Neuroscience

Human Brain Barcodes

Darryl Shibata

Department of Pathology, University of Southern California, Keck School of Medicine, Los Angeles, United States

Reviewed Preprint
v1 • October 2, 2024
Not revised

Abstract

Dynamic CpG methylation “barcodes” were read from 15,000 to 21,000 single cells from three human male brains. To overcome sparse sequencing coverage, the barcode had ~31,000 rapidly fluctuating X-chromosome CpG sites (fCpGs), with at least 500 covered sites per cell and at least 30 common sites between cell pairs (average of ~48). Barcodes appear to start methylated and record mitotic ages because excitatory neurons and glial cells that emerge later in development were less methylated. Barcodes are different between most cells, with average pairwise differences (PWDs) of ~0.5 between cells. About 10 cell pairs per million were more closely related with PWDs < 0.05. Barcodes appear to record ancestry and reconstruct trees where more related cells had similar phenotypes, albeit some pairs had phenotypic differences. Inhibitory and excitatory neurons both showed evidence of tangential migration with related cells in different cortical regions. fCpG barcodes become polymorphic during development and can distinguish between thousands of human cells.

eLife Assessment

This study presents a **valuable** conceptual approach that cell lineage can be determined using methylation data. However, the evidence supporting the claims of the author is currently **inadequate**. If the author could carry out some additional experiments as well as explore alternative explanations for the current data, this approach could be of broad interest to neuroscientists and developmental biologists.

https://doi.org/10.7554/eLife.101163.1.sa4

Introduction

Cell lineages outline tissue development. Complete fate maps are possible by direct observation for small organisms such as C. elegans, but various elegant experimental fate markers are employed for larger tissues and longer time intervals (1). For human tissues, prior experimental manipulations are impractical, and genomic alterations are employed. Somatic mutations mark subclones and their fates can be reconstructed with DNA sequencing (molecular clock hypothesis). Recent advances in single cell technologies potentially allow fate map reconstruction at single cell resolution.

Here we show how fluctuating CpG (fCpG) DNA methylation (2) can be used as dynamic barcodes to study human brain development using single cell epigenomes annotated with their locations and phenotypes (3). DNA methylation patterns are usually copied between cell

divisions, but replication errors are much higher compared to base replication, allowing for more differences between daughter cells. DNA methylation modulates expression and their patterns can be used to infer cell phenotypes (3,4), but most fCpG sites are present outside of genes or in unexpressed genes. Criteria for our fCpG barcode are as follows: 1) a defined initial pattern in a progenitor cell; 2) polymorphic changes upon cell division; 3) adequate polymorphism to distinguish between most cells; and 4) capability to record ancestry.

The brain has several features that facilitate barcode development and validation.

Foremost, there is extensive single cell methylation data, with thousands of cells annotated by locations and phenotypes (3). Although billions of cells are present in an adult brain, lineage trees are compact because growth is largely neonatal. The brain also allows for serial "stopwatch" barcode sampling because development follows a caudal to rostral pattern, and groups of neurons characteristically stop dividing and differentiate at different times and locations (5). Brainstem neurons emerge early (6) and their barcodes should most resemble the initial progenitor state, whereas the stopwatch runs longer for excitatory neurons that appear later in development. To facilitate presentation, barcode performance is summarized as follows: The brain fCpG barcode initializes as predominately methylated in the progenitor cell and becomes polymorphic with more diverse barcodes in excitatory neurons that emerge later in development. The barcode becomes sufficiently polymorphic to uniquely distinguish between most sampled brain cells, and barcoded cells organize into lineage trees.

Results

fCpG barcode identification

Barcode development was limited by the sparse single cell data (3), with < 5% of CpG sites sequenced, often with only a single read. Sparse coverage was mitigated with X- chromosome fCpG sites because only a single read can infer a binary (0,1) state in male individuals. Autosomal CpG sites require at least 2 reads to infer three possible states (0, 0.5, 1). The X-chromosome also simplifies the identification of polymorphic fCpGs because many neurons have different binary states if average methylation is between 0.25 and 0.75 in bulk WGBS adult male neurons reference data (4).

CpG sites (N~116,000), with average methylation between 0.25 and 0.75 in bulk neurons from seven males (4), were further filtered by discarding more stable CpG sites with average methylation less than 0.2 or more than 0.8 for all cells, inhibitory neurons, and excitatory neurons in brain H02. The ~79,000 CpGs were further filtered to remove sites with average methylation less than 0.3 or greater than 0.7 in brain H01, and ~31,000 fCpG sites were used for analysis.

fCpG site methylation appears neutral because they are predominately intergenic, with 16% within genes or promoters (File S1). Epigenomes from 15,434 to 21,836 cells were downloaded from three male brains with a general criterion of allc.tsv.gz file sizes 90 mb or larger (**Table 1**). Analyzed cells had at least 500 fCpGs (average ~1,100), with pairwise distances (PWDs) calculated between cell pairs when at least 30 fCpGs were comparable (average ~48 fCpGs per cell pair). Each cell, annotated by its provided phenotype and location, is characterized by its fCpG methylation level and its PWDs from other cells. A PWD of 0 is a perfect match and 0.5 indicates randomization. fCpG methylation was variable between cells with averages of ~58% for all three brains (**Fig 1A**). The 73 to 197 million possible cell pair comparisons revealed polymorphic barcodes with average PWDs of ~0.47 between cells (**Fig 1B**).

Fig 1

fCpG barcode methylation

A) Barcode methylation was variable between cells with averages ~ 50%

B) Most cells had different barcodes with average PWDs ~ 0.5.

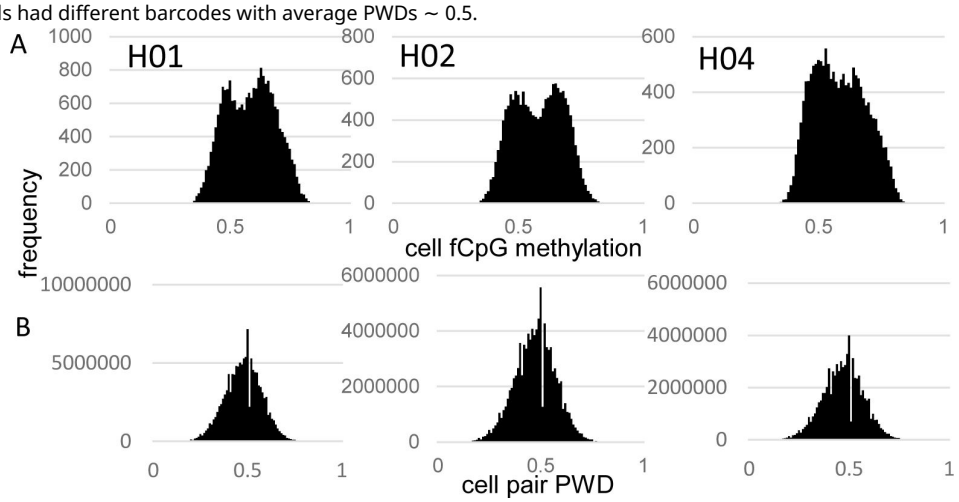


Table 1

Brain data

brain	age	cells	ave fCpG per cell	ave Meth per cell	ave PWD between cell pairs	cell pairs per million	fCpG per pair	closely related pairs (PWD<0.05)	closely related pairs per million	fCpG per nearest neighbor pair
H01	42y	21,836	1170	0.58	0.47	197	51	1385	7.0	35
H02	29y	16,161	1128	0.58	0.47	99	48	743	7.5	34
H04	58y	15,434	1060	0.58	0.47	73	45	1078	14.8	35
between brains										
H02-H01					0.47	281	49	785	2.8	35
H02-H04					0.47	178	46	653	3.7	34
H04-H01					0.47	254	47	943	3.7	35

fCpG barcodes initialize methylated and change with cell division

Brain patterns were similar, and data are presented for H01, with H02 and H04 presented in Figs S1 and S2. Methylation was variable between cells of the same type and average methylation was highest in the brain stem (pons, thalamus), intermediate for inhibitory neurons, and lowest for excitatory neurons, non-neuronal cells, and cerebellar cells (**Fig 2A**). Outer layer cortical excitatory neurons (L2_3) that are made later during development were less methylated than inner cortical excitatory neurons (L4_6) that appear earlier. Predominately methylated individual fCpG sites were common in brainstem neurons, less frequent in inhibitory neurons, and rare in excitatory neurons (**Fig 2B**). Predominately unmethylated individual fCpGs were common in glial and hippocampal cells.

The methylation hierarchy is consistent with a barcode initialized with predominately methylated fCpGs in a progenitor cell. Barcodes becomes progressively demethylated and are fixed when their cells stop dividing and differentiate, which occurs at different times and places during brain development. Simplistically, the brainstem with mature neurons (6) forms early in development, followed by inhibitory neurons in the ganglionic eminences, and then excitatory neurons and glial cells in the cortex. Barcode methylation follows this temporal development and reconstruct when specific neuron types start to appear and reach their adult contents (**Fig 2C**). For example, after barcodes flip from ~100 to ~70% methylated, most adult brainstem and inhibitory neurons are present but excitatory neurons are fewer, with very few adult outer layer (L2_3) neurons. Outer excitatory and glial progenitor cells are present (7), but their barcodes continue to demethylate until they stop dividing and differentiate later in development. This stopwatch like pattern, with more demethylated barcodes in later appearing cell types, was present in all three adult brains.

fCpG barcodes are polymorphic

A progenitor cell barcode should become increasingly polymorphic with subsequent divisions. This pattern was observed, with average barcode PWDs lowest in the brainstem, intermediate between inhibitory neurons, and highest for excitatory neurons (**Fig 3A**). Most cells had different barcodes, with an overall average PWD of ~0.47 (**Fig 1B**). Cells of the same phenotype were more similar with lower average PWDs (**Fig 3A**, 3B), suggesting they are more related to each other and have common progenitors.

The human brain has billions of cells and relatively few cells were sampled from each region. Consistent with sparse sampling, cell pairs with smaller PWDs (< 0.05) were rare. To help distinguish between ancestry and chance, cells within and between brains were compared (**Table 1**). Closely related cell pairs were ~2.9 times more frequent within a brain (average ~9.8 per million) compared to between brains (average ~3.4 per million). Closely related cells had fewer matching fCpG sites (~35 compared to ~48 for all cell pairs) and were more common early in development when barcodes are more methylated (**Fig 3C**), indicating that lower barcode complexity favors matching. Overall, fCpG barcodes are sufficiently polymorphic to distinguish between most adult brain cells.

Brain lineage trees

It should be possible to reconstruct human brain development if barcodes record ancestry. fCpG barcodes from ~1,000 brain cells with different phenotypes yield lineage trees that resemble caudal to rostral development (**Fig 4**). The trees are rooted by a progenitor with a fully methylated barcode, and branches progressively yield brainstem neurons, a subset of excitatory lower (L4_6) neurons, thalamic neurons (THM), inhibitory neurons, cerebellar cells, and non-neuronal cells. Excitatory neurons branch last, and hippocampal neurons (CA, DG) that may divide postnatally (8) were at the terminus. Cells generally grouped by phenotype, with some early

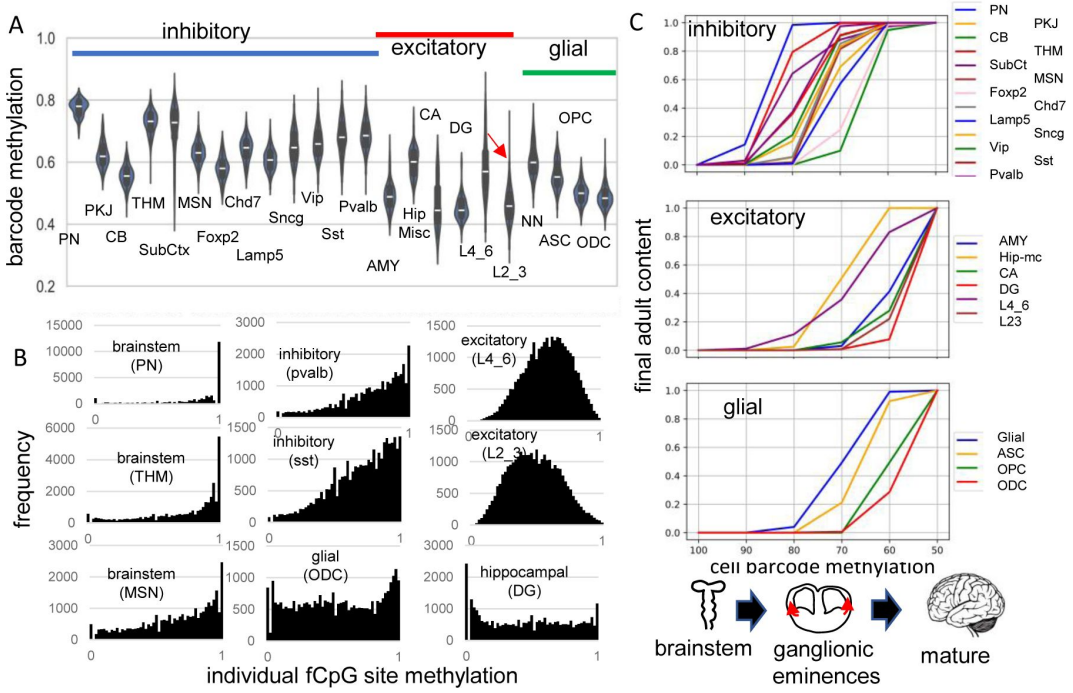


Fig 2

Higher fCpG barcode methylation in earlier emerging cells

A) Average barcode methylation was higher in the brainstem and inhibitory neurons. Barcode methylation was lower for excitatory neurons, cerebellar, and glial cells. Notably, average methylation was lower for outer cortical (L2_3) compared to earlier appearing inner (L4_6) cortical excitatory neurons. Abbreviations are as in reference 3, with L2_3 all outer and L4_6 all inner cortical excitatory neurons, and NN are non-neuronal cells other than ASC, OPC and ODC.

B) Most fCpGs appear to start methylated in a progenitor because nearly all individual fCpGs are methylated in inhibitory neurons in the brain stem (PN, THM, MSN). Many fCpGs in inhibitory neurons (pvalb, sst) are still predominately methylated. Few fCpGs in excitatory neurons that differentiate later in development are fully methylated. Glial cells that also emerge late in development, and hippocampal cells that may divide postnatally had variable methylation with both highly methylated and unmethylated fCpGs.

C) Barcodes become fixed when their cells stop dividing and differentiate. Barcode methylation levels can indicate when neurons emerge and are correlated with a cartoon of physical caudal to rostral brain development. fCpGs infer that inhibitory neurons made in the ganglionic eminences appear before and reach their final adult content before most cortical excitatory neurons and glial cells. Notably, lower cortical layer neurons can be detected earlier in life relative to later appearing outer cortical neurons that reach adult levels late in development. Brain contents inferred by adult barcodes may differ from actual neonatal brains because neurons that die during development are not sampled.

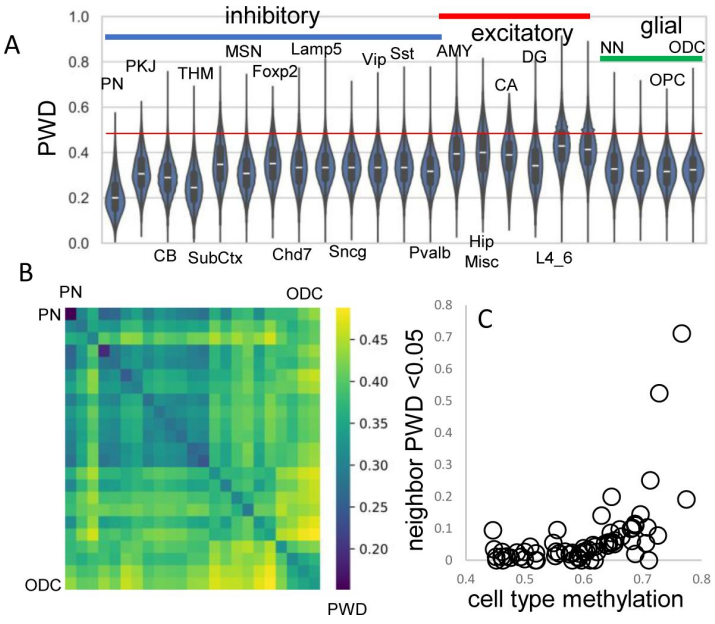


Figure 3

Related cell pairs

A) Most cells had different barcodes with average PWDs between cell pairs of ~ 0.5. Cell pairs of the same phenotype had different barcodes but were on average more related to each other.
B) Heatmap showing that cells of the same phenotype are more related.
C) Cells that emerge early in development are more related and more methylated. Closely related nearest neighbors (PWD <0.05) are numerically more common for more methylated cell types.

appearing excitatory neurons admixed among inhibitory neurons. Similar trees were observed for H02 and H04, albeit with less separation between inhibitory and excitatory neurons for H04 (**Fig 4A**). Barcode lineage trees are largely consistent with expected sequential neuronal differentiation.

Cell lineage fidelity and cortical migration

Barcodes could record neuronal differentiation and migration. Uncertain for mouse and human development is whether inhibitory and excitatory neurons originate from shared or distinct progenitors (3,9,10). Lineage fidelity can be quantified by comparing most closely related cells or nearest neighbor cell pairs with PWDs < 0.05. Lineage fidelity was high (>90%) for inhibitory neurons (**Fig 5A**). Excitatory lineage fidelity was slightly lower, indicating that some excitatory and inhibitory neurons may share common progenitors (10). Lineage trees indicate common progenitors are present earlier in development, and excitatory neurons that appear later do not have many closely related inhibitory neighbors (**Fig 4B**). The barcodes documented the known switching between inhibitory neuron subtypes (**Fig 5B**). Brainstem, cerebellar, excitatory, and non-neuronal cells had less subtype lineage fidelity.

Barcodes can also infer migration because their neurons are annotated by their adult locations. Trees (**Fig 4C**) indicate that most neurons sampled from the brainstem, cerebellar and hippocampal regions are related and localized to their respective regions. Inhibitory neurons were scattered throughout the cortex, consistent with their differentiation in the ganglionic eminences and subsequent tangential migration to the cortex. Nearest neighbor inhibitory cortical neuron pairs were found in the same cortical region ~25% of the time (**Fig 5C** and **5D**). Nearest neighbor excitatory neuron pairs were also scattered throughout the cortex, but less than inhibitory neurons, and were in the same cortical region ~50% of the time.

The poor ability to detect localized excitatory neuron radial cortical migration with ~1,000 cell whole brain trees (**Fig 4C**) may reflect that sparse sampling is unlikely to include multiple neurons from the same small clonal region (radial unit hypothesis (11)). Greater localized excitatory neuron migration was seen when trees were reconstructed with more (~2,800) neurons, while inhibitory neurons still showed scattered tangential migration (**Fig 4D**). Neurons of the same subtype were still more related. Hence, lineage trees appear to increase their resolution with more cells, albeit related lower and upper excitatory neuron pairs were still uncommon, which may reflect the unlikely chance of sampling very small radial clonal units.

fCpG methylation and post-mitotic epigenetic remodeling

After progenitors stop dividing, differentiation occurs through epigenetic remodeling and neuron specific methylation (12). fCpG patterns could reflect this post-mitotic remodeling because neurons of the same phenotype generally have more similar barcodes (**Fig 3A**, B). To help separate ancestry from post-mitotic phenotypic differentiation, barcodes, and the gene methylation tsne coordinates used to phenotype the neurons (3), were compared (**Fig 5E**). Cell pairs with closely related fCpG barcodes had similar phenotypic methylation patterns, but many cells with small phenotypic differences had very different barcodes. Hence, fCpG barcodes do not correlate well with post-mitotic epigenetic remodeling because phenotypically similar cells can be unrelated.

Discussion

Dynamic barcodes would be useful to study human tissues, but testing their performance is difficult. Ideally, samples obtained at different times would document how they change. The brain facilitates barcode validation because it periodically stores neurons that stop recording at

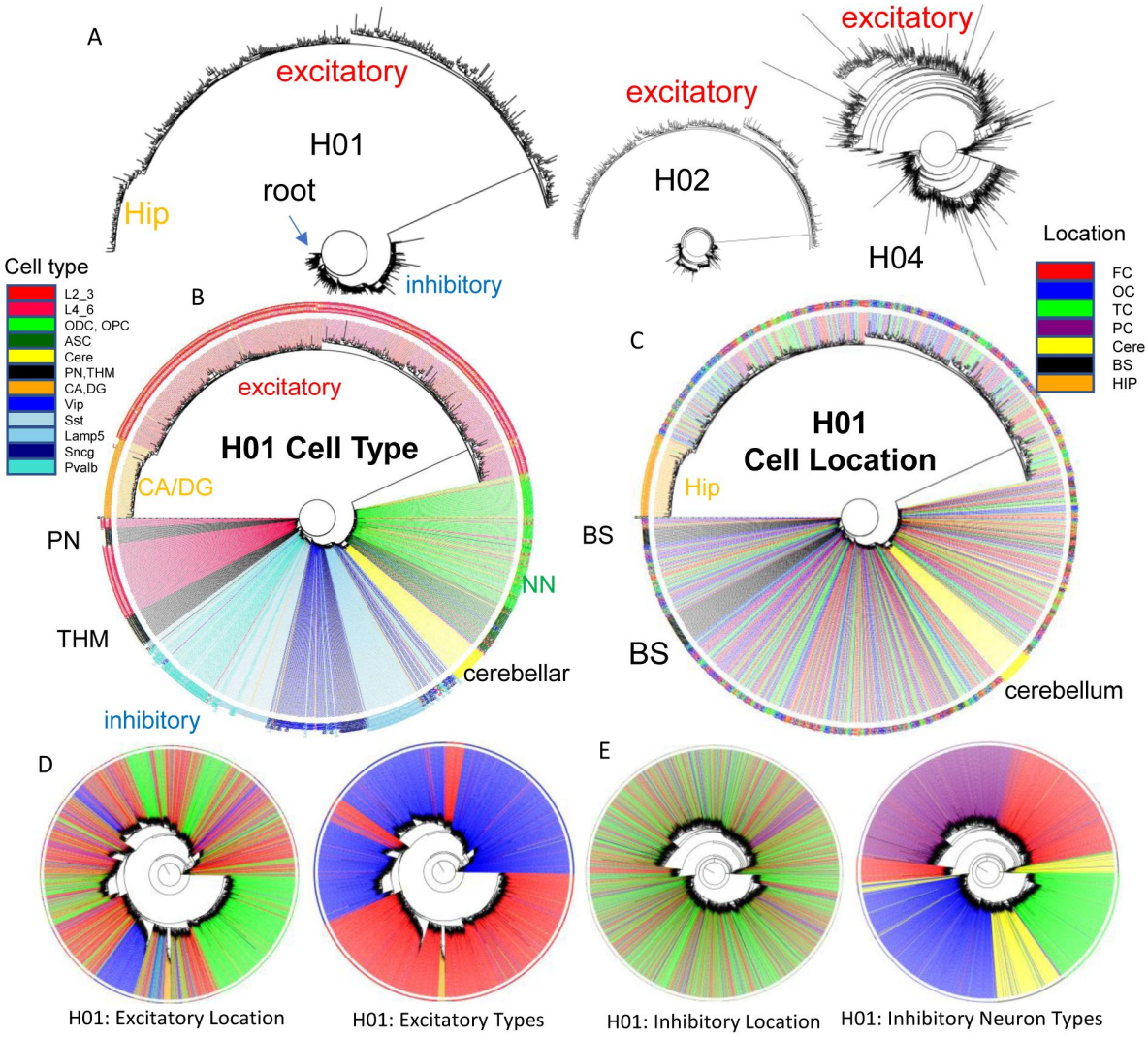


Figure 4

Brain lineage trees

A) Barcodes from 960 cells form lineage trees using IQtree (19) that are rooted by a fully methylated progenitor and generally follow caudal to rostral brain development, with sequential branching of brainstem neurons, inhibitory neurons, cerebellar neurons, and excitatory neurons, with hippocampal neurons furthest from the start. Trees are similar between the brains, with H04 inferring less distance between inhibitory and excitatory lineages. The degree of confidence was generally low with the data and numbers of bootstraps, with bootstrap branch support typically less than 15%.
B) H01 tree with labeled cell types. Neuron types generally clustered by phenotypes with closely branching excitatory and inhibitory neurons more common earlier in development.
C) H01 tree with labeled cell locations. Related inhibitory and excitatory neurons can be found in different parts of the brain (FC = frontal (red), TC = temporal, OC = occipital (blue), PC = parietal, HIP = hippocampus (orange), cere = cerebellum (yellow), BS = brainstem (black).
D) H01 tree with ~2,853 cortical excitatory neurons has more evidence of localized radial migration because related neurons are more often found in the same cortical region. Excitatory neurons cluster by subtype, and closely related lower and upper excitatory neurons were still few.
E) H01 tree with ~2,847 cortical inhibitory neurons still retains evidence of tangential migration with related neurons scattered throughout the cortex. Inhibitory neurons cluster by subtype with switching between some closely related pairs.

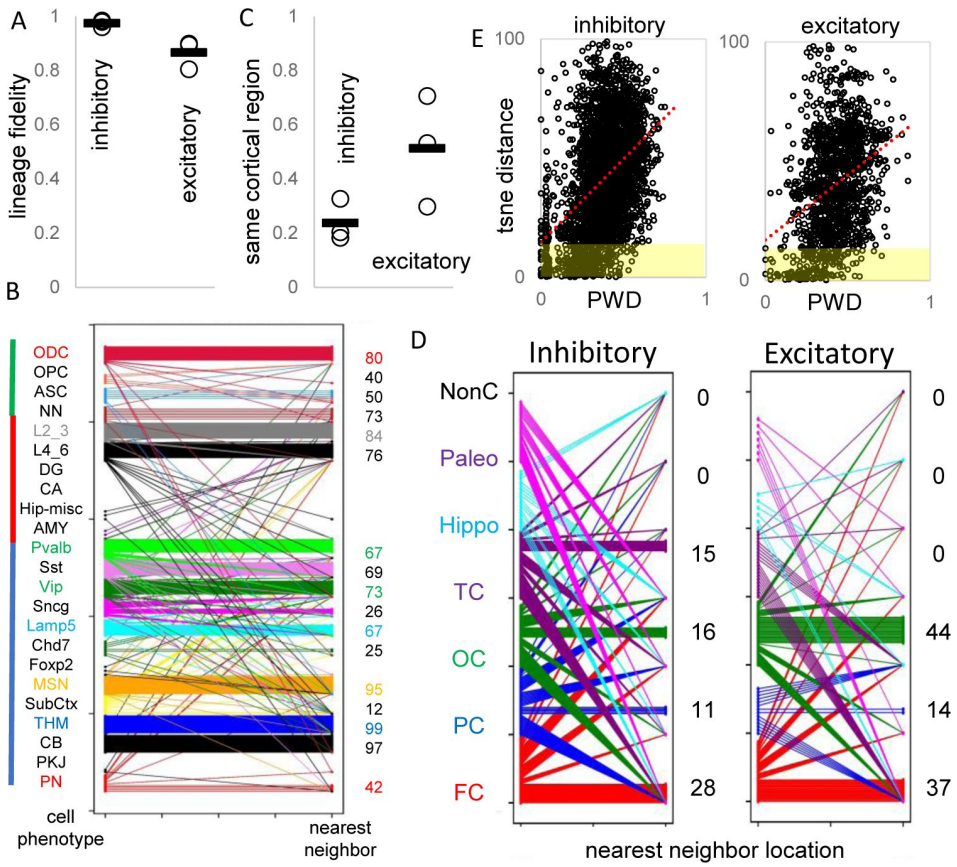


Figure 5

Lineage fidelity, migration, and differentiation

A) Inhibitory neurons have high lineage fidelity because nearest neighbor pairs (PWD <0.05) were nearly always both inhibitory neurons. Excitatory neurons had less lineage fidelity because a nearest neighbor was more often an inhibitory neuron. Data are for all three brains.

B) Nearest neighbor inhibitory neuron pairs often had subtypes differences. More lineage fidelity was generally present for brainstem and excitatory neurons. Numbers indicate percent lineage subtype fidelity.

C) Nearest neighbor inhibitory and excitatory neuron pairs showed evidence of tangential migration because they were found in different cortical regions. Data are for all three brains.

D) Nearest neighbor neurons were scattered in the cortex. Numbers indicate percent location fidelity. (NonC = non-cortical location, Paleo = paleocortex)

E) Phenotypic differences between inhibitory or excitatory neuron were measured by Euclidian distances between their annotated tsne coordinates (3). Although cells with more similar phenotypes (shaded yellow) were generally more related, cells with similar phenotypes could be distantly related, and related cells could have different phenotypes.

relatively defined times and locations (1,13). Specific neuron subsets recovered from the adult brain allow for sampling through time and before birth.

This serial sampling strategy facilitated fCpG barcode validation. The barcode started predominately methylated in multiple individuals and became sufficiently polymorphic to distinguish between thousands of neurons. Barcode changes appear to represent replication errors because they reconstruct lineage trees consistent with caudal to rostral brain development. Barcode methylation indicates when different neurons that survive to adulthood appear in the neonatal brain (**Fig 2C**).

The current barcode indicates that inhibitory and excitatory neurons have relative distinct progenitors, consistent with the lineage dendrograms reconstructed with neuron specific methylation of the same data (3). There was also evidence for common inhibitory and excitatory progenitors (10), primarily for earlier emerging excitatory neurons. Tangential migration was also detected, manifested by inhibitory neurons with closely related barcodes in different cortical regions. Tangential excitatory neuron migration was also detected, albeit related excitatory neurons were more localized than inhibitory neurons. Tangential migration is also seen with sequencing studies that find neurons with specific mutations in multiple brain regions (14–17).

fCpGs more efficiently distinguish between cells than mutations due to higher replication error rates. Although average methylation decreases with time, both demethylation and remethylation are likely because fully demethylated neurons were not observed, and balanced fluctuating methylation is inferred in other tissues when CpG sites are ~50% methylated in bulk tissues (2). More adult divisions in brain cancers did not saturate the barcode with average fCpG methylation ~50% (Fig S3). Fluctuating methylation complicates lineage tracing but backmutations can be modeled for ancestral reconstructions. Lineage resolution could be improved by combining mutations and fCpGs.

Weaknesses of this study including very sparse cell sampling and lack of uniform CpG sites comparisons between neurons. Inferred lineage trees (**Fig 4**) had relatively low statistical support for their branches. Like many human fate markers studies, it is difficult to independently verify accuracy. However, preliminary studies are largely consistent with brain development and sequential stopwatch like neurogenesis. Technical improvements such as targeted bisulfite sequencing of a limited number of informative fCpGs could lead to more consistent coverage and less expensive sequencing of more neurons. Single cell measurements of small numbers of fCpGs, and snMCode cell type specific CpG sites (3), could efficiently reconstruct human brain lineages. A barcode of 100 fCpGs has enough complexity (2^{100} or ~$1X10^{30}$) to potentially distinguish between most excitatory neurons, with less resolution early in development when cells are inherently more related.

The analysis of more brains can verify that a fCpG barcode starts predominately methylated in most individuals. A common initialized state could facilitate standardized human fate maps and comparisons between individuals. Many polymorphisms linked to brain abnormalities such as autism are in neuronal proliferation, migration, and maturation pathways (18), and this preliminary survey indicates lineage heterogeneity between individuals (**Fig 4A**). fCpG barcodes have been applied to the intestines, endometrium, and blood (2), and could be found for multiple other tissue types, helping to unravel human development and aging.

Methods

Brain Single Cells

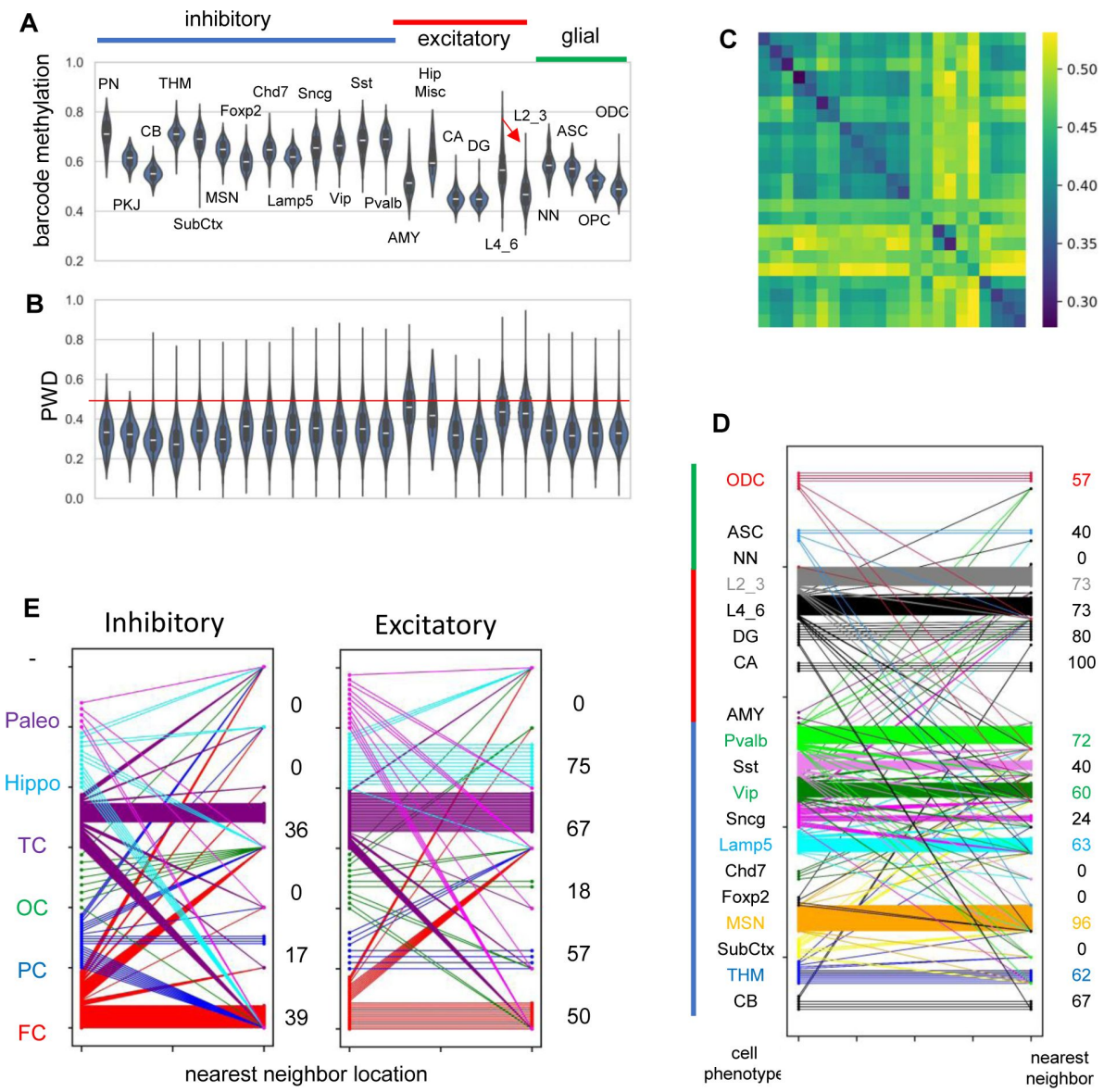
Single cells with their annotations and methylation at each fCpG site were read from single cell files downloaded from GEO (GSE215353) and supplemental files from reference 3. Lists of fCpG sites and data summarized for the Figures are in Supplemental File 2. The cells and methylation at the fCpG sites are in Supplemental Files 3-5. PWDs were calculated between all cell pairs with at least 30 matching fCpG sites, with PWD data matrices in Supplemental File 6.

IQtree

IQtree (19) tree was downloaded and run on a server with 64 cpus and 32 GB of memory. The model (GTR2+FO+G4) accounts for backmutation and binary data with missing values. Bootstraps were 1,000 per tree with 3,000 iterations for whole brain (~1,000 cells) trees and 1,000 iterations for inhibitory or excitatory (~2,800 neurons) trees. Trees (.treefile) were displayed with FigTree (*http://tree.bio.ed.ac.uk/software/figtree/*) with truncation of long branches (generally fewer than 10) for display purposes. The cells used for the trees are in Supplemental File 7.

Acknowledgements

This work was supported by grants from the NIH (P01CA196569 and CA271237). I thank Drs. Trevor Graham and Heather Grant for useful discussions, and Omar Khan and Nikhil Krishnan for initial studies. The author thanks all of the researchers that helped produced the high quality very valuable and freely available data used for analysis.



Supplemental Figure 1

H02 data

A) Barcode methylation for different cell types

B) PWDs between cells of the same type

C) PWDs between cell types

D) Lineage cell type fidelity between nearest neighbor pairs (PWD<0.05)

E) Location fidelity between nearest neighbor pairs (PWD<0.05)

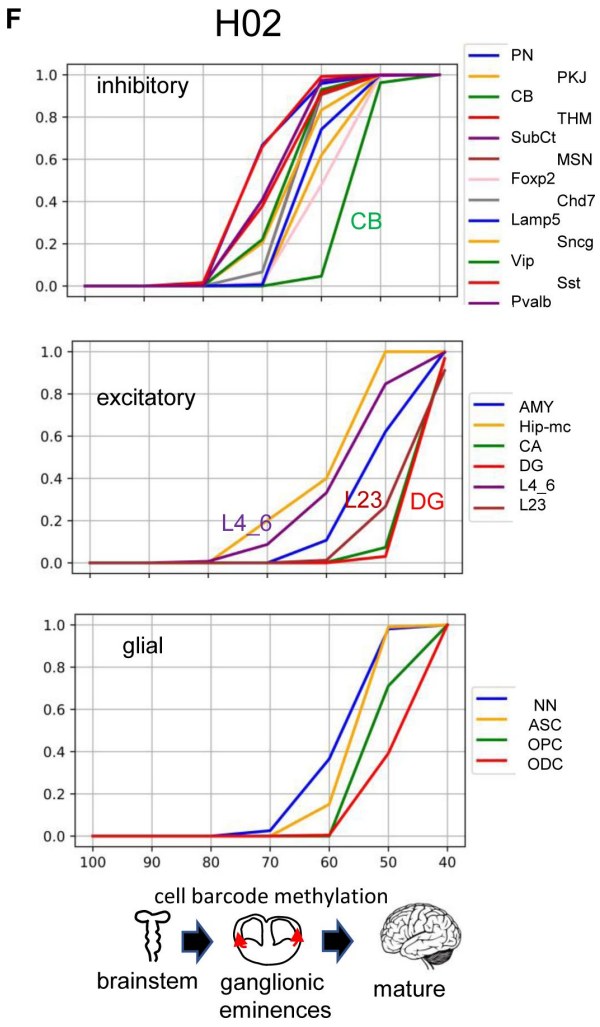
F) Barcode methylation versus final adult brain content indicates that inhibitory neurons appear first and reach their adult levels before excitatory or glial cells.

G) Ancestral tree with 1,001 cells rooted at a fully methylated progenitor shows sequential branching with excitatory, then brain stem, inhibitory and cerebellar neurons, then glial cells, and finally excitatory neurons with hippocampal neurons at the end.

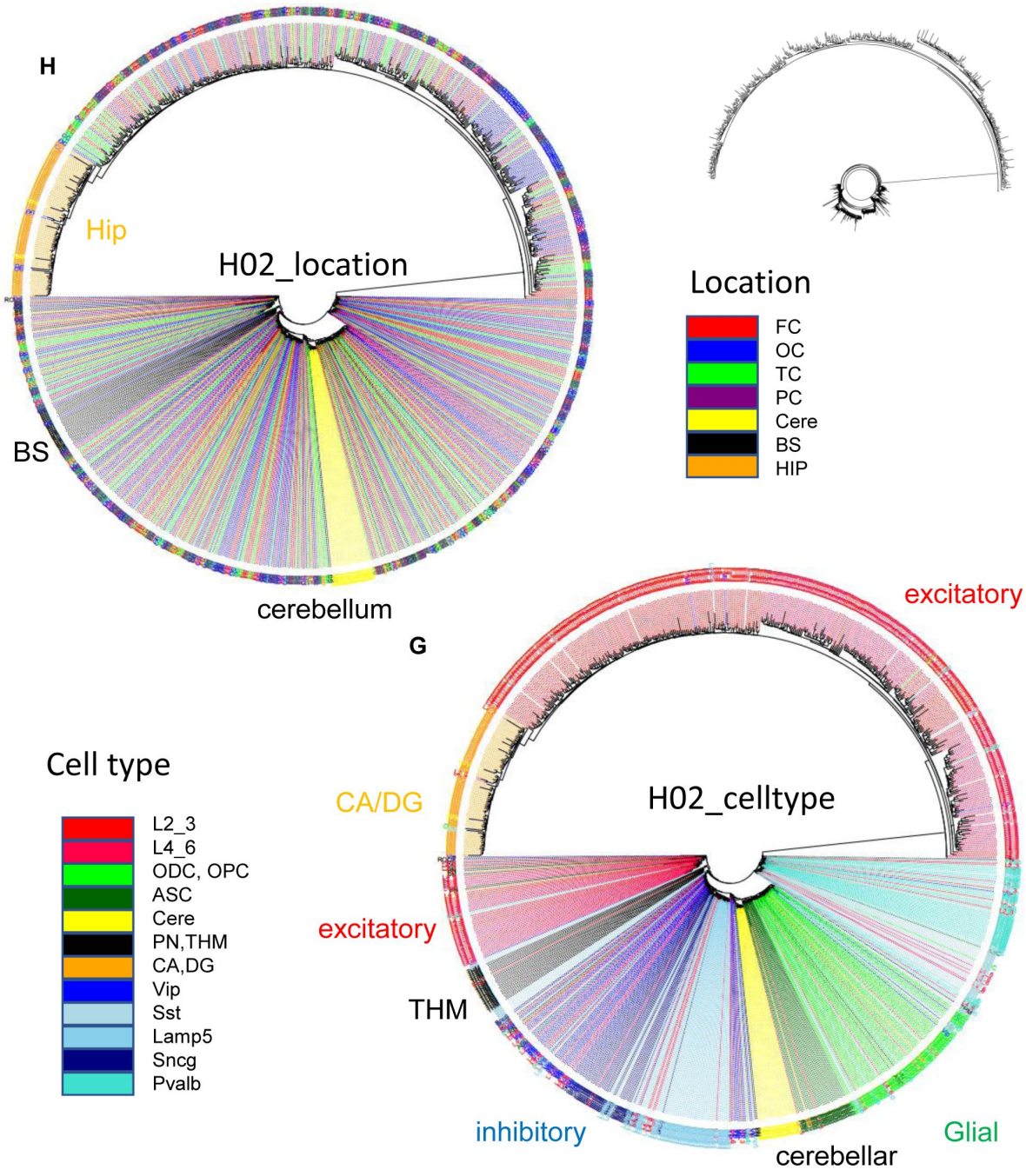
H) Related cells colocalize for brain stem, cerebellar, and hippocampal neurons. Inhibitory neurons are more scattered. Excitatory neurons are also scattered with some localization within the cortex (see I & J for trees with more neurons)

I) Ancestral tree with more (2,806) excitatory neurons shows more localization between related neurons in the hippocampus, occipital and temporal cortex. Related neurons also tend to have similar phenotypes.

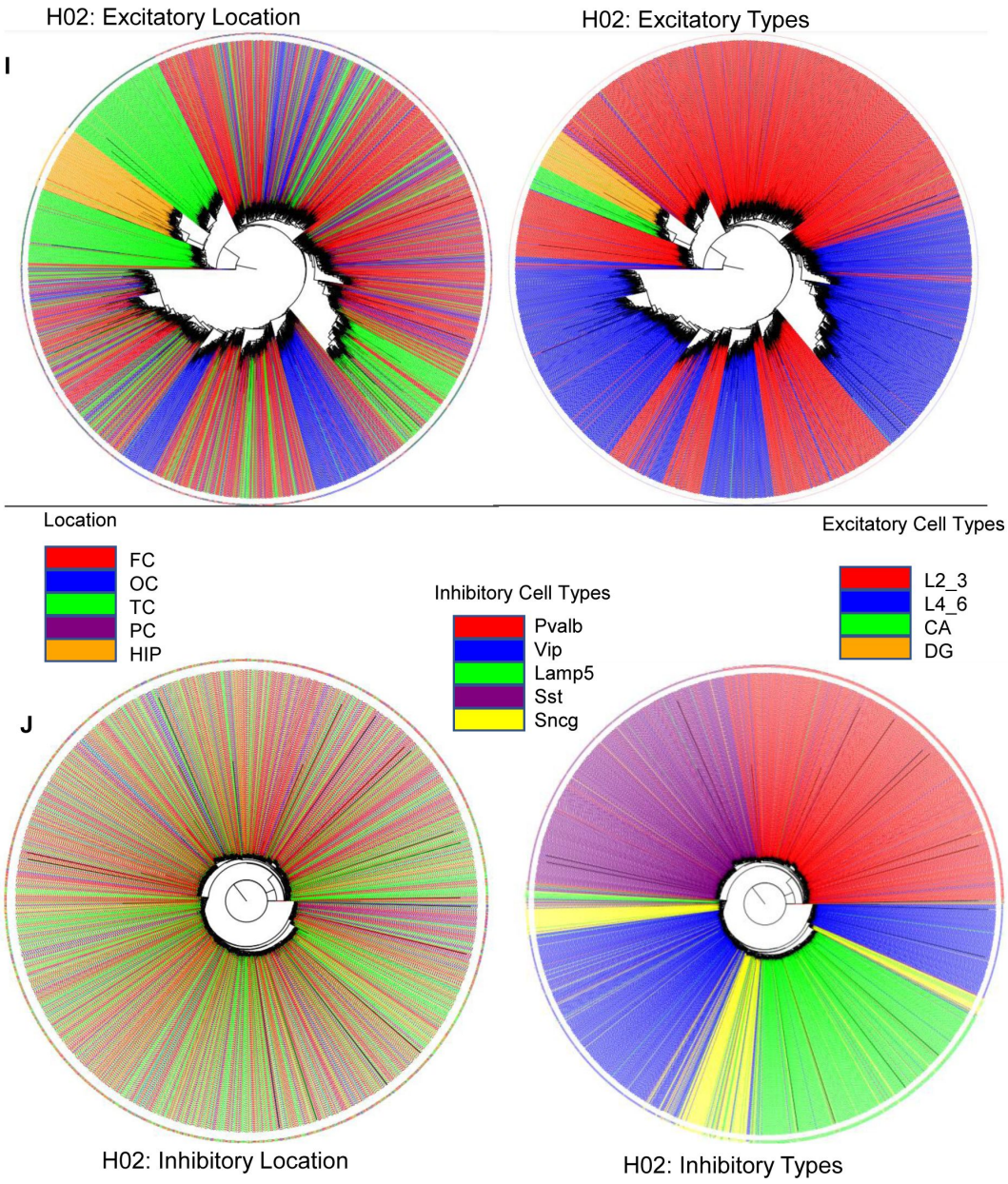
J) Ancestral tree with more (2,788) inhibitory neurons still shows scattering between related neurons. Related neurons tend to have similar phenotypes.



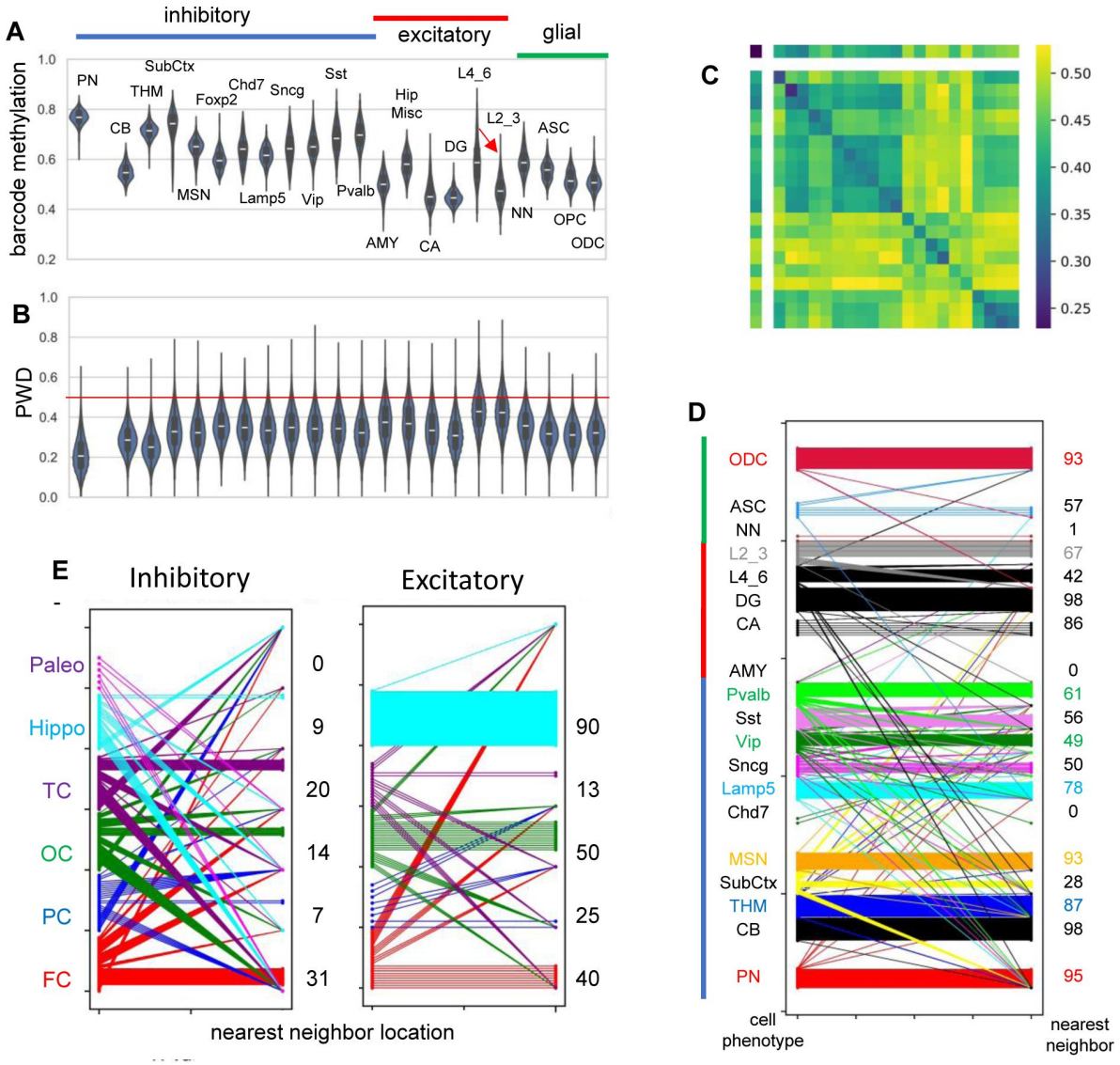
Supplemental Figure 1 (continued)



Supplemental Figure 1 (continued)



Supplemental Figure 1 (continued)



Supplemental Figure 2

H04 data

A) Barcode methylation for different cell types

B) PWDs between cells of the same type

C) PWDs between cell types

D) Lineage cell type fidelity between nearest neighbor pairs (PWD<0.05)

E) Location fidelity between nearest neighbor pairs (PWD<0.05)

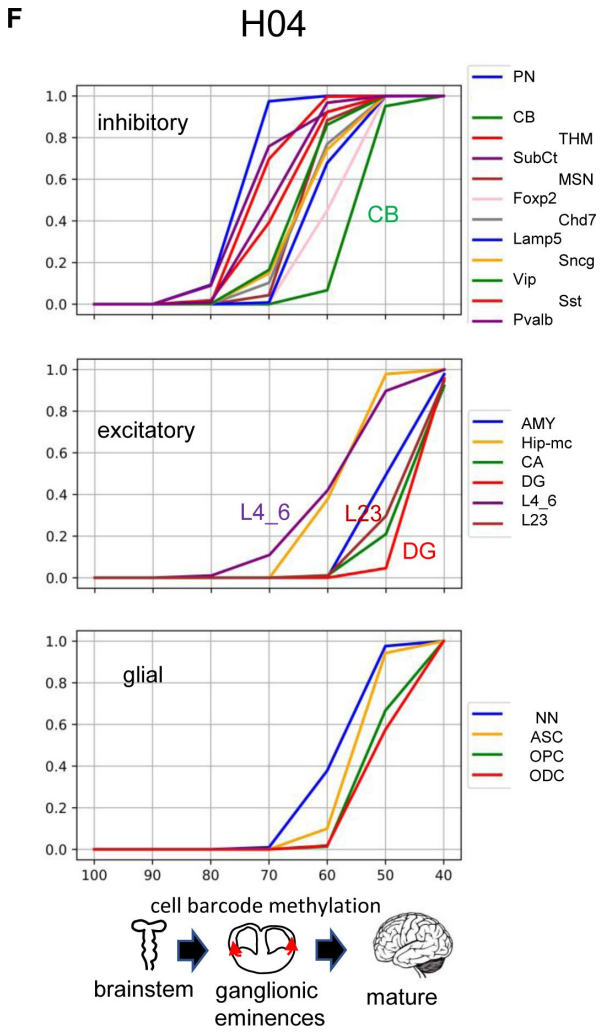
F) Barcode methylation versus final adult brain content indicates that inhibitory neurons appear first and reach their adult levels before excitatory or glial cells.

G) Ancestral tree with 1,033 cells rooted at a fully methylated progenitor shows sequential branching with excitatory, then brain stem, inhibitory and cerebellar neurons, then glial cells, and finally excitatory neurons with hippocampal neurons at the end.

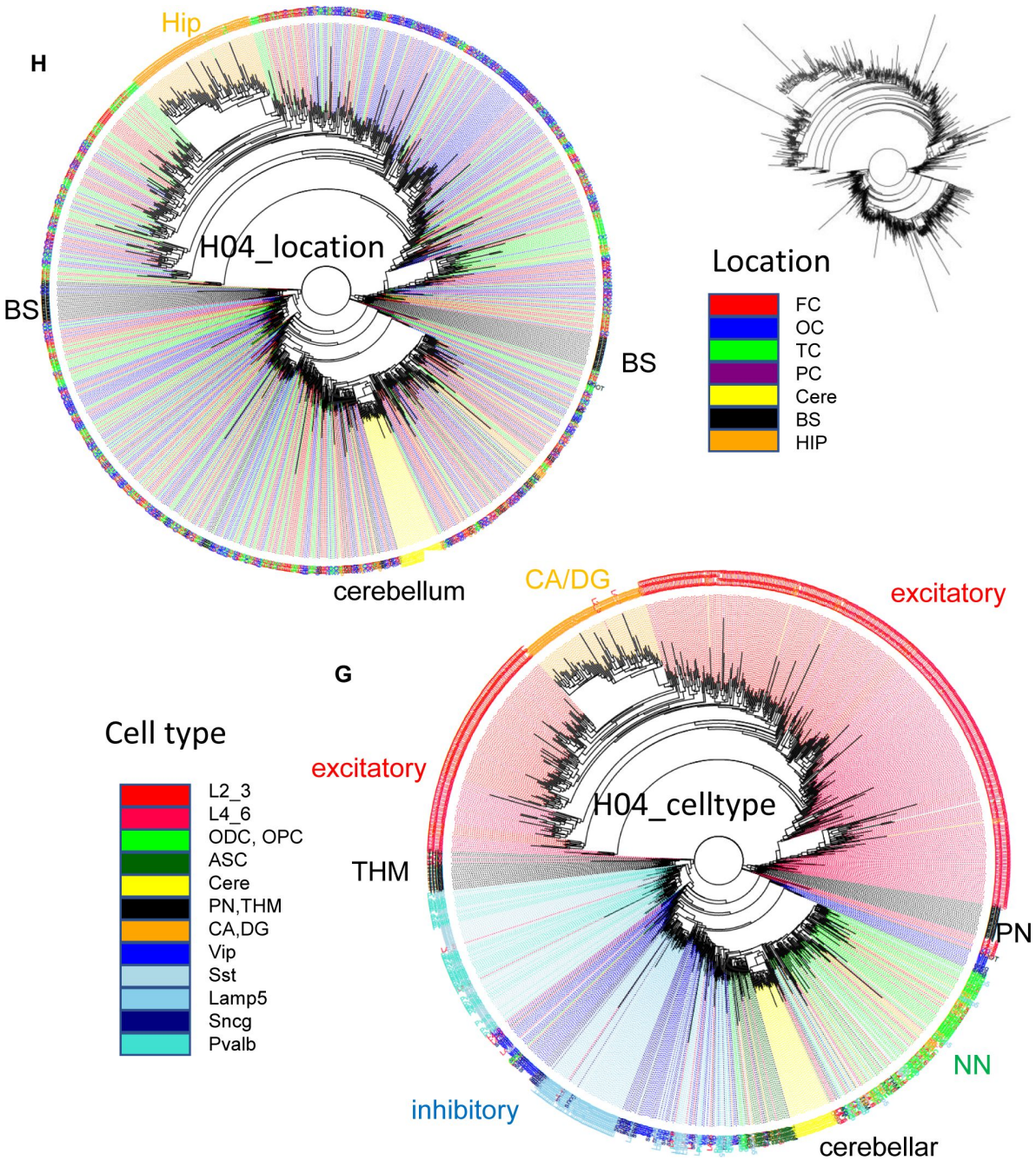
H) Related cells colocalize for brain stem, cerebellar, and hippocampal neurons. Inhibitory neurons are more scattered. Excitatory neurons are also scattered with some localization within the cortex (see I & J for trees with more neurons)

I) Ancestral tree with more (2,797) excitatory neurons shows more localization between related neurons in the hippocampus, occipital and temporal cortex. Related neurons also tend to have similar phenotypes.

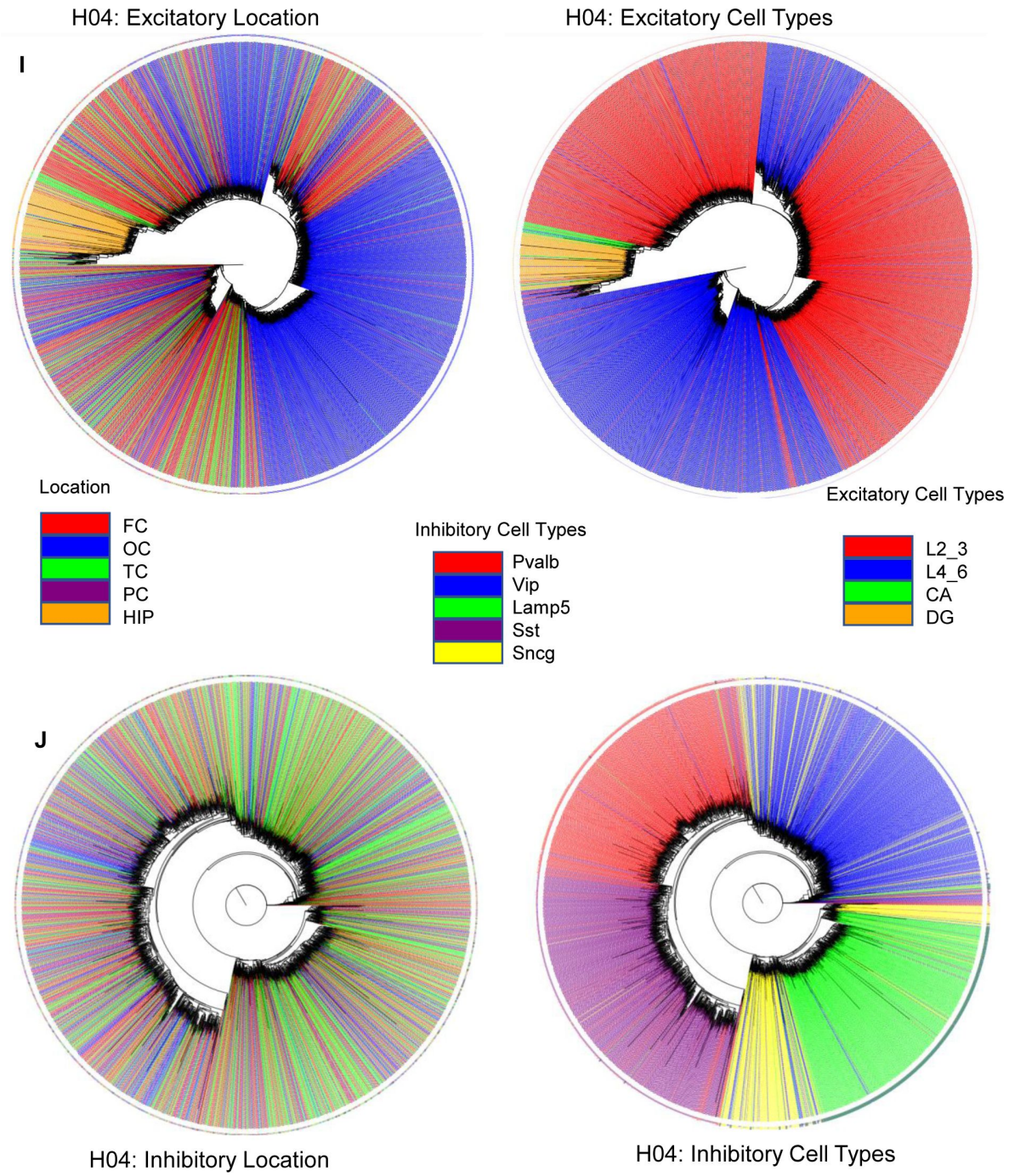
J) Ancestral tree with more (2,752) inhibitory neurons still shows scattering between related neurons. Related neurons tend to have similar phenotypes.



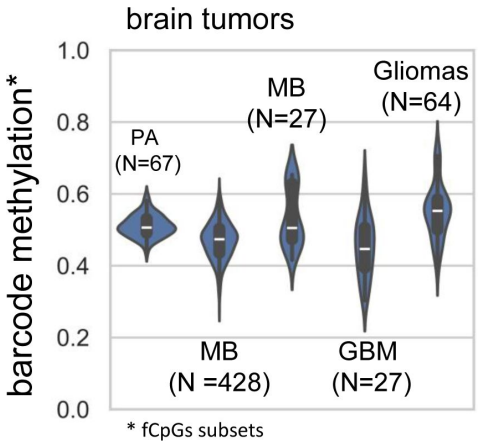
Supplemental Figure 2 (continued)



Supplemental Figure 2 (continued)



Supplemental Figure 2 (continued)



Supplemental Fig 3

Brain tumor fCpG methylation:

fCpG sites were matched with brain tumor data from methylation arrays. Average methylation is displayed for each brain tumor sample. PA= pilocytic astrocytoma, GSE44684, 145 fCpGs covered; MB=medulloblastoma, GSE193646, 434f CpGs covered; MB=medulloblastoma, GSE63669, 883 fCpGs covered; GBM=glioblastoma multiforme, GSE109399, 1,138 fCpGs covered; Gliomas=various gliomas, GSE248471, 1,138 fCpGs covered. Includes both male and female tumors.

References

1. McKenna A., Gagnon J. A (2019) **Recording development with single cell dynamic lineage tracing** *Development* **146**

2. Gabbutt C. *et al.* (2022) **Fluctuating methylation clocks for cell lineage tracing at high temporal resolution in human tissues** *Nature Biotechnology* **40**:720–730

3. Tian W. *et al.* (2023) **Single-cell DNA methylation and 3D genome architecture in the human brain** *Science* **382**

4. Loyfer N. *et al.* (2023) **A DNA methylation atlas of normal human cell types** *Nature* **613**:355–364

5. Stiles J., Jernigan T. L (2010) **The basics of brain development** *Neuropsychology Review* **20**:327–348

6. Fan X. *et al.* (2020) **Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development** *Science Advances* **6**

7. Eze U. C., Bhaduri A., Haeussler M., Nowakowski T. J., Kriegstein A. R (2021) **Single- cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia** *Nature Neuroscience* **24**:584–594

8. Gage F. H (2002) **Neurogenesis in the adult brain** *Journal of Neuroscience* **22**:612–613

9. Bandler R. C. *et al.* (2022) **Single-cell delineation of lineage and genetic identity in the mouse brain** *Nature* **601**:404–409

10. Delgado R. N., Allen D. E., Keefe M. G., Mancia Leon W. R., Ziffra R. S., Crouch E. E., Alvarez-Buylla A., Nowakowski T. J (2022) **Individual human cortical progenitors can produce excitatory and inhibitory neurons** *Nature* **601**:397–403

11. Rakic P (1988) **Specification of cerebral cortical areas** *Science* **241**:170–176

12. Ciceri G., et al. (2024) **An epigenetic barrier sets the timing of human neuronal maturation** *Nature* **626**:881–890

13. Finlay B. L., Darlington R. B (1995) **Linked regularities in the development and evolution of mammalian brains** *Science* **268**:1578–1584

14. Lodato M. A. *et al.* (2015) **Somatic mutation in single human neurons tracks developmental and transcriptional history** *Science* **350**:94–98

15. Breuss M. W. *et al.* (2022) **Somatic mosaicism reveals clonal distributions of neocortical development** *Nature* **604**:689–696

16. Chung C. *et al.* (2024) **Cell-type-resolved mosaicism reveals clonal dynamics of the human forebrain** *Nature* **629**:384–392

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

18. Pan Y. H., Wu N., Yuan X. B (2019) **Toward a better understanding of neuronal migration deficits in autism spectrum disorders** *Frontiers in Cell and Developmental Biology* **7**

19. Nguyen L. T., Schmidt H. A., von Haeseler A., Minh B. Q. (2015) **IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies** *Molecular Biology and Evolution* **32**:268–274

Editors

Reviewing Editor
Genevieve Konopka
University of Texas Southwestern Medical Center, Dallas, United States of America

Senior Editor
Timothy Behrens
University of Oxford, Oxford, United Kingdom

Reviewer #1 (Public review):

Summary:

In this manuscript, Shibata describes a method to assess rapidly fluctuating CpG sites (fCpGs) from single-cell methylation sequencing (sc-MeSeq) data. Assuming that fCpGs are largely consistent over time with changes induced by inheritable events during replication, the author infers lineage relationships in available brain-derived sc-MeSeq. Supplementing current lineage tracing through genomic and mitochondrial mosaic variants is an interesting concept that could supplement current work or allow additional lineage analysis in existing data.

However, the author failed to convincingly show the power of fCpG analysis to determine lineages in the human brain. While the correlation with cellular division and distinction of cell types appears plausible and strong, the application to detect specific lineages is less convincing. Aspects of this might be due to a lack of clarity in presentation and erroneous use of developmental concepts. However, without addressing these problems it is challenging for a reader to come to the same conclusions as the author.

On the flip side, this novel application of fCpGs will allow the re-use of existing sc-MeSeq to infer additional features that were previously unavailable, once the biological relevance has been further elucidated.

Strengths:

(1) Novel re-analysis application of methylation data to infer the status of fCpGs and the use as a lineage marker.

(2) Application of this method to an innovative existing data set to benchmark this framework against existing developmental knowledge.

Weaknesses:

(1) Insufficient clarity when presenting results (this includes an incredible shortness of the methods section making an informed assessment very difficult). This makes it hard to fully

grasp and evaluate the presented results.

(2) Inconsistent or erroneous use of neurodevelopmental concepts which hinders appropriate interpretation of the results.

(3) Lack of consideration for alternative explanations for the observed data (i.e., considering fCpGs as a cellular division clock that diverges over 'time').

https://doi.org/10.7554/eLife.101163.1.sa3

Reviewer #2 (Public review):

The manuscript by Shibata proposed a potentially interesting idea that variation in methylcytosine across cells can inform cellular lineage in a way similar to single nucleotide variants (SNVs). The work builds on the hypothesis that the "replication" of methylcytosine, presumably by DNMT1, is inaccurate and produces stochastic methylation variants that are inherited in a cellular lineage. Although this notion can be correct to some extent, it does not account for other mechanisms that modulate methylcytosines, such as active gain of methylation mediated by DNMT3A/B activity and activity demethylation mediated by TET activity. In some cases, it is known that the modulation of methylation is targeted by sequence-specific transcription factors. In other words, inaccurate DNMT1 activity is only one of the many potential ways that can lead to methylation variants, which fundamentally weakens the hypothesis that methylation variants can serve as a reliable lineage marker. With that being said (being skeptical of the fundamental hypothesis), I want to be as open-minded as possible and try to propose some specific analyses that might better convince me that the author is correct. However, I suspect that the concept of methylation-based lineage tracing cannot be validated without some kind of lineage tracing experiment, which has been successfully demonstrated for scRNA-seq profiling but not yet for methylation profiling (one example is Delgado et al., nature. 2022).

(1) The manuscript reported that fCpG sites are predominantly intergenic. The author should also score the overlap between fCpG sites and putative regulatory elements and report p-values. If fCpG sites commonly overlap with regulatory elements, that would increase the possibility that these sites being actively regulated by enhancer mechanisms other than maintenance methyltransferase activity.

(2) The overlap between fCpG and regulatory sequence is a major alternative explanation for many of the observations regarding the effectiveness of using fCpG sites to classify cell types correctly. One would expect the methylation level of thousands of enhancers to be quite effective in distinguishing cell types based on the published single-cell brain methylome works.

(3) The methylation level of fCpG sites is higher in hindbrain structures and lower in forebrain regions. This observation was interpreted as the hindbrain being the "root" of the methylation barcodes and, through "progressive demethylation" produced the methylation states in the forebrain. This interpretation does not match what is known about methylation dynamics in mammalian brains, in particular, there is no data supporting the process of "progressive demethylation". In fact, it is known that with the activation of DNMT3A during early postnatal development in mice or humans (Lister et al., 2013. Science), there is a global gain of methylation in both CH and CG contexts. This is part of the broader issue I see in this manuscript, which is that the model might be correct if "inaccurate mC replication" is the only force that drives methylation dynamics. But in reality, active enzymatic processes such as the activation of DNMT3A have a global impact on the methylome, and it is unclear if any signature for "inaccurate mC replication" survives the de novo methylation wave caused by DNMT3A activity.

(3) Perhaps one way the author could address comment 3 is to analyze methylome data across several developmental stages in the same brain region, to first establish that the signal of "inaccurate mC replication" is robust and does not get erased during early postnatal development when DNMT3A deposits a large amount of de novo methylation.

(4) The hypothesis that methylation barcodes are homogeneous among progenitor cells and more polymorphic in derived cells is an interesting one. However, in this study, the observation was likely an artifact caused by the more granular cell types in the brain stem, intermediate granularity in inhibitory cells, and highly continuous cell types in cortical excitatory cells. So, in other words, single-cell studies typically classify hindbrain cell types that are more homogenous, and cortical excitatory cells that are much more heterogeneous. The difference in cell type granularity across brain structures is documented in several whole-brain atlas papers such as Yao et al. 2023 Nature part of the BICCN paper package.

(5) As discussed in comment 2, the author needs to assess whether the successful classification of cell types (brain lineage) using fCpG was, in fact, driven by fCpG sites overlapping with cell-type specific regulatory elements.

(6) In Figure 5E, the author tried to address the question of whether methylation barcodes inform lineage or post-mitotic methylation remodeling. The Y-axis corresponds to distances in tSNE. However, tSNE involves non-linear scaling, and the distances cannot be interpreted as biological distances. PCA distances or other types of distances computed from high-dimensional data would be more appropriate.

https://doi.org/10.7554/eLife.101163.1.sa2

Reviewer #3 (Public review):

Summary:

In the manuscript entitled "Human Brain Barcodes", the author sought to use single-cell CpG methylation information to trace cell lineages in the human brain.

Strengths:

Tracing cell lineages in the human brain is important but technically challenging. Lineage tracing with single-cell CpG methylation would be interesting if convincing evidence exists.

Weaknesses:

As the author noted, "DNA methylation patterns are usually copied between cell division, but the replication errors are much higher compared to base replication". This unstable nature of CpG methylation would introduce significant problems in inferring the true cell lineage. The unreliable CpG methylation status also raises the question of what the "Barcodes" refer to in the title and across this study. Barcodes should be stable in principle and not dynamic across cell generations, as defined in Reference#1. It is not convincing that the "dynamic" CpG methylation fits the "barcodes" terminology. This problem is even more concerning in the last section of results, where CpG would fluctuate in post-mitotic cells.

https://doi.org/10.7554/eLife.101163.1.sa1

Author response:

I thank the Senior Editor and the three reviewers for their consideration and careful assessments, which I find fair and justified. I agree the evidence is inadequate that single cell

fluctuating CpG DNA methylation allows for human neuron lineage tracing. I agree with Reviewer #1 that fCpGs essentially function as “a cellular division clock that diverges over time”, but that fCpG methylation also records ancestry because cells with more similar patterns should be more related than cells with different patterns. However, as noted, there are alternative explanations that could explain fCpG DNA methylation pattern neuronal differences, or potentially obscure ancestry recorded by replication errors. Lineage tracing with fCpG methylation previously appeared possible in human intestines, endometrium, and blood, and potentially a similar approach could be used to reconstruct human brain cell ancestries.

I intend to revise the manuscript in a few weeks to address points raised by reviewers. These include a) editing to improve clarity and correct neurodevelopmental concepts, and b) adding a supplement that explains in much more detail how fCpG methylation may record cell divisions and ancestries. As recommended, additional “experiments” will be added including a) an analysis of single cell zygote to inner cell mass data to illustrate how fCpG brain barcode methylation changes between cell divisions very early in development before neurogenesis, and b) an analysis of newly released single cell brain aging data (Chien et al., 2024, Neuron 112, 2524–2539, August 7, 2024) that should help address issues of reproducibility and barcode stability over time. The evidence for lineage tracing will still be incomplete, but the modifications should help support the idea that fCpG methylation can record somatic cell ancestries.

https://doi.org/10.7554/eLife.101163.1.sa0