cortex and potentially until after birth for inhibitory neurons of neocortex, and produces neurons with aberrant genomes. Estimated trends programmed cell death based on literature [22, 72] are displayed, which establishes a large wave of cell death (∼30%) near birth, a time of extensive neuronal connectivity and maturation, followed by ongoing elimination of neurons that do not form appropriate synaptic connections. Created with BioRender.com.

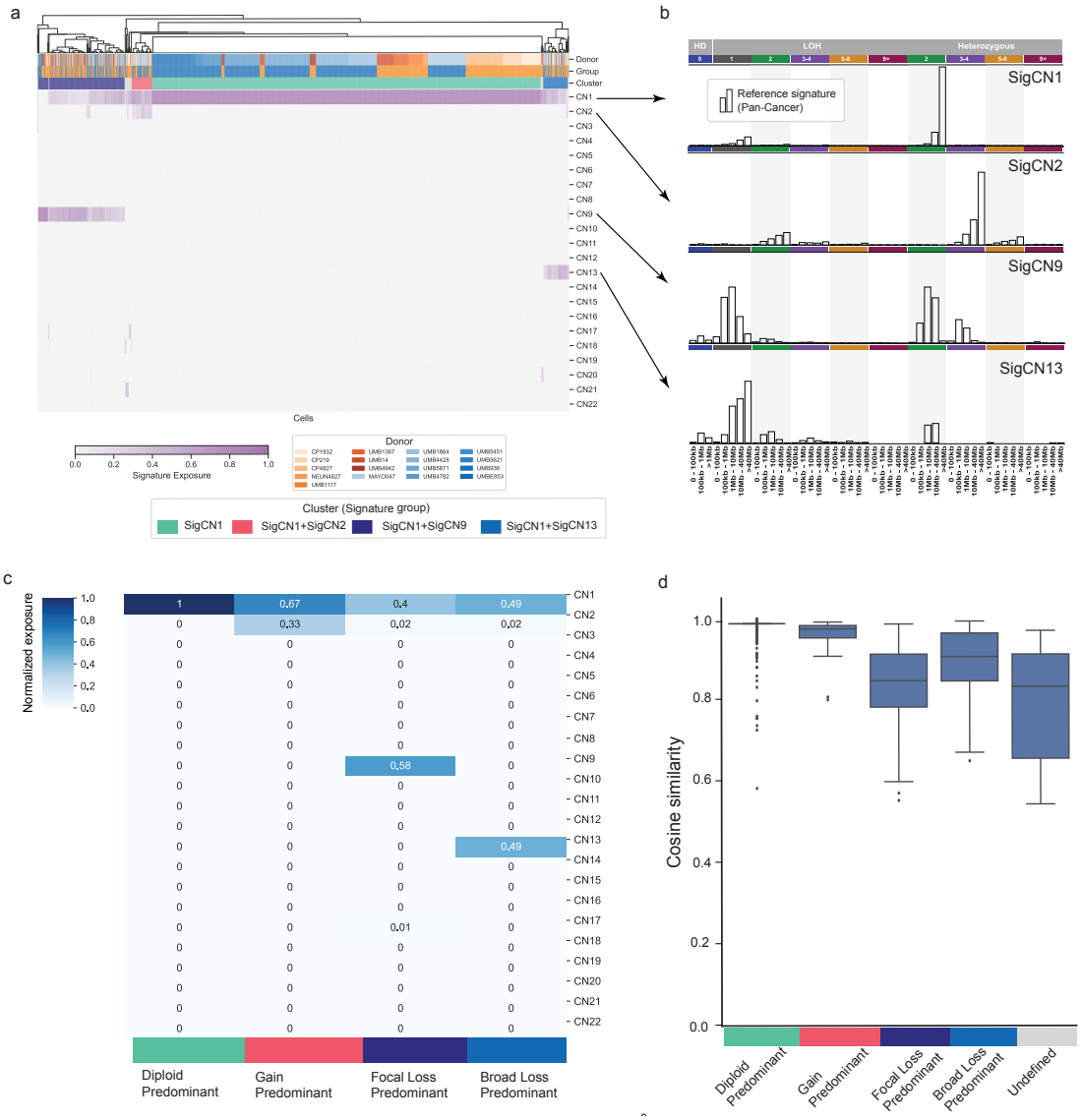
Extended Data Figures



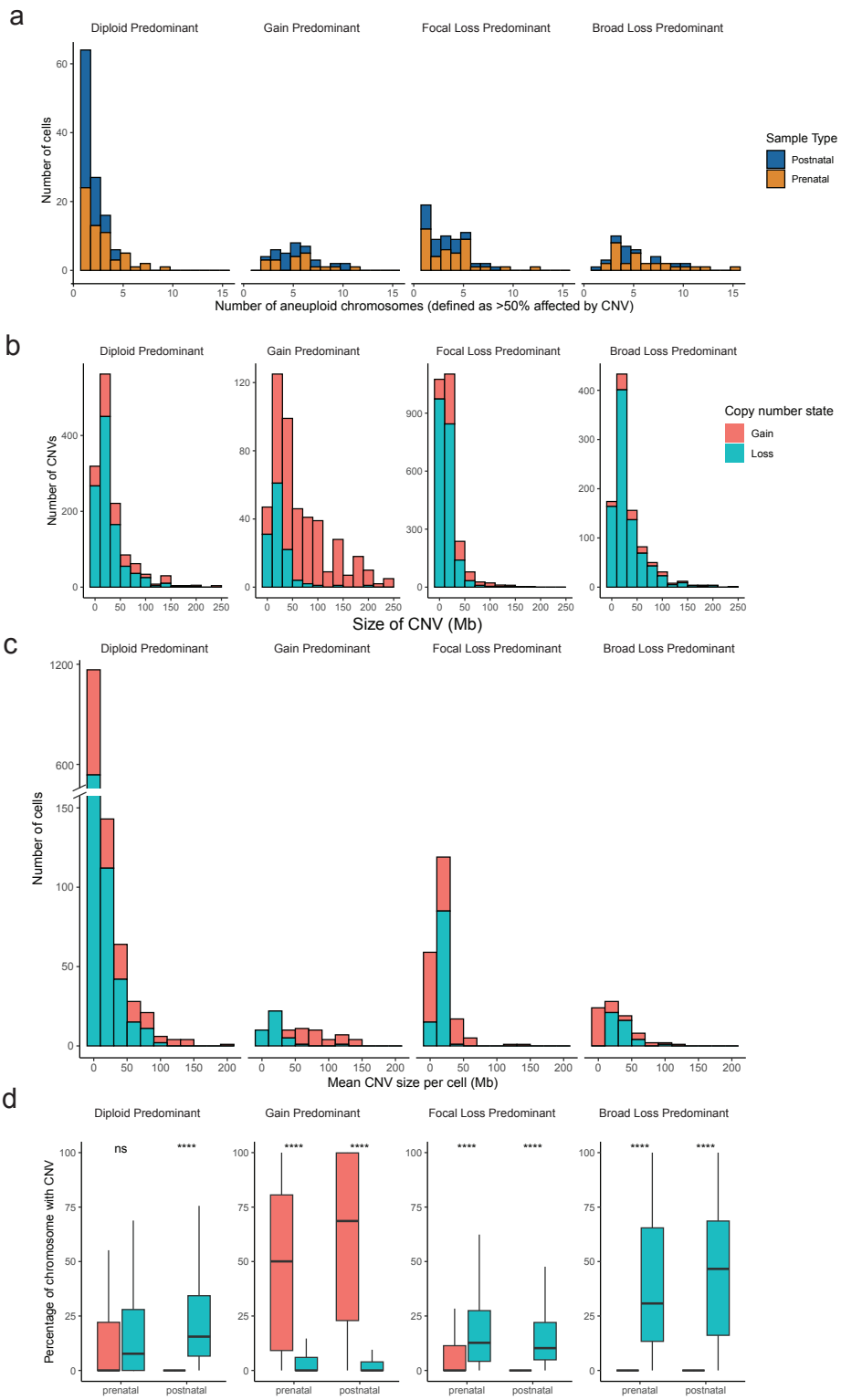
Extended Data Fig. 1 **a.** Nuclei isolation and confirmation of cell types from prenatal cortical plate dissection. Excitatory neuron (ExN), inhibitory neuron (InhN), intermediate progenitor (IP), radial glia (RG), neural progenitor (Pg), oligodendrocyte progenitor cell (OPC), microglia (Mic), endothelial cell (End). **b.** Single-cell isolation for postnatal neuronal nuclei ($NeuN^{High}$) after gating for 2N nuclei. This strategy has previously been shown to result in a nearly pure neuronal population [33]. **c.** Dimensionality reduction of binned read coverage across single-cell genomes of neural progenitor cell line (NPC) results in distinct clusters that identify high quality genomes. From left to right, UMAP colored by: (top) leiden clustering, (middle) NPC nuclei state inferred from staining, and (bottom) cell type (NPC or neuron) after label transfer. **d-f.** Representative copy number profile for QC Clusters 0-2. Only QC Cluster 0 cells were used for subsequent analyses. **g-h.** Coverage uniformity (measured by MAPD) and total reads for NPC and single neuron from human donors. MAPD: Median of the Absolute values of all Pairwise Differences. **i.** Total number of reads per single neuron genome across human donors. **j.** MAPD of single neuron genomes from pre- and postnatal human donors.



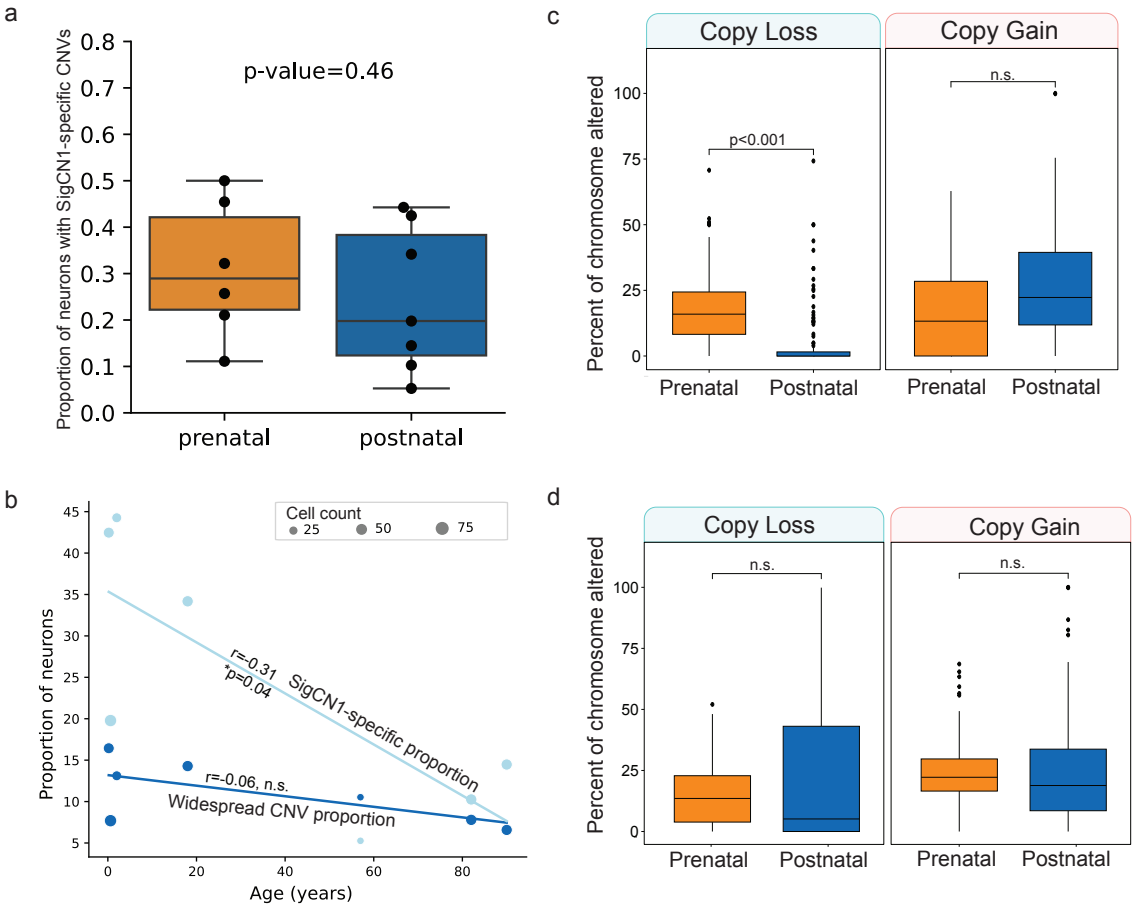
Extended Data Fig. 2 **a.** Assessment of Fragment Overlap Density (FOD) to distinguish copy number states at diploid (female X) and haploid (male X) chromosomes. **b.** Assessment of FOD to distinguish between single copy loss (CN=1), normal (CN=2), and single copy gain (CN=3) states using simulated CNVs by mixing defined copies of male haploid X chromosomes. Plots are separated into high and low read depth (1X and 0.5X respectively) simulations such that FOD determination of copy number cannot rely on read depth in these scenarios. **c-d.** Evaluation of FOD filter using positive predictive value in simulated single copy gains (**c**) and single copy losses (**d**) at varying sequencing depths and CNV lengths. **e.** Distribution of CNV calls per neuron in the single neuron data before and after FOD filter, demonstrating substantial filtering of likely false positives. **f-g.** Examples of copy loss (**f**) and copy gain (**g**) from a single cell, showing copy number, FOD values, and fragment overlaps. Solid lines indicate FOD-confirmed CNV calls while dashed lines indicate FOD-rejected calls.



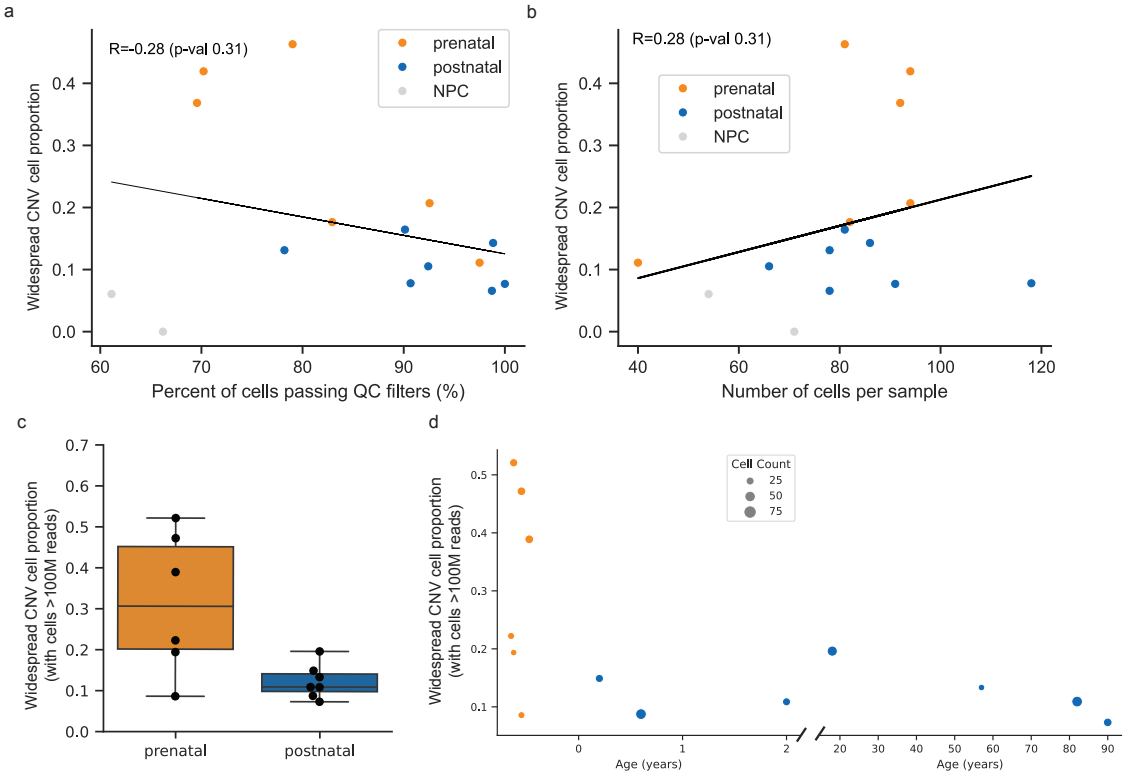
Extended Data Fig. 3 **a.** Heatmap of normalized exposures (i.e., relative contribution) of each COSMIC copy number signature to each neuron. Clusters are defined by hierarchical clustering of normalized exposures. **b.** Reference COSMIC copy number signatures that are most prevalent in human neurons. **c.** Mean normalized exposure of composite (e.g., SigCN1+SigCN9) COSMIC copy number signatures. **d.** Cosine similarity between composite COSMIC copy number signatures and neuronal CNV profiles.



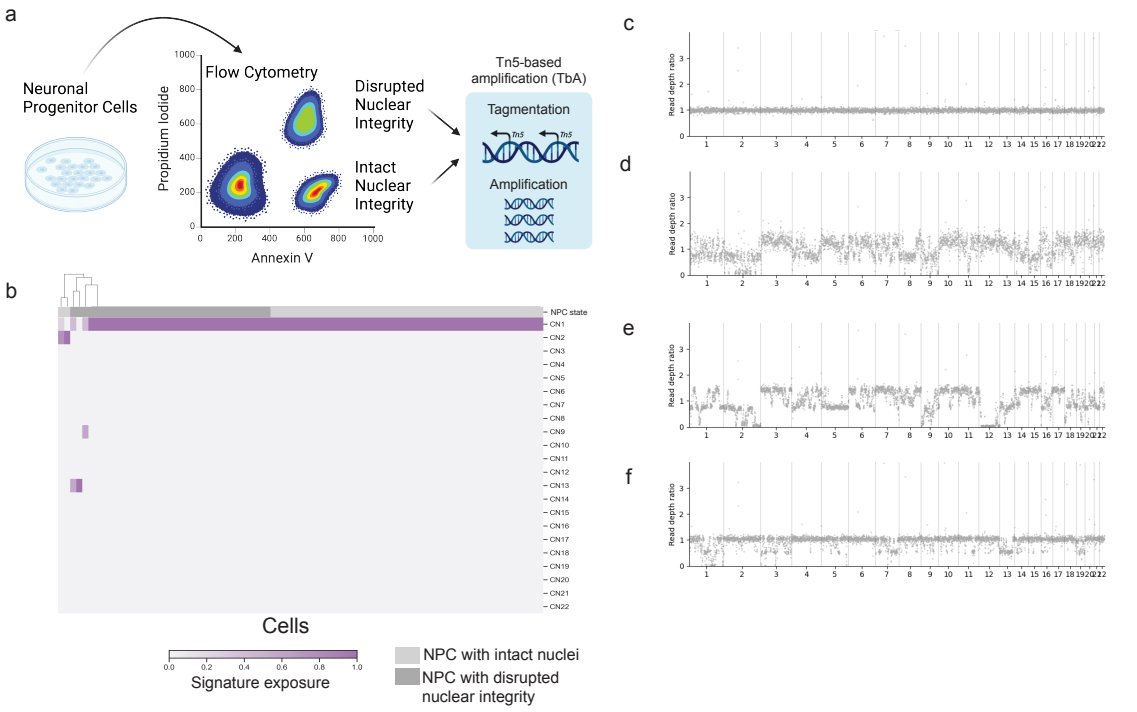
Extended Data Fig. 4 **a.** Distribution of aneuploid chromosomes in each signature cluster. **b.** Size distribution of CNVs in each signature cluster, accounting for cells that have at least 1 CNV. **c.** Average size distribution of CNVs in each cell. **d.** Average percentage of chromosome altered per cell, accounting for cells that have at least 1 CNV.



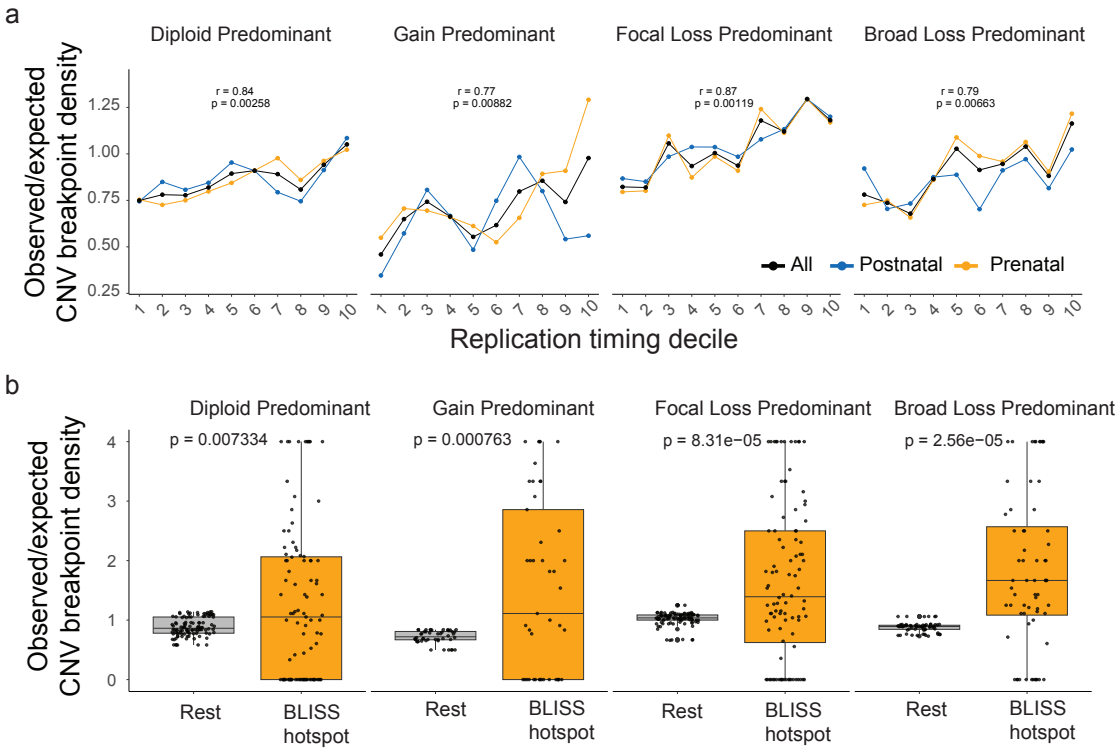
Extended Data Fig. 5 **a.** Proportion of neurons with SigCN1-specific CNVs observed in human prenatal versus postnatal neurons. Box plots indicate the median, first and third quartiles (hinges) and the most extreme data points no farther than 1.5X IQR from the hinge (whiskers); n=6 prenatal and n=7 postnatal brains. **b.** Trends of widespread or SigCN1-specific CNV neuron proportion during aging (n=7 postnatal brains). **c-d.** Average CNV size (expressed as percent of chromosome) for neurons with SigCN1-specific CNVs (**c**) and neurons with widespread CNVs (**d**).



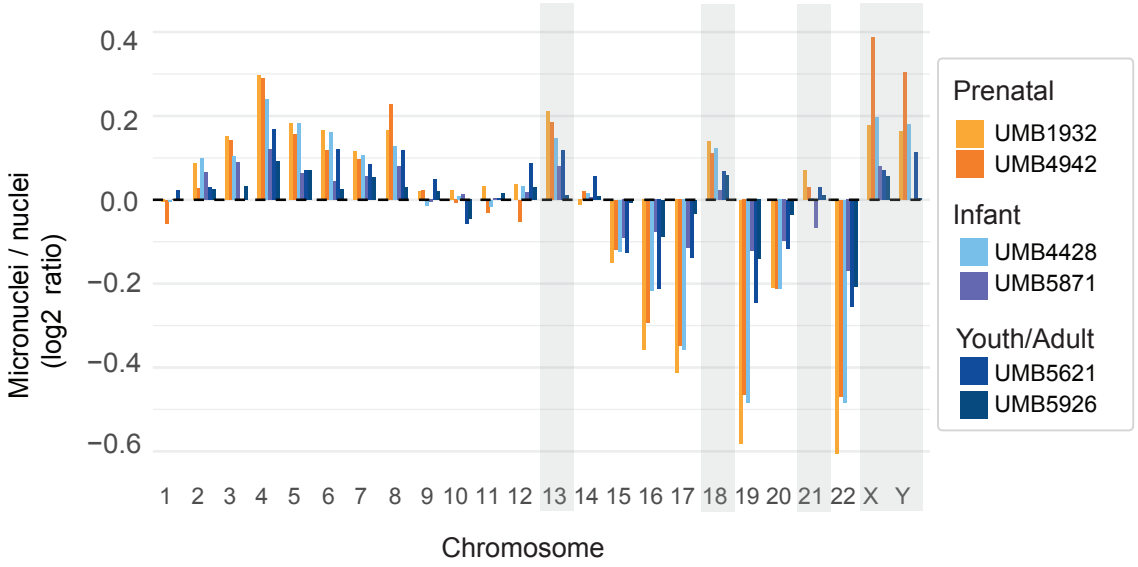
Extended Data Fig. 6 **a.** No significant correlation was found between sample quality (as defined by proportion of cells passing QC) and widespread CNV cell proportion. **b.** No significant correlation was found between size of final sample that passes QC and widespread CNV proportion. **c.** Widespread CNV neuron proportion when analyzing only high depth cells (i.e., cells with >100M reads). **d.** Widespread CNV neuron proportion trend with age when analyzing only high depth cells.



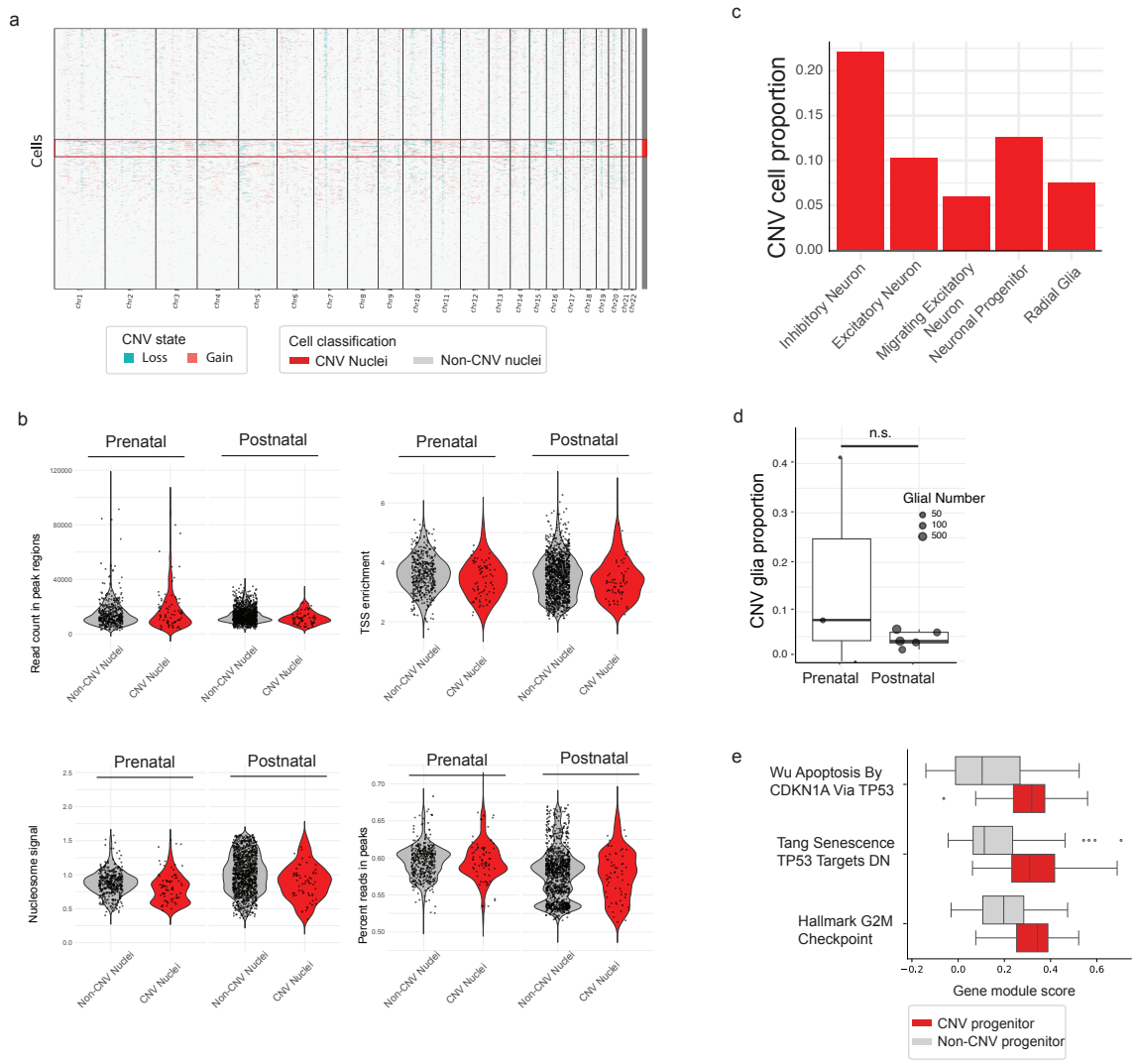
Extended Data Fig. 7 **a.** Schematic of sorting neural progenitor cells for those with disrupted vs intact nuclear integrity for TbA. To form an appropriate comparison to postmortem tissue, live cells were excluded. **b.** Signature decomposition of single cell copy number profiles from neural progenitor cells indicates their copy number patterns represent similar patterns as those in the human brain. **c-f.** Representative genome-wide copy number profiles from neural progenitor cells showing the four signature clusters: **c.** SigCN1 (*Diploid Predominant*), **d.** SigCN1+SigCN2 (*Gain Predominant*), **e.** SigCN1+SigCN13 (*Broad Loss Predominant*), and **f.** SigCN1+SigCN9 (*Focal Loss Predominant*).



Extended Data Fig. 8 a. Association between CNV breakpoints and replication timing for each copy number signature. **b.** Association between CNV breakpoints and physiologic double stranded breaks (BLISS hotspots) for each copy number signature.



Extended Data Fig. 9 Log2 ratio of micronuclei fraction to nuclei fraction read counts by chromosome. Grey shading highlights chromosomes 13, 18, 21, X, Y which deviate from the expected trend.



Extended Data Fig. 10 **a.** epiAneufinder CNV profiles from scATAC-seq data of a representative prenatal brain sample (29180_F4), showing chromosomal gains (red) and losses (teal). The horizontal red lines indicate a cluster of CNV neurons, cells with the highest CNV burden. **b.** Quality control characteristics of the CNV neurons are indicative of high quality cells, similar to the rest of the population. **c.** CNV cell proportions across different cell types, demonstrating CNV presence across various stages of neuronal maturity. **d** Box plot of CNV glia proportion in pre- and postnatal brains. **e.** Box plots of DGM score distribution in CNV progenitor cells in prenatal samples at midgestation. Box plots indicate the median, first and third quartiles (hinges) and the most extreme data points no farther than 1.5X IQR from the hinge (whiskers).

Tables

Donor ID	Age group	Sex	Age	Cells	Cells passed QC	Pass Rate (%)
4827	prenatal	F	14wk	94	87	92.6
1117	prenatal	M	15wk	82	68	82.9
1367	prenatal	M	15wk	81	64	79.0
219	prenatal	F	18wk	94	66	70.2
14	prenatal	F	18wk	40	39	97.5
1932	prenatal	M	21wk	92	64	69.6
4942	prenatal	F	25wk	75	71	94.7
4428	postnatal	M	0.2y	81	73	90.1
5871	postnatal	M	0.6y	91	91	100.0
1864	postnatal	F	2y	78	61	78.2
6032	postnatal	M	4y	17	14	82.4
4782	postnatal	M	18y	86	85	98.8
5926	postnatal	M	21y	12	11	91.7
5958	postnatal	M	22y	23	16	69.6
5621	postnatal	F	37y	34	29	85.3
5451	postnatal	M	57y	66	61	92.4
E853	postnatal	F	82y	118	107	90.7
047	postnatal	F	90y	78	77	98.7
Intact nuclei	NPC			71	47	66.2
Disrupted nuclei	NPC			54	33	61.1

Table 1: Sample table of all prenatal, postnatal, and cell line TbA libraries used in this study. Pass rate refers to percent of cells passing all quality controls (see methods). NPC: neural progenitor cell line.

	Micronuclei (manual)	Main nuclei (Cellpose)	Main nuclei (manual)
Image 1	26	265	251
Image 2	24	374	324
Image 3	15	457	361
Image 4	11	465	416
Image 5	27	443	349
Average	20.6	400.8	340.2

Table 2: Micronuclei and primary nuclei enumeration in the cortical plate. Enumerated across 5 different images.

Donor ID	Micronuclei mean coverage	Nuclei mean coverage	Micronuclei mean depth	Nuclei mean depth
1932	67.25	70.40	1.85	2.01
4428	62.69	80.88	1.58	3.19
4942	47.67	62.16	1.43	2.16
5621	34.95	69.42	0.93	2.97
5926	46.02	63.44	1.36	2.17
5871	17.24	59.65	0.43	2.04

Table 3: Sequencing statistics for paired micronuclei and nuclei isolated by sucrose gradient fractionation.

References

[1] Maury, E. A. *et al.* Somatic mosaicism in schizophrenia brains reveals prenatal mutational processes. *Science* **386**, 217–224 (2024).

[2] McConnell, M. J. *et al.* Intersection of diverse neuronal genomes and neuropsychiatric disease: The brain somatic mosaicism network. *Science* **356** (2017).

[3] 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).

[4] Bolton, H. *et al.* Mouse model of chromosome mosaicism reveals lineage-specific depletion of aneuploid cells and normal developmental potential. *Nat. Commun.* **7**, 11165 (2016).

[5] Currie, C. E. *et al.* The first mitotic division of human embryos is highly error prone. *Nat. Commun.* **13**, 6755 (2022).

[6] Starostik, M. R., Sosina, O. A. & McCoy, R. C. Single-cell analysis of human embryos reveals diverse patterns of aneuploidy and mosaicism. *Genome Res.* **30**, 814–825 (2020).

[7] Palmerola, K. L. *et al.* Replication stress impairs chromosome segregation and preimplantation development in human embryos. *Cell* **185**, 2988–3007.e20 (2022).

[8] van Echten-Arends, J. *et al.* Chromosomal mosaicism in human preimplantation embryos: a systematic review. *Hum. Reprod. Update* **17**, 620–627 (2011).

[9] Vanneste, E. *et al.* Chromosome instability is common in human cleavage-stage embryos. *Nat. Med.* **15**, 577–583 (2009).

[10] Rehen, S. K. *et al.* Constitutional aneuploidy in the normal human brain. *J. Neurosci.* **25**, 2176–2180 (2005).

[11] Sekar, S. *et al.* Complex mosaic structural variations in human fetal brains. *Genome Res.* **30**, 1695–1704 (2020).

[12] Rehen, S. K. *et al.* Chromosomal variation in neurons of the developing and adult mammalian nervous system. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13361–13366 (2001).

[13] Chronister, W. D. *et al.* Neurons with complex karyotypes are rare in aged human neocortex. *Cell Rep.* **26**, 825–835.e7 (2019).

[14] McConnell, M. J. *et al.* Mosaic copy number variation in human neurons. *Science* **342**, 632–637 (2013).

[15] Sun, C. *et al.* Mapping recurrent mosaic copy number variation in human neurons. *Nat. Commun.* **15**, 4220 (2024).

[16] Abrahams, B. S. *et al.* SFARI gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).

[17] Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

[18] Shohat, S., Ben-David, E. & Shifman, S. Varying intolerance of gene pathways to mutational classes explain genetic convergence across neuropsychiatric disorders. *Cell Rep.* **18**, 2217–2227 (2017).

[19] Frankle, W. G., Lerma, J. & Laruelle, M. The synaptic hypothesis of schizophrenia. *Neuron* **39**, 205–216 (2003).

[20] Spooren, W., Lindemann, L., Ghosh, A. & Santarelli, L. Synapse dysfunction in autism: a molecular medicine approach to drug discovery in neurodevelopmental disorders. *Trends Pharmacol. Sci.* **33**, 669–684 (2012).

[21] Blaschke, A. J., Staley, K. & Chun, J. Widespread programmed cell death in proliferative and postmitotic regions of the fetal cerebral cortex. *Development* **122**, 1165–1174 (1996).

[22] Wong, F. K. & Marín, O. Developmental cell death in the cerebral cortex. *Annual review of cell and developmental biology* **35**, 523–542 (2019).

[23] Yang, A. H. *et al.* Chromosome segregation defects contribute to aneuploidy in normal neural progenitor cells. *J. Neurosci.* **23**, 10454–10462 (2003).

[24] Peterson, S. E. *et al.* Aneuploid cells are differentially susceptible to caspase-mediated death during embryonic cerebral cortical development. *J. Neurosci.* **32**, 16213–16222 (2012).

[25] Voiculescu, B., Nat, R., Lin, E. & Iosef, C. Apoptosis in human embryo development: 1. cerebral cortex. *J. Cell. Mol. Med.* **4**, 284–288 (2000).

[26] Nikolić, M., Gardner, H. A. R. & Tucker, K. L. Postnatal neuronal apoptosis in the cerebral cortex: physiological and pathophysiological mechanisms. *Neuroscience* **254**, 369–378 (2013).

[27] Huttenlocher, P. R. Synaptic density in human frontal cortex - developmental changes and effects of aging. *Brain Res.* **163**, 195–205 (1979).

[28] Huttenlocher, P. R. & Dabholkar, A. S. Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178 (1997).

[29] Von Gudden, B. *Bernhard von Gudden's gesammelte und hinterlassene Abhandlungen* (1889).

[30] Wall, R. V., Basavarajappa, D., Klistoner, A., Graham, S. & You, Y. Mechanisms of transsynaptic degeneration in the aging brain. *Aging Dis.* **15**, 2149–2167 (2024).

[31] Finlay, B. L. Cell death and the creation of regional differences in neuronal numbers. *J. Neurobiol.* **23**, 1159–1171 (1992).

[32] Finlay, B. L. & Slattery, M. Local differences in the amount of early cell death in neocortex predict adult local specializations. *Science* **219**, 1349–1351 (1983).

[33] Miller, M. B. *et al.* Somatic genomic changes in single alzheimer's disease neurons. *Nature* **604**, 714–722 (2022).

[34] Zhao, Y. *et al.* High-resolution detection of copy number alterations in single cells with HiScanner. *bioRxiv* 2024.04.26.587806 (2024).

[35] Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).

[36] Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16910–16915 (2010).

[37] Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* **24**, 52–60 (2014).

[38] Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**, 984–991 (2022).

[39] De, S. & Michor, F. DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. *Nat. Biotechnol.* **29**, 1103–1108 (2011).

[40] Lu, J. *et al.* The distribution of genomic variations in human iPSCs is related to replication-timing reorganization during reprogramming. *Cell Rep.* **7**, 70–78 (2014).

[41] Ballarino, R. *et al.* An atlas of endogenous DNA double-strand breaks arising during human neural cell fate determination. *Sci Data* **9**, 400 (2022).

[42] Dileep, V. *et al.* Neuronal DNA double-strand breaks lead to genome structural variations and 3D genome disruption in neurodegeneration. *Cell* **186**, 4404–4421.e20 (2023).

[43] Adams, D. J. *et al.* Genetic determinants of micronucleus formation in vivo. *Nature* **627**, 130–136 (2024).

[44] Klaasen, S. J. *et al.* Nuclear chromosome locations dictate segregation error frequencies. *Nature* **607**, 604–609 (2022).

[45] Yano, S. *et al.* Propagation of neuronal micronuclei regulates microglial states. *bioRxiv* 2023.07.31.551211 (2023).

[46] Shimizu, N., Kanda, T. & Wahl, G. M. Selective capture of acentric fragments by micronuclei provides a rapid method for purifying extrachromosomally amplified DNA. *Nat. Genet.* **12**, 65–71 (1996).

[47] Tovini, L. & McClelland, S. E. Impaired CENP-E function renders large chromosomes more vulnerable to congression failure. *Biomolecules* **9** (2019).

[48] Nicolaides, K. H., Syngelaki, A., Ashoor, G., Birdir, C. & Touzet, G. Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. *Am. J. Obstet. Gynecol.* **207**, 374.e1–6 (2012).

[49] Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).

[50] Ramakrishnan, A. *et al.* epiAneufinder identifies copy number alterations from single-cell ATAC-seq data. *Nat. Commun.* **14**, 5846 (2023).

[51] Goodman, C. S. & Shatz, C. J. Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell* **72**, 77–98 (1993).

[52] Huang, Z. J. *et al.* BDNF regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell* **98**, 739–755 (1999).

[53] Pfisterer, U. & Khodosevich, K. Neuronal survival in the brain: neuron type-specific mechanisms. *Cell Death & Disease 2017 8:3* **8**, e2643–e2643 (2017).

[54] Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. Omim. org: Online mendelian inheritance in man (omim®), an online catalog of human genes and genetic disorders. *Nucleic acids research* **43**, D789–D798 (2015).

[55] Werling, D. M. *et al.* Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex. *Cell Rep.* **31**, 107489 (2020).

[56] Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210 (2019).

[57] Sondka, Z. *et al.* The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).

[58] Faust, T. E., Gunner, G. & Schafer, D. P. Mechanisms governing activity-dependent synaptic pruning in the developing mammalian CNS. *Nat. Rev. Neurosci.* **22**, 657–673 (2021).

[59] Becker, E. B. E. & Bonni, A. Cell cycle regulation of neuronal apoptosis in development and disease. *Prog. Neurobiol.* **72**, 1–25 (2004).

[60] Burek, M. J. & Oppenheim, R. W. Programmed cell death in the developing nervous system. *Brain Pathol.* **6**, 427–446 (1996).

[61] Kristiansen, M. & Ham, J. Programmed cell death during neuronal development: the sympathetic neuron model. *Cell Death Differ.* **21**, 1025–1035 (2014).

[62] Southwell, D. G. *et al.* Intrinsically determined cell death of developing cortical interneurons. *Nature* **491**, 109–113 (2012).

[63] Luquette, L. J. *et al.* Single-cell genome sequencing of human neurons identifies somatic point mutation and indel enrichment in regulatory elements. *Nat. Genet.* **54**, 1564–1571 (2022).

[64] Rohrback, S. *et al.* Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **115**,

520 10804–10809 (2018).

521 [65] Abascal, F. *et al.* Somatic mutation landscapes at single-molecule resolution. *Nature* **593**,
522 405–410 (2021).

523 [66] Dong, G. *et al.* Diverse somatic genomic alterations in single neurons in chronic traumatic
524 encephalopathy. *bioRxiv* 2025–03 (2025).

525 [67] Lodato, M. A. *et al.* Aging and neurodegeneration are associated with increased mutations in
526 single human neurons. *Science* **359**, 555–559 (2018).

527 [68] Xing, D., Tan, L., Chang, C.-H., Li, H. & Xie, X. S. Accurate SNV detection in single cells
528 by transposon-based whole-genome amplification of complementary strands. *Proc. Natl. Acad.*
529 *Sci. U. S. A.* **118** (2021).

530 [69] Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for inter-
531 preting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550
532 (2005).

533 [70] Bayés, À. *et al.* Characterization of the proteome, diseases and evolution of the human
534 postsynaptic density. *Nature neuroscience* **14**, 19–21 (2011).

535 [71] Herring, C. A. *et al.* Human prefrontal cortex gene regulatory dynamics from gestation to
536 adulthood at single-cell resolution. *Cell* **185**, 4428–4447.e28 (2022).

537 [72] Andersen, S. L. Trajectories of brain development: point of vulnerability or window of
538 opportunity? *Neuroscience & Biobehavioral Reviews* **27**, 3–18 (2003).

Methods

Isolation of neuronal nuclei

Human prenatal brain tissue was obtained from the NIH Neurobiobank from the frontoparietal lobes sectioned into 40μm sections with cryostat. Cortical plate was dissected using a microscalpel under a light microscope and dissociated into nuclei using gentle dounce homogenization in 1.5mL lysis buffer (10mM Tris-Hcl pH 8.0, 250mM sucrose, 25mM KCl, 5mM MgCl2, 0.1% Triton-X, 1X Compleat Mini Protease, 1mM DTT). Nuclei were pelleted by centrifugation for 10 minutes at 900G. Supernatant was removed and nuclei were resuspended in sorting buffer (1X PBS pH7.4, 0.8% BSA, 0.1μg/mL DAPI). Flow cytometry was performed, gating for FSC, SSC, and DAPI content, to isolate single nuclei. For postnatal brain nuclei, staining was performed with both DAPI and Alexa Fluor 488 conjugated Anti-NeuN at 1:1000 for 1 hour at 4°C (Sigma Aldrich MAB377X) as previously described [1].

Single nuclei whole genome library preparation

Single nuclei were lysed in microwells in 2μl lysis buffer (20mM NaCl, 20mM Tris-Hcl pH 8.0, 0.15% Triton-X, 1mM EDTA, 250mM DTT, 0.05ul Thermolabile Proteinase K (NEB P8111S)) at 30°C for 1 hour and proteinase inactivation at 55°C for 10 minutes. This step digests nucleosomes and improves accessibility of the genome. Tn5 was loaded per manufacturer instructions (Diagenode C01070010-20), and 1μl of Tn5 diluted 1000-fold was used per single nuclei. Tagmentation was performed using manufacturer buffers and instructions (Diagenode). Single-cell genome libraries were generated by 16 cycles of PCR using NexteraXT v2 indexes. Rows of 12 single cells were pooled for cleanup using Zymo Clean and Concentrator Kit. Each row of single cells was eluted in 42μl Zymo elution buffer for subsequent Ampure XP size selection at 0.8X. Four rows (48 cells) were pooled for sequencing in equimolar concentration, and sequencing was performed on Novaseq6000 (Novogene) for a target library size of 0.5Gb per single cell library. Illumina reads were aligned to the human reference with decoy sequence GRCh37d5 (hs37d5) using bwa-mem [2].

Single-cell library quality assessment

To ensure analysis of only high quality single-cell genomes, we first removed cells with <10M reads, and then used cell line data as the basis for transfer learning in order to distinguish high-quality from poor-quality sequencing profiles. The TbA libraries from single cells of a neuronal progenitor cell line (NPC) allowed us to establish a benchmark of sample quality. It is a labeled dataset of high-quality, recently viable cells as well as cells with compromised nuclear integrity through staurosporin treatment. To categorize cells computationally, we first normalized read counts for each cell in the NPC dataset using HiScanner [3]'s read-depth normalization function with bin size set to 500 kb, resulting in a cell-by-bin matrix. Next, we used scanpy [4] to scale the read counts, standardizing variance across cells, followed by principal component analysis (PCA) for dimensionality reduction. Cell-to-cell similarity was quantified by constructing a neighborhood graph with *scanpy.pp.neighbors*. Clustering was performed using the Leiden algorithm with a resolution parameter of 0.2. After visual inspection and correlating cluster-wise read count profiles with existing replication timing tracks, library sizes, and MAPD scores, we determined that the three clusters (QC clusters 0, 1 and 2) in NPC data are "high quality", "likely replicating" and "poor quality", respectively. We subsequently applied *scanpy.tl.ingest* to project these labels and embeddings from NPC to human brain single cell

TbA data with default parameters. Only cells labeled as "high quality" (cluster 0) were retained. In addition, we excluded batches with a passing rate (defined as the proportion of cells in cluster 0 relative to the total cell count in a given batch) below 65%, as we reasoned that low pass rates indicated poorer quality samples overall.

Single-cell CNV calling pipeline

After filtering at the cellular and sample level as above, we called CNVs on single-cell genomes using first a depth-only mode, followed by validation using fragment overlap density (FOD; Supplemental Methods) to improve calling precision.

CNV calling based on depth

Hiscanner [3] was applied in "rdr_only" mode with specific parameters optimized for our analysis. The bin size was set to 500 kb mappable positions. The hg19/GRCh37 genome was used as reference. Lambda was set to 16, and ploidy restriction was enabled (*max_wgd=1*). The multisample option was disabled (*multisample=false*) to process each cell individually. Hiscanner verison 1.4 was installed via pip install hiscanner. To ensure the reliability of detected CNVs, we implemented a stringent filtering approach: CNV candidates with a p-value ≥ 0.05 or those spanning fewer than 10 bins (approximately 5 Mb of mappable positions) were excluded. We also removed candidates within the pseudoautosomal regions (PARs) of chromosomes X and Y, as defined in the hg19/GRCh37 genome assembly.

CNV validation based on FOD

To refine CNV calls, the most confident diploid segments in each cell were identified as the five longest regions with a predicted copy number of 2, assuming no whole-genome doubling. The mean and standard deviation of FOD values from these regions established a cell-specific baseline for the diploid state. Z-scores were calculated for each CNV based on this baseline. CNV calls were retained if they met the thresholds of FOD Z-score > 1.96 for gains (CN > 2) or FOD Z-score $<$ -1.96 for losses (CN < 2). Cells with a diploid mean FOD of zero were excluded. Filtered CNVs were assigned their HiScanner-derived copy number, while calls failing to meet the threshold (FOD-rejected CNV) retained the diploid background value of 2.

Copy number signature analysis

To identify copy number signatures at the single-cell level, we used SigProfilerAssignment (version 0.1.4) in conjunction with SigProfilerMatrixGenerator (version 1.2.26) [5, 6]. Given the lack of allele-specificity in our single-cell copy number calls, we developed a conversion scheme to transform our integer copy number states into pseudo-allele-specific states. Specifically, we mapped copy numbers into allelic states as follows: 0 to 0|0, 1 to 1|0, 2 to 1|1, 3 to 2|1, 4 to 3|1, and so on up to a maximum copy number of 11 (10|1). This approach allowed us to build the input 48-channel copy number summary vectors that capture the complexity of copy number alterations while maintaining compatibility with existing signature analysis frameworks. SigProfilerAssignment was then executed using default parameters, utilizing the COSMIC version 3.4 pan-cancer signature panel (CN1-22, also referred to in our paper as SigCN1-22) [7] as reference.

To assess the quality of our signature analysis and explore patterns across cells, we performed hierarchical clustering on the signature exposures using Euclidean distance and average linkage. The resulting dendrogram revealed distinct groups of cells with similar signature exposure profiles, potentially indicating shared underlying genomic etiologies. We further evaluated the robustness of our signature assignments by examining the cosine similarity between the original copy number profiles and their reconstructions based on the assigned signatures. The mean cosine similarity was

calculated for each cluster of cells, providing a measure of how well the signature decomposition captured the copy number patterns within each group.

Breakpoint distribution analysis

To investigate the relationship between CNV breakpoints and genomic features, we analyzed replication timing and DNA double-strand break (DSB) distributions. For replication timing analysis, we utilized the replication timing track of BG02ES (human embryonic stem cell) line generated by the ENCODE project [8, 9] downloaded from the UCSC Genome Browser [10]. The average replication timing signal was calculated for each 500 kb mappable bin and normalized between 0 and 1. For DSB analysis, BLISS coordinates were downloaded from FigShare [11, 12], and DSB BED files for each replicate were intersected with the 500 kb genomic bins to quantify DSB signals, which were then normalized between 0 and 1.

To establish a background distribution for CNV enrichment analysis, 10,000 random CNVs were generated for each original CNV, matching their length distribution within autosomes (chr1-22). Peri-centromeric and telomeric regions were excluded to minimize biases, with centromeric regions defined by extending UCSC cytoband coordinates by ±5 Mb and telomeric regions defined as the first and last 5 Mb of each chromosome. Replication timing data were divided into deciles representing quantiles across the genome. For each decile, observed-to-expected (O/E) ratio was calculated by normalizing observed breakpoints (i.e., start and end sites of a CNV) to the expected number based on random simulations. Correlation between replication timing O/E ratios for CNVs from different categories (age group, signature group) was calculated, and statistical significance was evaluated using Pearson correlation tests.

To analyze the relationship between CNV breakpoints and DSBs, we compared CNV breakpoint distributions to BLISS data from six different neural cell preparations: two biological replicates each from human neuroepithelial stem cells (NES), neural progenitor cells (NPC), and post-mitotic neural cells (NEU). For each BLISS track, we identified hotspots of DSB activity and calculated (O/E) ratios comparing the frequency of CNV breakpoints overlapping these hotspots to randomly simulated background breakpoints, following the same permutation approach used for replication timing analysis. For statistical robustness, we aggregated data across NPC and NEU tracks (excluding NES due to its high correlation with replication timing [Pearson $r > 0.5$, $p < 0.05$]). The O/E ratios were capped at 4 for better visualization, and statistical significance was assessed using t-tests comparing aggregated O/E ratios between hotspot and non-hotspot regions within each signature group, with p-values adjusted for multiple testing using the Bonferroni correction.

CNV gene set enrichment analysis

To identify hot and cold spots in Diploid Predominant neurons, we adapted an approach similar to Sun *et al* [13]. We defined genomic bins of 5 Mb by aggregating the existing 500 kb bin coordinate set that was used in CNV calling, with any remaining small segments at the end of the chromosome discarded. Each 5 Mb bin was assigned an empirical overlap count, representing the number of overlaps with observed CNVs. We then performed 10,000 permutations, as outlined in the "Breakpoint distribution analysis," preserving the CNV sizes. For each permutation, the number of overlaps with random CNVs was also calculated, allowing us to generate a null distribution for each bin. P-values were calculated to assess whether the observed CNV overlaps deviated significantly from the random expectation, with separate p-values for enrichment ("hotspots") and depletion ("coldspots"). Adjusted p-values were computed using Bonferroni correction, and significant hotspots and coldspots were identified accordingly. Each 5 Mb bin was further evaluated for overlaps with gene sets, including housekeeping genes (Gene Ontology), loss of function intolerant genes (pLI [14]>0.95), SFARI [15] Autism genes, neurodevelopmental disorder [16], schizophrenia [16] gene sets, and cancer driver genes from COSMIC [17], and epilepsy (OMIM [18] symptom search "epilepsy"). Based on their adjusted p-values, bins were classified as hotspots, coldspots, or neutral regions, with a significance

cutoff of 0.05. To assess gene set enrichment within hotspots and coldspots, hypergeometric tests were performed to evaluate over-enrichment of each gene set within these regions compared to all genomic bins.

Single-nuclei multiomic ATAC/RNA library preparation

Nuclei were isolated as above and permeabilized after flow cytometry according to the Nuclei Isolation from Complex Tissues for Single Cell Multiome ATAC + Gene Expression Sequencing protocol from 10X Genomics (Demonstrated Protocol, CG000375). Chromium Next GEM Single Cell Multiome ATAC + Gene Expression libraries were prepared according to the manual with adaptations to also enable detection of complementary DNA strands (duplexes) (personal communications Andrea Kriz PhD). Duplex information was not utilized for this study.

Cell type and CNV annotation of single-cell ATAC/RNA data

The multiomic dataset consists of snRNA-seq and snATAC-seq data from prenatal and postnatal brain samples, produced following the 10X Genomics protocol. The multiome data was processed using the 10x Genomics Cell Ranger ARC pipeline (version 2.0.1). The multiomic dataset was analyzed with Seurat (version 5.0.1) [19] and Signac (version 1.12.0) [20] R packages. For both prenatal and postnatal brain samples, cells were filtered on several metrics. First, multiplet detection was performed on snATAC-seq data using AMULET, after which cells classified as multiplets were filtered out. Filtering thresholds for several metrics were then determined by visual inspection of rank plots of each metric for all samples. For total RNA molecule count, RNA feature count (number of genes), and total ATAC fragment count, a lower threshold of 1000 was used to remove empty background cells and no upper threshold was used given multiplet removal had already been performed. For both nucleosome signal and transcription start site (TSS) enrichment values from ATAC data, cells were removed with lower (0.32 for nucleosome signal and 1.75 for TSS enrichment) and upper (1.59 for nucleosome signal and 6.80 for TSS enrichment) thresholds determined by the rank plots. Additionally, a lower threshold of 0.5 was determined from the rank plots for the fraction of reads in peaks.

Cell type annotation based on snRNA-seq

Cell type annotation was performed by mapping each sample to a reference sample. We chose separate references for prenatal and postnatal due to the differences in cell type composition between prenatal and postnatal cells. For the prenatal samples, 29180_F4 was chosen as reference, and for the postnatal samples, the downsampled reference was used from a prior published dataset [21]. The prenatal reference was manually annotated by first performing unsupervised clustering on normalized, PCA-transformed RNA count data. The resulting clusters were then manually annotated using known marker genes for different cell types. A UMAP (Uniform Manifold Approximation and Projection) representation was generated for the prenatal reference. For the postnatal reference, the pre-existing annotations from Herring *et al* [21], based on clustering performed in the paper, were used. The remaining non-reference "query" samples were then projected onto the reference UMAPs using Seurat's MapQuery function. The MapQuery function transferred the manually annotated cell type labels from reference to query samples, resulting in complete cell type annotation for all samples in the multiome dataset.

CNV inference from snATAC-seq

CNV calling was performed using the R package epiAneufinder (version 1.0.3). [22], with a window size of 100 kb, and a minimum fragment number of 20,000 per sample. The parameter *minCNVsize* was set to 0, meaning that the minimum callable CNV size was 1 bin, i.e. 100 kb. After CNV calling was complete, the CNV calls were combined across all samples by taking the intersection of bins

in all samples. For the resulting bins, for each cell across all samples, the percentage of bins with a CNV (either gain or loss) was calculated. The resulting distribution of percentage of bins affected was used to determine a cutoff for our definition of CNV Neurons or CNV Cells. Tukey's rule was applied for identifying outliers, such that any cell with a percentage of bins greater than the third quartile plus 1.5 times the interquartile range (Q3 + 1.5IQR) was classified as an outlier. These outliers were then classified as CNV neurons, and non-outliers as non-CNV neurons.

Differential gene module (DGM) analysis

We analyzed the snRNAseq part of single-cell multiome data separately for prenatal samples (n=3; UMB4827, UMB1932, 29180_F4) and postnatal samples (n=5; APGEQ, U1790, U4643, UMB1278, UMB1864) due to differences in cell type composition. For each age group, sample-specific Seurat objects were merged into a combined dataset while preserving sample identifiers. Gene expression values were log-normalized with a scale factor of 10,000. To assess pathway activation, we calculated module scores using gene sets from the Molecular Signatures Database [23] (MSigDB), including Gene Ontology Cellular Component terms and curated gene sets (C2). For C2, we only included gene sets that matched the keywords "senescence," "apoptosis," "autophagy," and "caspase" – hereafter referred to as "cell elimination C2 pathways." Module scores were computed using Seurat's *AddModuleScore* function using default parameters, which calculates the average expression of each gene set in a given cell after controlling for gene expression program complexity by subtracting the aggregated expression of randomly selected control gene sets with similar average expression levels.

To test for associations between gene module scores and CNV neuron status, we implemented a stratified permutation testing approach. For each gene module, we calculated the observed difference in mean module scores between CNV and non-CNV cells (hereafter referred to as "absolute observed difference"). Statistical significance was assessed through 10,000 permutations, where CNV neuron labels were randomly shuffled within each stratum defined by sample and cell type, preserving the underlying data structure. P-values were calculated as the proportion of permuted differences that exceeded the observed difference in absolute value and were adjusted for multiple testing using Bonferroni correction. In order to avoid confounding by cell-type specific gene expression differences, this analysis was performed separately for each major cell type (excitatory neurons, inhibitory neurons, progenitors, and rest) and developmental stage (prenatal and postnatal).

Clustering of CNV neurons based on DGM and cell death pathways

To identify distinct subpopulations of CNV neurons based on their pathway activation patterns, we performed hierarchical clustering on the normalized gene module scores from a selected set of pathways. Since clustering on all available pathways could introduce noise from non-informative or weakly differential pathways, we focused on the top 10 most differentially activated gene modules from GOCC and cell elimination C2 pathways, ranked by the absolute observed difference (see DGM analysis section above for definition of "absolute observed difference" and the selection of cell elimination C2 pathways). The module score matrix was normalized by subtracting the mean and dividing by the standard deviation across cells. Hierarchical clustering was performed using average linkage with the cosine distance metric. We determined an optimal cluster number (n=2) through visual inspection of the dendrogram and cluster stability.

To visualize pathway activation patterns, we generated heatmaps using the seaborn *clustermap* function with row clustering enabled for gene sets and column clustering for cells. The resulting visualization revealed two main clusters of cells with distinct pathway activation patterns. Cell cluster assignments were validated through multiple approaches, including assessment of cluster sizes (Cluster 1: 1,667 cells; Cluster 2: 2,487 cells) and evaluation of enrichment of CNV neurons using Fisher's exact test. The association between cluster membership and CNV neuron status was quantified using odds ratios with 95% confidence intervals. Multiple testing correction was performed using the Bonferroni procedure.

Method References

[1] Evrony, G. D. *et al.* Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).

[2] Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).

[3] Zhao, Y. *et al.* High-resolution detection of copy number alterations in single cells with HiScanner. *bioRxiv* 2024.04.26.587806 (2024).

[4] Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

[5] Díaz-Gay, M. *et al.* Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. *Bioinformatics* **39** (2023).

[6] Bergstrom, E. N. *et al.* SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC Genomics* **20**, 685 (2019).

[7] Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**, 984–991 (2022).

[8] Thurman, R. E., Day, N., Noble, W. S. & Stamatoyannopoulos, J. A. Identification of higher-order functional domains in the human ENCODE regions. *Genome Res.* **17**, 917–927 (2007).

[9] Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 139–144 (2010).

[10] Perez, G. *et al.* The UCSC genome browser database: 2025 update. *Nucleic Acids Res.* (2024).

[11] Gruber, C. figshare - credit for all your research. https://figshare.com/. Accessed: 2024-11-27.

[12] Ballarino, R. *et al.* An atlas of endogenous DNA double-strand breaks arising during human neural cell fate determination. *Sci Data* **9**, 400 (2022).

[13] Sun, C. *et al.* Mapping recurrent mosaic copy number variation in human neurons. *Nat. Commun.* **15**, 4220 (2024).

[14] Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210 (2019).

[15] Abrahams, B. S. *et al.* SFARI gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).

[16] Werling, D. M. *et al.* Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex. *Cell Rep.* **31**, 107489 (2020).

[17] Sondka, Z. *et al.* The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).

[18] Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. Omim. org: Online mendelian inheritance in man (omim®), an online catalog of human genes and genetic

disorders. *Nucleic acids research* **43**, D789–D798 (2015).

[19] Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).

[20] Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with signac. *Nat. Methods* **18**, 1333–1341 (2021).

[21] Herring, C. A. *et al.* Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428–4447.e28 (2022).

[22] Ramakrishnan, A. *et al.* epiAneufinder identifies copy number alterations from single-cell ATAC-seq data. *Nat. Commun.* **14**, 5846 (2023).

[23] Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).