

Contrasting patterns of somatic mutations in neurons and glia reveal differential predisposition to disease in the aging human brain

Authors: Javier Ganz^{1,2,3,*}, Lovelace J. Luquette^{4,*}, Sara Bizzotto^{1,2,3,5*}, Craig L. Bohrson⁴, Hu Jin⁴, Michael B. Miller^{1,2,3,6}, Zinan Zhou^{1,2,3}, Alon Galor⁴, Peter J. Park^{4,7,#}, Christopher A. Walsh^{1,2,3,#}.

Affiliations:

¹ Division of Genetics and Genomics, Manton Center for Orphan Disease Research, Department of Pediatrics, and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA 02115, USA.

² Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA 02115, USA.

³ Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA.

⁴ Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA.

⁵ Sorbonne Université, Institut du Cerveau (Paris Brain Institute) ICM, Inserm, CNRS, Hôpital de la Pitié Salpêtrière, Paris, France.

⁶ Departments of Pathology and Neurology, Brigham and Women's Hospital, Boston, MA 02115, USA.

⁷ Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA.

* These authors contributed equally.

Correspondence to: christopher.walsh@childrens.harvard.edu; peter_park@hms.harvard.edu

SUMMARY

Characterizing the mechanisms of somatic mutations in the brain is important for understanding aging and disease, but little is known about the mutational patterns of different cell types. We performed whole-genome sequencing of 71 oligodendrocytes and 51 neurons from neurotypical individuals (0.4 to 104 years old) and identified >67,000 somatic single nucleotide variants (sSNVs) and small insertions and deletions (indels). While both cell types accumulate mutations with age, oligodendrocytes accumulate sSNVs 69% faster than neurons (27/year versus 16/year) whereas indels accumulate 42% slower (1.8/year versus 3.1/year). Correlation with single-cell RNA and chromatin accessibility from the same brains revealed that oligodendrocyte mutations are enriched in inactive genomic regions and are distributed similarly to mutations in brain cancers. In contrast, neuronal mutations are enriched in open, transcriptionally active chromatin. These patterns highlight differences in the mutagenic processes in glia and neurons and suggest cell type-specific, age-related contributions to neurodegeneration and oncogenesis.

KEYWORDS

Somatic mutations, glial cells, oligodendrocytes, oligodendrocyte precursor cells, aging, brain disorders, brain cancer, gliomagenesis, glioma.

1 INTRODUCTION

2 Somatic mutations accumulate in every tissue of the human body throughout life, via
3 mechanisms that depend on intrinsic tissue physiology and exogenous agents¹⁻⁸. Because human
4 tissues comprise diverse cell types with unique properties, quantifying cell type-specific rates
5 and mechanisms of somatic mutation is fundamental to understanding aging and disease
6 initiation at the tissue level. Although previous studies have addressed somatic mutations in
7 aging human neurons^{4,9-11}, mutations in glial cells—which represent more than half of the
8 cellular content of the brain and play primary roles in several brain disorders—have not yet been
9 examined. Abnormalities of white matter (WM), which consists mostly of glial cells, are
10 hallmarks of normal brain aging^{12,13} as well as neurodegenerative^{14,15} and neuropsychiatric
11 disorders¹⁶, but the causes of these changes are unknown. Furthermore, glial progenitor cells are
12 the cell-of-origin of many brain tumors¹⁷, and recent findings showed that WM in non-diseased
13 human brain is enriched with clonal oncogenic mutations compared to grey matter¹⁸.

14 Oligodendrocytes (OLs) are the main cell type of the WM, and OL dysfunction has been
15 reported in certain brain tumors, age-related disorders^{19,20}, psychiatric disorders^{21,22} and immune-
16 related multiple sclerosis²³. OL generation in humans begins during the second trimester of
17 gestation, peaks at birth and during the first years of life, and continues into adulthood, though at
18 reduced rates^{24,25}. Unlike neurons, which mostly arise before birth, OLs are replenished
19 throughout postnatal life by resident oligodendrocyte-precursor cells (OPCs)^{24,26}, with the rate of
20 replenishment diminishing with age²⁷. Dysregulation of proliferation and differentiation in the
21 OL lineage is involved in brain cancer, and OPCs are recognized as the cell of origin in some
22 gliomas²⁸⁻³¹. Thus, in contrast to neurons, OLs may be subject to mutational processes related to
23 DNA replication and can potentially undergo positive selection relevant for cancer insurgence³².

24 In this study, we assessed genome-wide rates and patterns of aging-related somatic
25 mutations in OLs compared to neurons isolated from the same individuals with single-cell
26 whole-genome sequencing (scWGS). In addition, we generated single-cell Assay for
27 Transposase-Accessible Chromatin with high-throughput sequencing (scATAC-seq) data from
28 these brains and integrated new as well as published⁸ single-cell RNA sequencing (scRNA-seq)
29 data obtained from individuals in the same cohort (Figure 1A). Joint analysis of these data
30 showed OL- and neuron-specific rates and patterns of somatic mutation accumulation, with DNA

31 replication and transcription playing significant roles in OL and neuronal mutagenesis,
32 respectively, and captured features of mutational processes in the differentiated OLs as well as in
33 precursor OPCs. The substantial differences in somatic mutation localization between these two
34 adjacent and interacting cell types informs cell type-specific contributions to distinct age-related
35 diseases.

36

37 **RESULTS**

38 **OLs accumulate somatic mutations at different rates than neurons**

39 OLs were isolated by antibody staining of nuclei prepared from post-mortem cortical
40 brain tissue, selecting SOX10-positive and NEUN-negative nuclei by fluorescence-activated
41 nuclear sorting (FANS). scRNA-seq performed on the sorted populations confirmed >99% purity
42 for both mature OLs and neurons sorted by NEUN positivity (Figure S1, **Methods**). Overall, 71
43 OLs were obtained from the prefrontal cortex (PFC) of 12 neurologically normal individuals
44 spanning 0.4 to 83 years of age (Table S1); 31 single-OL genomes were amplified by primary
45 template-directed amplification (PTA)^{10,33} and 40 were amplified by multiple-displacement
46 amplification (MDA) before PTA became available. Finally, we combined our single-OL
47 scWGS with 51 PTA-amplified neurons previously generated¹⁰ from 17 individuals, including 11
48 which overlap our OL cohort. Notably, due to the higher rate of technical artifacts caused by
49 MDA¹⁰, we focused on PTA-amplified samples except where indicated. Following scWGS,
50 somatic single-nucleotide variants (sSNVs) and small (1-30 base pair (bp)) insertions/deletions
51 (indels) were identified genome-wide using SCAN2, an algorithm we developed recently¹⁰
52 (Table S2). To focus on somatic mutations acquired during aging rather than development, high
53 allele frequency clonal sSNVs and indels were excluded by removing somatic calls supported by
54 one or more reads in matched 30-45X bulk DNA sequencing.

55 Compared to neurons, scWGS of OLs revealed higher yearly rates of sSNV accumulation
56 but lower rates of indel accumulation. As is the case with neurons and many other cell types<sup>3,4,9-
57 11,34</sup>, the increase in OL sSNV burdens was remarkably linear with respect to age, with a rate of
58 27 sSNVs/year (95% CI: 25.0-28.9), which is significantly greater than the neuronal rate of 16
59 sSNVs/year (CI: 14.9-17.4, Figure 1B; for the difference, $P = 9.6 \times 10^{-17}$, t -test for coefficients in
60 a linear mixed model (LMM t -test), see **Methods**). At birth, OLs contained 60% more sSNVs
61 per genome compared to neurons (intercept: 191 vs 119), though this difference was not

62 significant ($P = 0.15$, LMM t -test). Unlike sSNVs, indels accumulated more slowly in OLs than
63 in neurons (1.8 (CI: 1.56-2.07) versus 3.1 (CI: 2.61-3.61) indels/year, respectively, $P = 1.42 \times$
64 10^{-5} , LMM t -test, Figure 1B). Indel burdens at birth were comparable between cell types.
65 Deletions were more prevalent than insertions in both cell types, consistent with previous
66 reports^{10,35} (Figure S2A); however, OL indels were mostly single-bp deletions, while neurons
67 carried greater numbers of 2-4bp deletions and 1bp insertions (Figure S2B), likely representing
68 distinct mechanisms of indel generation.

69 OL and neuronal mutations showed opposite biases for genic regions, suggesting
70 different mechanisms of mutagenesis and different consequences for gene integrity. OL sSNVs
71 were enriched in intergenic regions (13%, $P < 10^{-4}$; all P -values for enrichment analyses based
72 on permutation tests, see **Methods**) and depleted in introns and coding regions (15.4%, $P < 10^{-4}$
73 and 19.7%, $P = 0.0024$, respectively) (Figure 1C). This pattern was replicated in MDA-amplified
74 OLs (Figure S2C, see **Methods**). Neuronal sSNVs were instead overrepresented in genes
75 (12.7%, $P = 0.0047$ in exons; 5.2%, $P < 10^{-4}$ in introns) and depleted in intergenic regions (5.4%,
76 $P < 10^{-4}$). Indels mirrored these patterns but often with greater effect sizes: in genes, OL indels
77 were depleted (34.8%, $P = 0.0656$ in exons; 4.9%, $P = 0.17$ in introns) whereas neuronal indels
78 were strongly enriched, especially in exons (78.6%, $P < 10^{-4}$ in exons; 19.3%, $P < 10^{-4}$ in
79 introns), as previously reported¹⁰. In general, a larger fraction of neuronal mutations were
80 predicted (by SnpEff³⁶, see **Methods**) to functionally impact genes (Figure 1D). Strikingly, the
81 rate of indels with the most severe gene-altering effects was ~3-fold higher in neurons than in
82 OLs. Due to the small number of mutations overall, no significant mutation enrichment or
83 depletion was detected for any individual genes after correction for multiple hypothesis testing
84 (Figure S2D).

85

86 **OL somatic mutations produce signatures associated with cell proliferation**

87 Analysis of mutational spectra and signatures indicated shared and cell type-specific
88 mutational mechanisms in OLs and neurons. The spectrum of OL sSNVs closely matched the
89 spectra of highly proliferative hematopoietic stem and progenitor cells (HSPCs, cosine similarity
90 0.946),^{9,34,37,38} whereas neurons were only marginally similar to HSPCs (cosine similarity 0.773,
91 Figure 2A). Surprisingly, the OL spectrum was less similar to neurons (cosine similarity 0.887)
92 than to HSPCs, likely reflecting a contribution of cell proliferation to OL somatic mutagenesis.

93 To explore mutagenic mechanisms, we identified single base substitution (SBS)
94 mutational signatures for the identified sSNVs using the COSMIC catalog (v3.1)³⁹. Ten of the
95 COSMIC SBS signatures passed our thresholds for activity in either OLs or neurons (Figure 2B,
96 **Methods**) and the remainder were removed to reduce overfitting. Signatures SBS5 and SBS89
97 were the most prevalent signatures in both cell types, and both signatures accumulated at higher
98 rates in OLs compared to neurons (SBS5, 10.2 versus 6.95 sSNVs/year, $P = 0.014$; SBS89, 5.5
99 versus 1.95, $P = 6 \times 10^{-11}$, Figure 2B-C, LMM t -test). SBS5 is a clock-like signature that
100 accumulates independently of cell division, whereas SBS89 was recently reported in colon
101 crypts³ but has no known etiology. Signatures SBS1 and SBS32 were strongly associated with
102 age in OLs ($P < 10^4$ for both signatures, LMM t -test) but were nearly absent in neurons ($P = 0.17$
103 and $P = 0.55$, LMM t -test). SBS32 is a C>T signature recently reported as a component of the
104 HSPC spectrum³⁴ and SBS1 is a clock-like signature associated with cell division⁴⁰ and
105 accumulated at rates of 2.63 sSNVs/year and 0.20 sSNVs/year in OLs and neurons, respectively
106 (Figure 2B-C). Since mature OLs are post-mitotic, SBS1 may be generated primarily during the
107 OPC stage; if true, the linear accumulation of SBS1 with age in mature OLs implies surprisingly
108 constant levels of OPC division during adult life. Only SBS16, a signature associated with
109 transcription, accumulated at a higher rate in neurons, with 1.59 sSNVs/year compared to 0.38
110 sSNVs/year in OLs, consistent with the enrichment of neuronal mutations in transcribed genomic
111 regions and in line with previous reports^{4,10}.

112 Two pairs of closely related OLs, which likely trace their ancestry to common OPCs,
113 illustrate the extent to which signatures of early- and late-life mutations differ, and represent
114 permanent maps of developmental lineages and age-related changes. Despite filtering high allele
115 frequency clonal sSNVs by removing mutations found in bulk tissues, two OL pairs from two
116 individuals (subjects 5559 (PTA OLs) and 5657 (MDA OLs), 19.8y and 82y, respectively)
117 shared unusually high levels of sSNVs (68 and 64 sSNVs, respectively, Figure 2D), indicating
118 common ancestry. In both subjects, the shared sSNVs were mostly C>T transitions at CpG sites
119 and fitting to the COSMIC signatures revealed a 38% contribution from the cell-division-related
120 signature SBS1. Each pair of OLs also contained similar numbers of private sSNVs (179 and 173
121 for subject 5559; 897 and 1017 for subject 5657), consistent with equal lifetimes for each of the
122 cognate OLs after the division of their most recent common ancestor (MRCA) OPC. The
123 mutational spectrum of private sSNVs was similar to the OL spectrum (Figure 2A,D) and, after

124 fitting to COSMIC and removing MDA-associated artifacts (**Methods**), the spectrum was
125 primarily explained by SBS89 and SBS5 (34% and 20%), followed by SBS12, SBS1 and SBS32
126 (15%, 9% and 9%, respectively, Figure 2D). We estimated the age at which the MRCA divided
127 in each subject by adjusting private sSNV counts for calling sensitivity and dividing by our
128 yearly sSNV accumulation rate (see **Methods**). This placed the MRCA of subject 5559 near
129 birth and the MRCA of subject 5657 between 0-20 years of age. Comparison of mutation spectra
130 with infant OLs (subjects aged 0-2 years old) provided orthogonal evidence that the MRCA
131 occurred during infancy. Indeed, the spectrum of shared sSNVs matched the spectrum of infant
132 OL sSNVs with a cosine similarity of 0.97 (Figure 2E). Crucially, the neuronal sSNV spectrum
133 from the same infant subjects contained far fewer SBS1-like C>Ts at CpG dinucleotides,
134 implying that the increase in SBS1 occurred in an OL-specific lineage and does not reflect early
135 clonal sSNVs that evaded our filters. We speculate that the most likely MRCA time coincides
136 with a burst of OL generation that occurs in the young human brain (0-10 years of age)²⁴. The
137 relationships of these two pairs of cells suggest that shared mutations mark a permanent forensic
138 lineage tree, while non-shared mutations represent a linear timer of when any two cells separate
139 from a common progenitor.

140 Indel (ID) signatures revealed shared and specific mutational processes further
141 distinguishing OLs from neurons (Figure 3A-C). ID4, a signature representing ≥ 2 bp deletions
142 and associated with transcriptional mutagenesis⁴¹, was most strongly correlated with age in
143 neurons, as previously reported^{9,10}, but was almost completely absent in OLs ($P = 0.534$, LMM
144 t -test; Figure 3C). ID5 and ID8, two clock-like indel signatures, were present in both cell types,
145 with ID8 correlated more strongly with age in neurons than in OLs. The two remaining clock-
146 like indel COSMIC signatures, ID1 and ID2, were not detected, but they are difficult to identify
147 in PTA data due to similarity with sequencing artifacts¹⁰. ID9, which is characterized by 1bp
148 deletions, was the most prevalent signature in OLs and accumulated at a rate of 0.54 indels/year;
149 in neurons, the accumulation was significantly lower at 0.20 indels/year ($P = 0.0017$, LMM t -
150 test). Interestingly, this ID9 signature is also found in a large fraction of adult gliomas⁴² as well
151 as in a considerable fraction of other brain tumors³⁹.

152

153 **OL SNVs are enriched in inactive genomic regions**

154 Our earlier observation that OL mutations were depleted in genes—opposite to the
155 pattern in neurons (Figure 1C)—suggested different determinants of mutagenesis in these two
156 cell types. Comparison of somatic mutation density to additional data types, including scRNA-
157 seq, scATAC-seq, replication timing, and chromatin marks, revealed that OL mutations are
158 enriched in chromatin that is either inaccessible, untranscribed, or which harbors repressive
159 histone marks—which we refer to as inactive chromatin—in striking contrast to neuronal
160 mutations. We first compared somatic mutation densities to gene expression levels from brain
161 scRNA-seq data for three subjects in our cohort (1465, 4638 and 4643; 40,083 PFC cells in total;
162 Figure 4A)⁸. OL sSNVs were depleted by 25%-32% in the top few deciles of expression
163 measured in OLs (Figure 4B, $P < 10^{-4}$; all P -values in this section are from permutation tests)
164 and were similarly depleted in all other cell types detected in our scRNA-seq data. The negative
165 association between transcription level and somatic mutation density in OLs was confirmed
166 using bulk brain RNA-seq data from the Genotype Tissue Expression Consortium (GTEx)⁴³
167 (Figure S3A). Indels in OLs were not significantly enriched or depleted, possibly due to a lack of
168 statistical power caused by the relatively low number of somatic indels in OLs (Figure 4B).

169 Next, brain scATAC-seq data representing ~82,000 cortical cells obtained from ten
170 subjects in our cohort (see **Methods**) revealed a strong depletion of OL sSNVs in open
171 chromatin (Figure 4C,D). In the decile of the genome with the highest chromatin accessibility
172 from OLs identified in scATAC-seq data, OL sSNVs were depleted by 29% ($P < 10^{-4}$). Slightly
173 weaker sSNV depletions were observed for the remaining cell types (mean 21% for the top
174 decile of chromatin accessibility), with OPCs showing the second strongest depletion signal
175 (24%, $P < 10^{-4}$). A weak but negative trend between OL indel density and chromatin
176 accessibility was also observed (Figure 4D).

177 Data from the Encyclopedia of DNA Elements (ENCODE)⁴⁴ and the Roadmap
178 Epigenomics Project⁴⁵ further confirmed enrichment of OL mutations in inactive chromatin.
179 First, OL sSNVs were significantly enriched in late-replicating regions of the genome (which
180 tend to be less transcriptionally active) as determined by RepliSeq data from the ENCODE
181 project (mean 39% in the latest replicated decile, $P < 10^{-4}$; Figure 4E, S3B). Comparison to
182 histone marks from the Roadmap Epigenomics Project revealed negative associations between
183 OL sSNVs and marks of open chromatin, transcription and active regulatory elements
184 (H3K27ac, H3K36me3, H3K4me1, H3K4me3 and H3K9ac) and positive associations with

185 repressive marks (H3K9me3 and H3K27me3)⁴⁶ (Figure 4F-G). Chromatin state annotations from
186 ChromHMM⁴⁶, which classify chromatin based on an ensemble of histone marks, further
187 confirmed the pattern of OL mutation enrichment in inactive or inaccessible genomic regions,
188 with OL sSNVs overrepresented in heterochromatin (state 9, 24% enrichment, $P = 0.0036$) and
189 quiescent regions (state 15, 11% enrichment, $P < 10^{-4}$), and depleted in the transcriptionally
190 active states 1-7 (Figure 4H). The strongest depletion of OL sSNVs across all genomic
191 covariates analyzed in this study was observed for active transcription start sites (ChromHMM
192 state 1, mean 50.4% depletion across brain tissues) and was not specific to any of the 13
193 Roadmap Epigenomics brain tissues (Figure S3C). An orthogonal dataset of active promoters in
194 neurons, oligodendrocytes, microglia and astrocytes from flow-sorted cell populations⁴⁷ further
195 confirmed the strong depletion of OL sSNVs in promoters (mean depletion 51%) and again there
196 was no marked preference for the cell type from which the promoters were measured (Figure 4I).

197 The distribution of neuronal mutations differed from OLs across all the genomic
198 covariates we tested: neuronal sSNV and indel rates increased with gene expression, chromatin
199 accessibility and active histone modifications and decreased with inactive histone modifications
200 (Figure 4A-G). Unlike OLs, somatic mutations in neurons were more specifically associated with
201 transcription levels measured in brain tissues (Figure S3A) and especially with single cell
202 transcriptomic and chromatin accessibility signals from neurons (Figure 4B,D). Neuronal
203 mutations showed little association with replication timing (Figure 4E), which is unsurprising
204 since most neuronal mutations are acquired in the post-mitotic state and clonal somatic mutations
205 were largely removed by our bulk filters.

206 To further understand the action of mutational processes in the two cell types, we
207 correlated SBS mutation signature exposures (rather than total mutation density) to the
208 previously discussed genomic covariates. In OLs, SBS1 density generally followed the patterns
209 of total mutation density, with positive associations with inactive chromatin and late replication
210 timing (Figure 5A). The distribution of SBS1 in neurons mimicked that of OLs and, in particular,
211 was strongly positively associated with replication timing, suggesting that neuronal SBS1 may
212 have accumulated during cell divisions in neurogenesis. SBS16 (a T>C signature associated with
213 transcriptional activity) exposure in neurons was positively associated with active histone marks,
214 gene expression, and chromatin accessibility levels from excitatory and inhibitory neurons
215 (Figure 5B). Consistent with the known transcribed-strand bias of SBS16, neuronal T>C

216 mutations exhibited the largest transcribed-strand bias of all mutation types in this study (Figure
217 S4). Interestingly, despite neurons being post-mitotic, SBS16 density trended negatively with
218 replication timing, likely reflecting higher gene density in early replicating regions.

219 Although SBS5 is the most prevalent signature in both OLs and neurons, it did not
220 accumulate in the same genomic regions in these two cell types, particularly with respect to
221 expression levels (Figure 5C). In OLs, patterns of SBS5 exposure showed little difference from
222 the aggregate somatic mutation density, with negative associations with active epigenetic marks,
223 gene expression and open chromatin and positive associations with inactive marks and late
224 replicating regions. However, in neurons, SBS5 was not significantly associated with any of the
225 covariates tested, with only a marginally significant positive trend with scRNA-seq expression in
226 excitatory neurons (top 33%, $P = 0.08$, Figure 5C). These observations suggest that either SBS5
227 is generated by cell type-specific mechanisms or that SBS5 may not be a fully decomposed
228 signature—in particular, it may be contaminated by the transcription-associated SBS16,
229 consistent with the marginally significant association with expression levels in neurons—as
230 previously suggested⁹.

231

232 **The OL mutation density profile resembles that of glial-derived tumors**

233 Patterns of somatic mutation in cancer often contain sufficient information to identify the
234 cell type from which a tumor emerged⁴⁸; thus, we explored whether our normal OL sSNV
235 densities resembled those from a large collection of cancer WGS data from the Pan-Cancer
236 Analysis of Whole Genomes (PCAWG) project⁴². OL sSNVs were positively correlated with
237 somatic mutation densities of all cancer types from PCAWG whereas neuronal sSNVs were not
238 correlated with any tumor type (Figure 6A). Specifically, for OL mutations, the highest
239 correlations observed corresponded to glial-derived tumors of the central nervous system (CNS)
240 for which OPCs are thought to be the cell of origin (CNS-GBM, glioblastoma multiforme and
241 CNS-Oligo, oligodendroglioma)²⁸⁻³¹.

242 Our scATAC-seq data allowed further cell type-specific evaluation of cancer sSNV
243 densities. Among all tumor types in PCAWG, GBM sSNV density was best predicted by OPC-
244 specific scATAC-seq tracks using a regression model, with 44% of variance in GBM sSNV
245 density explained (Figure 6B). This provides additional evidence that OPCs are the cell of origin
246 for GBM tumors and that scATAC-seq is a powerful approach for determining the cell-of-origin

247 for a tumor⁴⁸. Expression levels from scRNA-seq were far less effective in explaining cancer
248 mutation density, explaining only 6% of variance in the best case (Figure 6C).

249 Finally, we tested whether cancer-associated genes were more likely to be mutated in
250 OLs compared to neurons. For each tumor type, we determined the 100 most frequently mutated
251 genes and computed an odds ratio (OR) indicating whether mutations in OLs (OR > 1), neurons
252 (OR < 1) or neither cell type (OR=1) were more likely to occur in the frequently mutated genes.
253 In general, OL sSNVs were biased toward cancer-associated genes (OR near 1.1 for most cancer
254 types, Figure 6D), likely reflecting the general similarities between OL and cancer sSNV
255 mutation densities. However, OL sSNVs were clearly biased toward genes mutated in CNS
256 tumors, with the highest odds ratios observed for oligodendrogliomas (CNS-Oligo, OR=1.35, P
257 = 1.6×10^{-7} , Fisher's exact test) and pilocytic astrocytomas (CNS-PiloAstro, OR=1.31, P
258 = 3.1×10^{-6}). Analysis of the top N cancer mutated gene lists for $N = 10$ to 500 confirmed that these
259 findings did not depend on our choice of cutoff $N = 100$ (Figure S5). Altogether, the similarities
260 between OL and cancer mutation patterns suggest a contributing relationship between OL
261 mutations—especially those acquired at the OPC stage—and tumorigenesis.

262

263 **DISCUSSION**

264 Our integrative analysis of somatic mutations uncovered OL-specific mutational
265 processes during aging compared to neurons and suggests how these differences may predispose
266 to diseases such as cancer. Our study design provides a unique opportunity to explore how
267 different cell types sharing the same microenvironment for years—or even decades—can exhibit
268 contrasting features that ultimately shape the human brain. An additional advantage of our design
269 is that comparison of OLs and neurons measured by the same single-cell DNA sequencing
270 technology helps to rule out the possibility that differential mutation rates or genomic
271 distributions reflect technical artifacts or biased representation of specific genomic regions.

272 Somatic mutation burdens increase linearly in both OLs and neurons with age; however,
273 OLs accumulate 69% more sSNVs than neurons and 42% fewer indels. The apparent low indel
274 rate in OLs may reflect a high rate of indel mutagenesis in neurons compared to other cell types,
275 as reported by previous studies⁹. Some of the excess sSNVs in OLs (e.g., those attributed to
276 SBS1) are likely associated with cell division in ancestral OPCs⁴⁹, but due to lacking information
277 regarding the mechanisms of SBS5, SBS32 and SBS89 mutagenesis, it is not clear what

278 processes account for the remaining excess sSNV burden in OLs. A recent study highlighted the
279 importance of cell proliferation-independent sources of somatic mutations in normal cells and
280 hypothesized that the interplay between cell type-specific DNA damage and repair processes
281 may underlie differences in mutation burden between cell types⁹. Hence, less efficient DNA
282 repair processes in OLs—rather than additional DNA damage—may be a plausible explanation
283 for the excess OL sSNV burden compared to neurons. Follow-up studies mapping DNA repair
284 sites in OLs versus neurons might be needed to address this question⁵⁰⁻⁵². Another important
285 factor differentiating mutations in OLs and neurons is the strength of selective forces acting
286 during aging. Aging neurons cannot be subject to positive selection, and negative selection is
287 likely limited to highly deleterious mutations that induce cell death. OLs, on the other hand,
288 include mutations accumulated by postnatal cycling OPCs in which both positive and negative
289 selective effects may be present. Thus, the subset of OL somatic mutations acquired at the
290 ancestral OPC stage are of particular importance since they can expand clonally and amplify
291 deleterious effects.

292 OL mutations are more prevalent in transcriptionally inactive and/or inaccessible
293 chromatin and resemble patterns reported in cancer³² and other proliferative cells⁴⁸. Associations
294 between OL somatic mutation density and genomic covariates generally were not cell type- or
295 tissue-specific; however, it is important to interpret this observation in light of the fact that OL
296 mutations represent a combination of mutations accrued during the OPC and mature OL states.
297 Neuronal mutations were characterized by a strongly contrasting pattern of enrichment in
298 transcriptionally active, open chromatin with clear preference for genomic covariates measured
299 in brain tissue and, specifically, excitatory neurons.

300 Mutational signature analysis was helpful in interpreting some contributing factors to the
301 overall mutational burden and to its accumulation over time. SBS1 was prevalent in OLs and
302 nearly absent in neurons, consistent with previous characterizations of SBS1 as a cell division-
303 dependent mutational clock⁴⁰ and at odds with a recent study which estimated a nearly 10-fold
304 greater SBS1 rate in human neurons⁹. SBS5 made up the majority of mutations in both OLs and
305 neurons, but was distributed differently across the genome in the two cell types. One attractive
306 explanation for this is differential repair: i.e., that SBS5-associated DNA damage may occur
307 throughout the genome but be more efficiently repaired in certain genomic regions in a cell type-
308 specific manner. However, concerns of contamination by other signatures and other algorithmic

309 issues may yet be to blame; for example, the current COSMIC catalog was generated primarily
310 by cancer exomes and genomes, and signatures present in post-mitotic cells are likely to be
311 under-represented. In addition, despite the dozens of single cells we sequenced, the total number
312 of mutations is not large enough to confidently identify signatures that are present at low
313 exposures.

314 Our work demonstrates marked differences in somatic mutation accumulation between
315 neurons and OLs in the same tissue and reveals the mutational dynamics of OLs during
316 neurotypical brain aging. In contrast to neuronal mutations, OL mutations mimic features of
317 somatic mutations in cancers of the CNS, including (1) OL mutational signatures³⁹ and (2) the
318 distribution of mutations across the genome, particularly for tumors for which OPCs are believed
319 to be the cell of origin (GBM and oligodendroglioma). These observations emphasize the
320 contribution of mutations that occurred during the OPC stage to the mutation landscape of
321 mature OLs. Furthermore, the differing genomic regions enriched for somatic mutations in OLs
322 versus neurons suggests that cell type-specific mutation distributions may contribute to cell type-
323 specific predispositions to particular pathologies.

324

325 **ACKNOWLEDGMENTS**

326 We thank J. Neil and the University of Maryland Brain Bank for helping with the tissue
327 collection. We thank the donors and their families for their invaluable donations for the
328 advancement of science. We thank the Boston Children’s Hospital IDDRC Molecular Genetics
329 Core Facility, supported by NIH award U54HD090255 from the National Institute of Child
330 Health and Human Development. **Funding:** Supported by R01AG070921 and R01AG078929
331 from the NIA to CAW. CAW is an Investigator of the Howard Hughes Medical Institute. JG was
332 supported by a Basic Research Fellowship from the American Brain Tumor Association
333 BRF1900016 and by the Brain SPORE grant P50CA165952. S.B was supported by the Manton
334 Center for Orphan Disease Research at Boston Children’s Hospital, and is now supported by the
335 Horizon2020 Research and Innovation Program Marie Skłodowska-Curie Actions (MSCA)
336 Individual Fellowship (grant agreement no. 101026484 — CODICES). ZZ was supported by the
337 PRMRP Discovery Award W81XWH2010028 and Edward R. and Anne G. Lefler Center
338 Postdoctoral Fellowship.

339

340 **AUTHOR CONTRIBUTIONS**

341 J.G, S.B and L.J.L conceived the study; J.G and S.B performed all the experiments; L.J.L led
342 bioinformatic analyses helped by C.L.B H.J and A.G; J.G, and S.B contributed to bioinformatic
343 analysis interpretation; S.B contributed to scRNA-seq data analyses; M.M and Z.Z contributed
344 neuronal PTA data; C.A.W. and P.J.P. directed the research; J.G and L.J.L wrote the manuscript
345 greatly helped by S.B.

346

347 **COMPETING INTERESTS**

348 The authors declare no competing financial and/or non-financial interests.

349

350 **INCLUSION AND DIVERSITY**

351 One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in
352 their field of research or within their geographical location.

353

354

355 **FIGURE LEGENDS**

356

357 **Figure 1. Oligodendrocytes and neurons exhibit contrasting patterns of somatic mutation**
358 **accumulation.**

359 (A) Experimental strategy. Oligodendrocytes (OL; n=31 PTA, n=40 MDA) and neurons (n=51
360 PTA) were obtained from the brains of 17 neurotypical individuals (0-104 years of age) through
361 FANS using NEUN (neurons) and SOX10 (OL) antibodies. Single genomes were amplified
362 using PTA or MDA and non-clonal sSNVs and indels were called using SCAN2. Mutation
363 distributions were compared with scATAC-seq and scRNA-seq data obtained from a subset of
364 the 17 individuals.

365 (B) Extrapolated genome-wide sSNV and indel burdens for OLs and neurons as a function of
366 age. Trend lines are mixed-effects linear regression models (see **Methods**).

367 (C) Distribution of OL and neuronal sSNVs and indels in genic and intergenic regions.
368 Enrichment/depletion levels are calculated by comparison to a null distribution obtained by
369 randomly shuffling mutations across the genome (see **Methods**).

370 (D) Percent of somatic mutations with HIGH, MODERATE and LOW impact on genes, as
371 determined by SnpEff.

372 See also Figure S1 and Figure S2.

373

374 **Figure 2. Comparison of single base substitution (SBS) mutational signatures in human**
375 **oligodendrocytes and neurons.**

376 (A) SBS mutational spectra of neuronal and oligodendrocyte sSNVs identified in this study (left
377 column); the spectrum of hematopoietic stem and progenitor cells (HSPCs) identified in Lee-Six
378 et al.³⁷ and a signature derived from an analysis of human lymphocytes (Machado et al.³⁸).

379 Cosine similarities are shown for each pair of spectra.

380 (B) The number of somatic mutations, after extrapolation to genome-wide burdens, attributed to
381 each COSMIC SBS signature by least squares fitting for each PTA single OL and neuron.

382 Subjects are ordered from young (left) to elderly (right). A reduced COSMIC catalog of 10
383 signatures with evidence of activity in either OLs or neurons was used to minimize overfitting
384 (see **Methods**).

385 (C) Same as (B), but plotted against age. Each point represents one single cell. Trend lines are
386 mixed-effects linear regression models (see **Methods**).

387 (D) Schematic of two pairs of related OLs and estimates of the time of division for each pair's
388 most recent common ancestor (MRCA). Insets show the SBS spectrum and contributions of
389 COSMIC signatures for sSNVs acquired before division of the MRCA (shared, top left) and
390 sSNVs acquired after division of the MRCA (private, top right). Spectra in panels (D) and (E)
391 are reconstructions obtained by summing exposure-weighted signatures after fitting the infant
392 mutations to the COSMIC SBS catalog (see **Methods**).

393 (E) Mutational spectra of all mutations from infant (0-2 years of age) subjects in PTA neurons
394 (left) and PTA OLs (right). The similarity in spectra between infant OL mutations and shared OL
395 mutations supports the estimated time of most recent common ancestor (MRCA) division.

396

397

398 **Figure 3. Comparison of insertion and deletion (ID) COSMIC signatures in human**
399 **oligodendrocytes and neurons.**

400 (A) Spectra of somatic indels from human OLs and neurons using the 83-dimensional indel
401 classification scheme from COSMIC.

402 (B) Contribution of COSMIC indel signatures to each single OL and neuron. One bar represents
403 one single cell; cells are ordered according to age with the youngest individuals on the left and
404 eldest individuals on the right.

405 (C) Same as (B), but signature exposure is plotted against age for each single cell; each point
406 represents one cell. Trend lines are mixed-effects linear regression models (see **Methods**). ID5
407 and ID8 are annotated as clock-like signatures in COSMIC.

408

409 **Figure 4. Oligodendrocyte somatic mutations are associated with inactive chromatin while**
410 **neuronal mutations associate with active chromatin.**

411 (A) Representative UMAP plot of scRNA-seq from one normal brain from our cohort (subject
412 1465) with cell type annotations.

413 (B) Enrichment analysis of somatic mutations vs. scRNA-seq transcription level. The genome is
414 divided into 1 kilobase, non-overlapping windows and each window is annotated with an average
415 gene expression level per cell type; windows that are <80% covered by a gene are discarded. The
416 remaining windows are classified into 10 deciles, with 1 representing the least transcribed and 10
417 representing the most transcribed. In each decile, the observed number of somatic SNVs and
418 indels is compared to a null distribution of mutations obtained by randomly shuffling mutation
419 positions (see **Methods**). Enrichment analyses show somatic mutation density vs. transcription
420 level from each of the cell types identified in our scRNA-seq.

421 (C-D) Same as (A) for scATAC-seq from the brains of 10 subjects from this cohort.

422 (E) Somatic mutation density vs. replication timing as measured by ENCODE RepliSeq; lines
423 represent average enrichment across 15 cell lines.

424 (F-G) Somatic mutation density vs. 5 epigenetic marks related to gene activity (F) and two
425 repressive epigenetic marks (G) measured in dorsolateral prefrontal cortex tissue (Roadmap
426 Epigenomic Project, reference epigenome E073).

427 (H) Enrichment of somatic mutations vs. functional genomic regions identified by ChromHMM
428 in reference epigenome E073. Numbers in parenthesis indicate the ChromHMM state number.

429 (I) Enrichment of somatic mutations vs. active enhancers and promoters identified in Nott et al.
430 ⁴⁷ for several brain cell types.

431 See also Figure S3.

432

433

434 **Figure 5. Enrichment of mutational signatures in active and inactive chromatin.**

435 Enrichment analysis of somatic mutations attributed to SBS1 (A), SBS16 (B) or SBS5 (C)—
436 rather than total mutation density—vs. the genomic covariates presented in Figure 4. To reduce
437 noise caused by the smaller number of mutations attributed to each specific mutational signature,
438 the genome has been divided into three quantiles rather than ten. OLs are not plotted for SBS16
439 due to noisy measurements; none of these omitted OL SBS16 enrichment tests achieved
440 significance at the $P < 0.05$ level.

441 See also Figure S4.

442

443 **Figure 6. Patterns of oligodendrocyte sSNVs correlate with somatic mutation density in**
444 **cancer.**

445 (A) Correlation of OL and neuronal sSNV mutation density with cancer mutation density. For
446 each cell type and cancer type, the genome was tiled with non-overlapping 1 MB bins and
447 numbers of mutations per bin were tabulated. CNS tumors are colored: CNS-Oligo,
448 oligodendroglioma, red; CNS-PiloAstro, pilocytic astrocytoma, purple; CNS-GBM, glioblastoma
449 multiforme, orange; CNS-Medullo, medulloblastoma, black.

450 (B-C) Mutation density for each tumor type was fit using a linear regression to cell type-specific
451 single cell chromatin accessibility signals (B) and single cell expression levels (C) using the
452 same 1 MB bins described in panel (A). For each tumor type and cell type, the fraction of
453 variance in tumor mutation density explained (R^2) by each cell type is shown.

454 (D) Comparison of OL and neuron somatic mutation rates in frequently mutated cancer genes.
455 For each tumor type in PCAWG (y-axis), the 100 most frequently mutated genes were
456 determined. For each tumor-specific set of 100 cancer genes (G_T) an odds ratio (OR) is computed
457 such that $OR > 1$ indicates that OL mutations are more likely to occur in G_T and $OR < 1$ indicates
458 neuronal mutations are more likely to occur in G_T . Formally, $OR = (\# \text{ OL sSNVs in } G_T / \# \text{ genic}$
459 $\text{ OL sSNVs not in } G_T) / (\# \text{ neuron sSNVs in } G_T / \# \text{ genic neuron sSNVs not in } G_T)$.

460 See also Figure S5.

461

462

463 **METHODS**

464

465 **Data and code availability**

466 Newly generated sequencing data for MDA and PTA single-cell WGS of
467 oligodendrocytes, single-cell ATAC-seq, and single-cell RNA-seq are available on dbGaP,
468 accession number phs001485.v3.p1. Previously generated scWGS, matched bulk controls,
469 and scRNA-seq data for individual 1465 are also available for download at dbGaP, accession
470 number phs001485.v3.p1. Scripts used for analyses in this manuscript are available on
471 Zenodo (<https://doi.org/10.5281/zenodo.7508802>).

472

473 **Human tissues and DNA samples**

474 All human tissues were obtained from the NIH NeuroBioBank at the University of
475 Maryland. Frozen post-mortem tissues from 12 neurologically normal individuals were obtained
476 as part of a previous study⁴. Samples were processed according to a standardized protocol
477 (<http://medschool.umaryland.edu/btbank/method2.asp>) under the supervision of the NIH
478 NeuroBioBank ethics guidelines. Bulk DNA was extracted using the QIAamp DNA Mini kit
479 with RNase A treatment.

480

481 **Nuclear sorting and whole-genome amplification**

482 Isolation of single nuclei using fluorescence-activated nuclear sorting (FANS) for NEUN
483 and SOX10 was performed using a modified version of a previously described protocol^{53,54}.
484 Briefly, nuclei were prepared by dissecting fresh-frozen human brain tissue previously stored at -
485 80°C, dissolved on ice in chilled nuclear lysis buffer (10mM Tris-HCl, 0.32M Sucrose, 3mM
486 MgAc₂, 5mM CaCl₂, 0.1mM EDTA, pH 8, 1mM DTT, 0.1% Triton X-100) using a Dounce
487 homogenizer. Lysates were layered on top of a sucrose cushion buffer (1.8M Sucrose, 3mM
488 MgAc₂, 10mM Tris-HCl, pH 8, 1mM DTT) and ultra-centrifuged for 1 hour at 30,000rcf. Pellets
489 containing nuclei were resuspended in ice-cold PBS 1X supplemented with 3mM MgCl₂, then
490 filtered, blocked in PBS 1X supplemented with 3mM MgCl₂ and 3% Bovine Serum Albumin
491 (blocking solution), and stained with an anti-NEUN antibody (Millipore MAB377) previously
492 used for neuronal nuclei isolation^{4,53}, anti-SOX10 antibody (Novus NBP2-59621R), and DAPI.
493 Other antibodies targeting the OL population were also evaluated, KLK6 (Bioss bs-5870R) and
494 CNP (Bioss bs-1000R). Nuclei were washed once with blocking solution, centrifuged at 500rpm
495 for 5 minutes and resuspended again in blocking solution. Nuclear sorting was performed in one
496 nucleus per well in 96-well plates and whole-genome amplification was performed using
497 Primary Template-directed Amplification (PTA) following manufacturer guidelines. Libraries
498 for sequencing were generated using the KAPA HyperPlus kit (Roche) using dual indexes and
499 were sequenced across 5 lanes of Illumina NovaSeq6000 (2x150bp), targeting 20x coverage
500 (75Gbp)/sample.

501

502 **Sorting purity assessment**

503 Three populations were sorted, DAPI, NEUN+, SOX10+/NEUN-, from one 17y
504 individual (Supp. Fig. S1). 10x scRNAseq was used and 6,000-10,000 nuclei from each

505 population were sorted into individual tubes containing RT mastermix, and immediately
506 processed for GEM generation, barcoding, and cDNA amplification, following manufacturer
507 instructions. Each library prep was submitted to paired-end single indexing sequencing on a lane
508 of Illumina HiSeqX to obtain ~50,000 read pairs per nucleus. Analysis of scRNA seq data was
509 performed with Loupe Cell Browser 4.q software provided by 10x genomics. Sorting purity is
510 critical when performing single-cell whole-genome (scWGS) studies, hence we evaluated purity
511 of a larger number of cells by scRNA-seq. As a control, evaluation of 3,447 DAPI+ sorted nuclei
512 obtained from a mix of grey and white matter (PFC) from a 17yr male identified all the cell-
513 types anticipated for this region and with significant presence of OLs following WM
514 expectation⁵⁵. The purity of 3,739 NEUN+ sorted nuclei was nearly 100% (Supp. Fig. S1A-C)
515 with ~1% (40 out of 3,739 nuclei) presence of a PLP1/MBP/MOG+ OL population, and 0.1% (3
516 out of 3,739 nuclei) of NOSTRIN+ endothelial cells. NEUN sorted nuclei can be broadly
517 classified into 60% excitatory and 40% inhibitory neurons consistent with recent reports of
518 excitatory/inhibitory ratios⁵⁶. Evaluation of 9,227 SOX10+/NEUN- sorted nuclei confirmed
519 100% mature OLs purity, with the absence of other cell-type markers expression (Supp. Figure
520 S1A-C). The SOX10+/NEUN- sorted nuclei showed homogenous distribution of classic mature
521 OL-markers such as PLP1, MOG, MALAT1, among others. Although SOX10 is expressed in all
522 stages of OL development, including in OPCs, our strategy consistently recovered only mature
523 OLs.

524

525 **10x scRNA-seq**

526 scRNA-seq was performed using the 10X Genomics Chromium Next GEM Single Cell
527 3 \square Reagent Kit v3.1. Fresh frozen human brain tissue from the prefrontal cortex of individuals
528 UMB1465, UMB4638 and UMB4643 was processed to obtain nuclear pellets. Briefly, tissue
529 was dissociated on ice in chilled nuclear lysis buffer (10 mM Tris-HCl, 0.32 M Sucrose, 3 mM
530 MgAc₂, 5 mM CaCl₂, 0.1 mM EDTA, pH 8, 1 mM DTT, 0.1% Triton X-100) using a dounce
531 homogenizer. Homogenates were layered on top of a sucrose cushion buffer (1.8 M Sucrose, 3
532 mM MgAc₂, 10 mM Tris-HCl, pH 8, 1 mM DTT) and ultra-centrifuged for 1 hour at 30,000 rcf.
533 Pellets containing nuclei were resuspended in 250 μ l ice-cold 1X PBS supplemented with 3 mM
534 MgCl₂, 3% Bovine Serum Albumin (BSA) and 0.2 U/ μ l RNase inhibitor (Thermo Fisher
535 Scientific ref.10777019), then filtered. After filtering, suspension volume was completed at 1 ml

536 using the same solution, and nuclei were stained with DAPI before sorting to select for intact
537 nuclei. Some of the UMB1465 samples were additionally stained with the following antibodies:
538 two samples with anti-NEUN antibody (Millipore MAB377) for neuron sorting, one sample each
539 for anti-CX43/GJA1 (Novus Biologicals, FAB7737R-1 00UG AF647), anti-SOX9 (Abcam,
540 ab196450 AF488) and anti-GFAP (Millipore, MAB3402 AF647) to enrich for glial cells, and
541 one sample with anti-SOX10 (Novus Biologicals, NBP2-59621 AF647) for oligodendrocyte
542 sorting. 10,000 to 15,000 single nuclei were sorted for each experiment directly in a tube
543 containing the 10X RT mastermix, and immediately processed for gel-bead in emulsion (GEM)
544 generation, barcoding, cDNA amplification and library preparation following manufacturer
545 instructions. Each library preparation was submitted for paired-end single indexing sequencing
546 on Illumina HiSeqX or NovaSeq6000 targeting ~50,000 read pairs per nucleus.

547

548 **10x scRNA-seq data analysis**

549 scRNA-seq data were demultiplexed using `bcl2fastq`. scRNA-seq FASTQ files were
550 then processed using the 10X Genomics `cellranger count` pipeline for gene expression to
551 perform alignment to hg19, barcode counting, UMI counting, and generation of feature-barcode
552 matrices. Cell Ranger filtered count matrices were used for downstream analysis using Seurat
553 3.0⁵⁷. For each library, we further filtered for cells with > 200 and < 3000 genes and <5%
554 mitochondrial genes, and genes with <10,000 UMI counts and >3 cells. RNA counts were
555 normalized using the LogNormalize method and the 2,000 most highly variable features were
556 identified using the `vst` method. Data were then scaled by regressing out the percentage of
557 mitochondrial genes. We then performed non-linear dimensional reduction and clustering. To
558 remove doublets from our datasets, we ran `DoubletFinder`⁵⁸ using optimal parameters as per the
559 `paramSweep` function. Finally, cell-type identities were assigned to each cluster in the Uniform
560 Manifold Approximation and Projection (UMAP) based on expression of known brain cell-type
561 markers⁸.

562

563 **10x scATAC-seq**

564 Nuclei from 10 individuals (1278, 1465, 4638, 4643, 5087, 5219, 5559, 5817, 5823,
565 5871) from our aging cohort were obtained from the same brain region as used for single cell

566 whole-genome amplification. Tissue was processed as described in nuclear sorting, and nuclei
567 were re-suspended in diluted nuclei buffer provided by the manufacturer. Nuclei derived from
568 different individuals were processed for transposition separately, before loading to the 10x
569 Chromium Controller for GEM generation, barcoding, and library construction, as per
570 manufacturer instructions. Libraries were submitted for paired-end dual index sequencing on one
571 flow cell of Illumina S2 NovaSeq6000 (100 cycles) to obtain ~50,000 reads per nucleus.

572

573 **10x scATAC-seq data analysis**

574 Sequencing data were demultiplexed using `bcl2fastq` and `mkfastq`.
575 `cellranger-atac count v1.1.0` was run separately on the resulting FASTQ files for each
576 scATAC-seq library (one per individual) with default parameters and the vendor-provided hg19
577 reference. Results from the individual library analyses (Cell Ranger output files
578 `fragments.tsv.gz` and `singlecell.csv` from each library) were then merged by
579 `cellranger-atac aggr --normalize-depth`. scATAC-seq data were analyzed by
580 Signac v1.1.0 and Seurat v3 following the authors' instructions. Briefly, the merged Cell Ranger
581 output was imported via `Read10X_h5` and `CreateChromatinAssay`; analyzed by
582 `RunTFIDF`, `FindTopFeatures`, `RunSVD` and `RunUMAP` with LSI reduction; and integrated
583 with our scRNA-seq to assign cell types via `GeneActivity`, `FindTransferAnchors` and
584 `TransferData`.

585

586 **Single neuron whole genome sequencing data**

587 Sequencing data for 36 PTA-amplified single neurons and matched bulk from the 12
588 individuals from which OLs were harvested were downloaded from dbGaP accession
589 `phs001485.v3.p1`. These BAM files were re-analyzed by SCAN2 jointly with OLs as described
590 in *Somatic mutation calling with SCAN2*. For a single additional individual (5171), no PTA data
591 were generated for either neurons or OLs; in this one case, sequencing data for MDA-amplified
592 single neurons from the prefrontal cortex and a matched bulk were downloaded from the same
593 dbGaP accession as the PTA data (see Table S1). SCAN2 somatic SNV calls, indel calls, and
594 genome-wide burden estimates from 15 previously published PTA-amplified single neurons
595 from 5 additional neurotypical individuals (individuals 4976, 5451, 5572, 5666, and 5943) were
596 downloaded from https://github.com/parklab/SCAN2_PTA_paper_2022.

597 **Somatic mutation calling with SCAN2**

598 SCAN2 v1.1 was run on the 12 individuals from which PTA OLs were collected and
599 individual 5171 (MDA only). First, a cross-sample panel (required for indel calling with
600 SCAN2) was built for the 12 individuals with PTA neurons and OLs by running `scan2`
601 `config` with parameters `--analysis makepanel --gatk=gatk3_joint`; the
602 GRCh37 human reference genome with decoy `hs37d5` (`--ref`), dbSNP v147 common (`--`
603 `dbSNP`) and 1000 Genomes phase 3 SHAPEIT2 phasing panel (`--shapeit-refpanel`) as
604 described in ref. 10; one `--sc-bam` argument for each PTA BAM; and one `--bam` argument
605 for each of the 12 matched bulk BAMs. After generation of the cross-sample panel, SCAN2 v1.1
606 was run for each of the 12 individuals and individual 5171. For each individual, all PTA OLs
607 (and MDA OLs for individuals 1278, 5871, 5171, and 5657) and PTA neurons (MDA neurons
608 for individual 5171) were analyzed together in a single SCAN2 run. `scan2 config` was run
609 with the parameters `--analysis=call_mutations --gatk=gatk3_joint --`
610 `abmodel-samples-per-step=20000 --abmodel-refine-steps=4 --`
611 `abmodel-n-cores=10` and the same GRCh37 reference, dbSNP and phasing panels used
612 above. The cross-sample panel created above was specified via the `--cross-sample-panel`
613 parameter. Finally, SCAN2 mutation signature-based rescue was run in two batches, one for
614 PTA neuron and one for PTA OL calls, using `scan2 config --analysis=rescue --`
615 `rescue-target-fdr=0.01` followed by `scan2 rescue`. Mutation signature-based
616 rescue was not run on MDA neurons or OLs. For neurons only, VAF-based SCAN2 calls (i.e.,
617 excluding signature-rescued SCAN2 calls) from the 15 previously analyzed PTA single neurons
618 (see *Single neuron whole genome sequencing data*) were added to the rescue process via `--`
619 `add-muts`.

620 **Total mutation burden estimation and aging models**

621 SCAN2 provides estimates of the total somatic SNV and indel burden in each sample
622 (i.e., the estimated total number of mutations per cell adjusted for sensitivity of mutation calling).
623 These estimates were obtained from each SCAN2 output RDA file by first loading the file in R,
624 then running `object@mutburden[["snv"]]$burden[2]`. Indel burdens were recovered
625 by replacing "snv" with "indel". To account for variability within and between individuals,
626 mixed-effects linear models were used to estimate the aging-related rates of somatic SNV and

627 indel accumulation. These models were fit by the R `lme4` package using
628 `lmer(genome.burden ~ age*celltype + (1|individual))`, where `celltype`
629 is either `oligo` or `neuron`, `individual` is one of the 12 individual IDs for which PTA OLS
630 were available (i.e., excluding individual 5171) and `age` is the numeric age of `individual`.
631 Confidence intervals were estimated by `confint`. For linear mixed models, statistical tests of
632 significance comparing each coefficient, interaction term and intercept to a null hypothesis of 0
633 were calculated by the `lmerTest` R package, which uses a *t*-test based on the Satterthwaite
634 approximation. Throughout the text, these *t*-tests are referred to as LMM *t*-tests.

635 **SnpEff annotation**

636 Both VAF-based and mutation signature-rescued SCAN2 somatic mutation calls were
637 annotated for functional impact via SnpEff version 4.3t using the hg19 database. Reported
638 functional impacts were taken from the first ANN field in the SnpEff annotated VCF. When
639 computing the fraction of mutations with LOW, MODERATE, and HIGH impact (Figure 1D),
640 duplicate and clustered mutations were removed as described in *Somatic mutation enrichment*
641 *analysis*.

643 **Somatic mutation enrichment analysis**

644 Enrichment analyses following the methodology described in ref. 10 were carried out to
645 determine whether somatic mutations accumulate preferentially in certain areas of the genome.
646 First, somatic mutations were filtered to remove duplicates and clusters of mutations in single
647 cells. For exact duplicate mutations (i.e., same position and base change or indel), if the duplicate
648 mutations all occur in cells from a single individual (suggesting a clonal mutation), a single
649 representative mutation was retained; otherwise, if the duplicate mutations span multiple
650 individuals (suggesting artifacts), all mutations were discarded. Clusters of mutations, defined as
651 any mutation within 50 bp of another mutation in the same single cell, suggest underlying
652 alignment artifacts or structural variants and were removed before enrichment analysis.
653 Duplicate and clustered mutations were determined separately for MDA and PTA mutation calls.
654 For MDA OLS only, all SCAN2 mutation calls from the 20 OLS from infant brains were
655 additionally filtered prior to duplicate and cluster filtering. These were removed because the
656 mutation burden of young OLS is too small to sufficiently outnumber MDA technical artifacts.

657 The only enrichment analysis using MDA OL calls is the genic region enrichment analysis in
658 Figure S2C.

659 Next, the genome was divided into regions based on a single covariate (described in more
660 detail below) and the number of somatic mutations in each region R was compared to the number
661 of mutations in R after the locations of the mutations were randomly permuted across the
662 genome. The permutation process was repeated 10,000 times and the final enrichment value E_R
663 was given by dividing the number of observed somatic mutations in each region R by the average
664 number of permuted mutations in R . For enrichment analysis of mutational signatures, the above
665 steps were followed, except: (1) somatic mutations called by SCAN2's mutation signature-based
666 rescue method were not used and (2) rather than counting the number of mutations in each region
667 R , the set of mutations in R was fit to the reduced COSMIC SBS or ID catalog (see *COSMIC*
668 *signature catalog activity thresholds reduction*) by non-negative least squares (using
669 `lsqnonneg` from the `pracma` R package) and the exposure value for each signature was used
670 in lieu of mutation counts. To obtain two-sided tests of enrichment significance for each region
671 R , a permutation test strategy was used. Enrichment values $E_R^{(i)}$ were computed for each of the
672 10,000 permutations i and a P -value was derived from the fraction of permutation sets with more
673 extreme enrichment values than the observed mutations. To avoid P -values of 0, a minimum of P
674 = 0.0001 was enforced, i.e.,

$$P_R = \max\left(1/10000, \frac{|\{i: |\log E_R^{(i)}| > |\log E_R|\}|}{10,000}\right).$$

675 To define regions R for enrichment analysis, a subset of the genome with anomalous
676 sequencing depth was first identified and removed from subsequent analyses. The human
677 reference genome GRCh37 with decoy sequences hs37d5 was divided into non-overlapping
678 windows of 100 bp and the average sequencing depth across all PTA neurons and OLs as output
679 by SCAN2 (file path:
680 `path/to/scan2_output/depth_profile/joint_depth_matrix.tab.gz`) was
681 computed for each 100 bp window. Windows with low average depth (<6 reads averaged across
682 all PTA cells) or excessive average depth (in the top 2.5% of average depth) were classified as
683 anomalous. Genomic regions were next derived from non-quantitative genomic covariates
684 (genes, Figure S2D; exons, introns, and intergenic spaces, Figure 1C; ChromHMM classes,

685 Figure 4F; and cell type-specific promoters and enhancers, Figure 4G) and quantitative
686 covariates (GTEx transcription levels, Figure S3A; scRNA-seq transcription levels, scATAC-seq
687 accessibility, RepliSeq replication timing and histone mark levels, Figures 4B,D-G). For non-
688 quantitative covariates, regions R were defined by the union of genomic intervals for each unique
689 covariate state (e.g., all exons or all regions annotated as ChromHMM state 1) and anomalous
690 windows were subtracted from these unions. For quantitative covariates, the genome was first
691 tiled with non-overlapping 1 kbp windows (corresponding to 10 100 bp windows from the
692 anomalous sequencing depth analysis). 1 kbp windows containing >2 anomalous sequencing
693 depth windows were discarded. For each remaining 1 kbp window i , a single quantitative value
694 V_i was derived for each covariate in a covariate-dependent manner (described in detail for each
695 covariate below). The distribution of values V_i were then discretized into $n = 10$ (for enrichment
696 analysis of total mutation burden) or $n = 3$ quantiles (for enrichment analysis of mutation
697 signatures) and each window i was assigned its quantile rank Q_i . Finally, a region R_Q was defined
698 for each quantile $Q = 1 \dots n$ by taking the union of windows with rank Q . Genomic regions and
699 the resulting enrichment analyses were always performed using one covariate at a time.

700 **Genomic covariates for enrichment analysis**

701 *Gene regions.* GENCODE genes version 26 was downloaded from
702 [https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/releas](https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_26/GRCh37_map)
703 [e_26/GRCh37_map](https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_26/GRCh37_map)
704 [ping/gencode.v26lift37.annotation.gtf.gz](https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_26/GRCh37_map). A single region was defined for
705 each protein coding gene using GTEx's transcript collapse script
706 ([https://raw.githubusercontent.com/broadinstitute/gtex-](https://raw.githubusercontent.com/broadinstitute/gtex-pipeline/master/gene_model/collapse_annotation.py)
707 [pipeline/master/gene_model/collapse_annotation.py](https://raw.githubusercontent.com/broadinstitute/gtex-pipeline/master/gene_model/collapse_annotation.py)) and used for the per-
708 gene enrichment analysis in Figure S2D. Exonic, intronic and intergenic regions were then
709 derived from the GENCODE GTF records for the same set of genes. When multiple records
710 overlapped a single locus, status was prioritized as follows: CDS > UTR > exon > intron >
711 upstream/downstream > intergenic. Finally, loci classified as CDS, UTR or exon were classified
712 as exonic, intron as intronic, and all other classes as intergenic.

713 *ChromHMM chromatin states.* 15-state ChromHMM annotations were downloaded for
714 epigenome ID E073 (dorsolateral prefrontal cortex) from
715 <https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentat>

716 ions/ChmmModels/coreMarks/jointModel/final/E073_15_coreMarks_mne
717 monics.bed.gz.

718 *Promoters and enhancers for flow sorted brain cell types*. Active promoter and enhancer
719 elements were extracted from Supplementary Table 5 of ref. 58 available at
720 https://www.science.org/doi/suppl/10.1126/science.aay0793/suppl_
721 [file/aay0793-nott-table-s5.xlsx](https://www.science.org/doi/suppl/10.1126/science.aay0793/suppl_file/aay0793-nott-table-s5.xlsx). Duplicate lines in these tables were removed
722 prior to analysis.

723 *GTEX transcription levels*. Median gene expression levels from 54 tissue types were downloaded
724 from the GTEX project at
725 [https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTE](https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEX_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz)
726 [x_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz](https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEX_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz). For
727 each tissue type, the median transcription level of each gene G was mapped to the genome by
728 applying it over G 's collapsed transcript (see *Gene regions* above). Only 1 kbp genomic
729 windows with $\geq 80\%$ coverage by gene transcripts were retained. If multiple genes overlapped
730 the same window, the maximum TPM value was used for the window.

731 *scRNA-seq transcription levels*. Cell type-annotated gene-expression matrices for each scRNA-
732 seq library were concatenated column-wise and average expression levels for each gene were
733 calculated for each cell type. Gene names were then matched to the GTEX gene model and
734 transcription levels for each cell type were mapped to the genome as described above for GTEX
735 transcription levels.

736 *scATAC-seq accessibility*. ATAC-seq transposition events output by cellranger-atac (file:
737 `fragments.tsv.gz`) were first separated by cell type (see *10x scATAC-seq data analysis*)
738 and then converted to BED format. The BED file of fragments for each cell type was then
739 converted to bedGraph format using `bedtools genomecov -bga` and finally to bigWig
740 format by `bedGraphToBigWig`. The bigWig signal files were then mapped to the 1 kbp
741 genome tiles described in *Somatic mutation enrichment analysis* by
742 `bigWigAverageOverBed`.

743 *Replication timing*. WaveSignal RepliSeq bigWigs were downloaded from
744 [http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncod](http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeUwRepliSeq/wgEncodeUwRepliseq{cell_line}WaeSignalRep1.bigWig)
745 [eUwRepliSeq/wgEncodeUwRepliseq{cell_line}WaeSignalRep1.bigWig](http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeUwRepliSeq/wgEncodeUwRepliseq{cell_line}WaeSignalRep1.bigWig) for
746 15 cell_lines (all available at the time of writing): BG02ES, BJ, GM06990, GM12801,

747 GM12812, GM12813, GM12878, HUVEC, HeLa-S3, HepG2, IMR90, K562, MCF-7, NHEK
748 and SK-N-SH. The bigWig signal files were then mapped to the 1 kbp genome tiles described in
749 *Somatic mutation enrichment analysis* by `bigWigAverageOverBed`; quantile values were
750 then reversed so that $Q = 1$ corresponded to the earliest replication timing quantile.
751 *Histone marks.* bigWig signal files representing ChIP-seq fold-change versus a no-IP control
752 were downloaded for 13 Roadmap reference epigenomes annotated as brain tissue from
753 [https://egg2.wustl.edu/roadmap/data/byFileType/signal/consolidat](https://egg2.wustl.edu/roadmap/data/byFileType/signal/consolidated/mac2signal/foldChange/{epigenome_ID}-{histone_mark}.fc.signal.bigwig)
754 [ed/mac2signal/foldChange/{epigenome_ID}-](https://egg2.wustl.edu/roadmap/data/byFileType/signal/consolidated/mac2signal/foldChange/{epigenome_ID}-{histone_mark}.fc.signal.bigwig)
755 [{histone_mark}.fc.signal.bigwig](https://egg2.wustl.edu/roadmap/data/byFileType/signal/consolidated/mac2signal/foldChange/{epigenome_ID}-{histone_mark}.fc.signal.bigwig) for 7 histone_mark values H3K27ac,
756 H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9ac and H3K9me3. The bigWig signal
757 files were then mapped to the 1 kbp genome tiles described in *Somatic mutation enrichment*
758 *analysis* by `bigWigAverageOverBed`.

759 **COSMIC signature catalog activity thresholds and reduction**

760 To reduce the possibility of overfitting by signatures that may not be active in our cell
761 types, all signature fits were conducted on a subset of the COSMIC v3.1 SBS and ID catalogs.
762 The catalogs were reduced by several heuristics designed to remove signatures with low
763 likelihood of activity in our data. The first heuristic is a step-forward procedure in which each
764 step i determines the single COSMIC signature that minimizes the residual (`resid.norm`) of
765 the non-negative least-squares fit (`lsqnonneg`) to the (signature channel) \times (sample) input
766 matrix compared to the signatures present in step $(i - 1)$. At each step, this reduction in residual
767 is compared to the residual using 0 signatures (i.e., the norm of the input matrix) to determine the
768 percent reduction of the signature selected at each step i . The second heuristic is designed to find
769 evidence of age-related accumulation of each signature. First, the (signature channel) \times (sample)
770 input matrix is fit by `lsqnonneg` to the entire COSMIC SBS or ID catalog to determine
771 exposure levels $E_{i,j}$ for each sample i and signature j . For each signature j , a linear model $E_{i,j} \sim$
772 age_i is fit and the significance of the test for slope=0 (t -test) is recorded. Finally, the significance
773 values for all signatures tested are corrected for multiple hypothesis testing by R's `p.adjust`
774 method using Holm's correction. The third heuristic measures the percent contribution P_j of each
775 signature j to the total spectrum; i.e., $P_j = \sum_i E_{i,j} / \sum_{i,j} E_{i,j}$. These three heuristics are computed
776 separately for neurons and OLs using VAF-based SCAN2 somatic mutation calls (to avoid

777 signature biases of signature-rescued SCAN2 calls). For each set of heuristics, signature j is
778 retained if either: heuristic (1) exceeds a 50% reduction in the residual; or if heuristic (2) < 0.01
779 after multiple hypothesis correction and heuristic (3) exceeds 1% of the total spectrum. The final
780 set of retained signatures is the union of signatures retained in either neurons or OLs.

781 **Analysis of related oligodendrocyte pairs**

782 Shared somatic SNVs were determined for each pair of single cells within each
783 individual. A shared sSNV was defined as a SCAN2 call present in at least one of the two cells
784 and for which 2 or more mutation supporting reads appear in the other cell. Private mutations
785 were SCAN2 calls that: (1) did not meet the shared mutation criteria and (2) were supported by 0
786 reads and a total depth of 6 or greater in the paired cell. These heuristics identified two pairs of
787 oligodendrocytes with exceptionally high numbers of shared sSNVs (5559-Oligo-5 and 5559-
788 Oligo-8, PTA-amplified OLs from individual 5559; 5657-GliaLC-4-F11 and GliaLC-4-G10,
789 MDA-amplified OLs from individual 5657).

790 Because the pair of OLs from individual 5657 were MDA-amplified, it was necessary to
791 remove artifacts associated with MDA, which follow a specific SBS signature termed Signature
792 B¹⁰. For signature analysis, removal of this signature was achieved by adding Signature B to the
793 reduced COSMIC catalog before fitting by `lsqnonneg` and then discarding exposures to
794 Signature B. Although PTA induces a much lower artifact burden than MDA, a similar
795 correction was performed to remove PTA artifacts from the spectra presented in Figure 2D,E: the
796 PTA artifact signature was added to the reduced COSMIC catalog, mutations were fit by
797 `lsqnonneg` and finally the reconstructed spectrum was created by summing COSMIC
798 signatures, weighted by their exposures, and excluding the PTA artifact weights.

799 To estimate time to the most recent common ancestor (MRCA) split, several corrections
800 to account for sensitivity and shared vs. private misclassification were required. To quantify
801 shared vs. private misclassifications, we applied the shared and private criteria for somatic SNVs
802 to germline heterozygous SNPs (hSNPs) known to be present in the individual, which should
803 always be shared between single cells from the same individual. The number of shared sSNVs
804 was then multiplied by $(1 + \text{fraction of hSNPs classified as private})$. The number of private
805 sSNVs was then reduced by the amount added to the shared sSNV count. The opposite error, a
806 private sSNV classified as shared, should occur rarely since it requires a random artifact to

807 intersect with a true mutation; we thus assumed this rate to be approximately 0. For the pair of
808 MDA OLs, the number of private sSNVs was reduced by the fraction of the private sSNV
809 spectrum attributed to the MDA artifact Signature B after fitting to the COSMIC SBS catalog
810 with Signature B. SCAN2's estimate of the total mutation burden (which accounts for mutation
811 calling sensitivity) was also reduced by the Signature B exposure fraction and SCAN2's calling
812 sensitivity re-estimated by $S = (\# \text{ SCAN2 calls}) / \text{adjusted total burden}$. The final corrected
813 shared sSNV and private sSNV counts reported in Figure 2D were computed by dividing by S .
814 The time to MRCA was estimated by counting mutations forward from the initial zygote (i.e.,
815 using the number of corrected shared sSNVs) and backward (i.e., using the number of corrected
816 private sSNVs). In each case, the corrected number of mutations was converted to a time in years
817 by first subtracting the intercept of our OL sSNV aging model and then dividing by the slope of
818 the OL aging model.

819 **Cancer mutation density**

820 PCAWG cancer somatic SNV and indel mutation catalogs were obtained in MAF format from
821 the ICGC portal
822 (https://dcc.icgc.org/releases/PCAWG/consensus_snv_indel).
823 Hypermuted tumors within each tumor type were identified by Tukey's method and mutations
824 from these tumors were removed. Next, cancer mutations for each sample were mapped to the
825 100 bp windows used for detecting anomalous sequencing depth (described in *Somatic mutation*
826 *enrichment analysis*) and the count of mutations in each window was then normalized by the
827 total number of mutations in that sample. Finally, a track representing the mutation density for
828 each tumor type was created by summing normalized window counts across samples from the
829 same tumor type and written in bigWig format via `rtracklayer's export .bw` function.

830 Because 100 bp or 1 kbp windows contain too few mutations for meaningful correlation
831 analyses with our neuron and OL somatic mutations, the per-tumor bigWig signal files with 100
832 bp resolution were mapped to a non-overlapping 1 Mbp genome tiling by
833 `bigWigAverageOverBed`. 1 Mbp windows with anomalous sequencing depth were then
834 removed following the same requirements for the 1 kbp tiling windows described in *Somatic*
835 *mutation enrichment analysis*. Total neuron and OL mutation counts were also determined over
836 these 1 Mbp windows and correlations between somatic mutation density in OLs, neurons and
837 each tumor type were shown in Figure 6A. A similar downscaling of signals from 1 kbp

838 resolution to 1 Mbp resolution via `bigWigAverageOverBed` was required for comparing
839 cancer mutation densities to scRNA-seq and scATAC-seq signals.

840

841 **Cancer gene odds ratio analysis**

842 Since the sizes of individual genes are usually too small for detecting mutation
843 enrichment or depletion with our catalog of somatic mutations from OLs and neurons, we
844 created a larger genomic region by considering sets of genes. PCAWG mutations (without
845 hypermutated samples, as described in *Cancer mutation density*) were mapped to genes by
846 SnpEff as described above. For each tumor type, a list of genes ordered by number of somatic
847 mutations intersecting the gene was constructed. For each tumor type, 50 progressively larger
848 genomic regions corresponding to the union of the top 10, 20, ..., 500 genes were created. In
849 each region R , the rates of OL mutations and neuron mutations impacting the region were
850 compared using the odds ratio

$$OR_R = \frac{\# \text{ oligo mutations in } R / \# \text{ oligo mutations not in } R}{\# \text{ neuron mutations in } R / \# \text{ neuron mutations not in } R}$$

851 Thus, $OR_R > 1$ implies a preference for OL mutations in the genes represented in R and $OR_R < 1$
852 implies a preference for neuronal mutations in R . Figure 6D presents odds ratios for the top 100
853 genes in each tumor type and the full analysis is presented in Figure S5.

854

855 **BIBLIOGRAPHY**

- 856 1. Genovese, G., Kahler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F.,
857 Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and
858 blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 371, 2477-2487.
859 10.1056/NEJMoa1409405.
- 860 2. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley,
861 R.C., Mermel, C.H., Burt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis
862 associated with adverse outcomes. *N Engl J Med* 371, 2488-2498.
863 10.1056/NEJMoa1408617.
- 864 3. Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R.J., Sanders, M.A., Moore, L.,
865 Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M., et al. (2019). The landscape
866 of somatic mutation in normal colorectal epithelial cells. *Nature* 574, 532-537.
867 10.1038/s41586-019-1672-7.
- 868 4. Lodato, M.A., Rodin, R.E., Bohrson, C.L., Coulter, M.E., Barton, A.R., Kwon, M., Sherman,
869 M.A., Vitzthum, C.M., Luquette, L.J., Yandava, C.N., et al. (2018). Aging and
870 neurodegeneration are associated with increased mutations in single human neurons.
871 *Science* (New York, N.Y) 359, 555-559. 10.1126/science.aao4426.
- 872 5. Martincorena, I., Fowler, J.C., Wabik, A., Lawson, A.R.J., Abascal, F., Hall, M.W.J., Cagan,
873 A., Murai, K., Mahbubani, K., Stratton, M.R., et al. (2018). Somatic mutant clones
874 colonize the human esophagus with age. *Science* 362, 911-917.
875 10.1126/science.aau3879.
- 876 6. Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge,
877 D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High
878 burden and pervasive positive selection of somatic mutations in normal human skin.
879 *Science* 348, 880-886. 10.1126/science.aaa6806.
- 880 7. Moore, L., Leongamornlert, D., Coorens, T.H.H., Sanders, M.A., Ellis, P., D'Entropio, S.C.,
881 Dawson, K.J., Butler, T., Rahbari, R., Mitchell, T.J., et al. (2020). The mutational
882 landscape of normal human endometrial epithelium. *Nature* 580, 640-646.
883 10.1038/s41586-020-2214-z.
- 884 8. Bizzotto, S., Dou, Y., Ganz, J., Doan, R.N., Kwon, M., Bohrson, C.L., Kim, S.N., Bae, T.,
885 Abyzov, A., Network, N.B.S.M., et al. (2021). Landmarks of human embryonic
886 development inscribed in somatic mutations. *Science* 371, 1249-1253.
887 10.1126/science.abe1544.
- 888 9. Abascal, F., Harvey, L.M.R., Mitchell, E., Lawson, A.R.J., Lensing, S.V., Ellis, P., Russell,
889 A.J.C., Alcantara, R.E., Baez-Ortega, A., Wang, Y., et al. (2021). Somatic mutation
890 landscapes at single-molecule resolution. *Nature*. 10.1038/s41586-021-03477-4.
- 891 10. Luquette, L.J., Miller, M.B., Zhou, Z., Bohrson, C.L., Zhao, Y., Jin, H., Gulhan, D., Ganz, J.,
892 Bizzotto, S., Kirkham, S., et al. (2022). Single-cell genome sequencing of human neurons
893 identifies somatic point mutation and indel enrichment in regulatory elements. *Nat*
894 *Genet* 54, 1564-1571. 10.1038/s41588-022-01180-2.

- 895 11. Miller, M.B., Huang, A.Y., Kim, J., Zhou, Z., Kirkham, S.L., Maury, E.A., Ziegenfuss, J.S.,
896 Reed, H.C., Neil, J.E., Rento, L., et al. (2022). Somatic genomic changes in single
897 Alzheimer's disease neurons. *Nature* 604, 714-722. 10.1038/s41586-022-04640-1.
- 898 12. Gunning-Dixon, F.M., Brickman, A.M., Cheng, J.C., and Alexopoulos, G.S. (2009). Aging of
899 cerebral white matter: a review of MRI findings. *Int J Geriatr Psychiatry* 24, 109-117.
900 10.1002/gps.2087.
- 901 13. Liu, H., Yang, Y., Xia, Y., Zhu, W., Leak, R.K., Wei, Z., Wang, J., and Hu, X. (2017). Aging of
902 cerebral white matter. *Ageing Res Rev* 34, 64-76. 10.1016/j.arr.2016.11.006.
- 903 14. Lee, S.H., Kim, S.S., Tae, W.S., Lee, S.Y., Choi, J.W., Koh, S.B., and Kwon, D.Y. (2011).
904 Regional volume analysis of the Parkinson disease brain in early disease stage: gray
905 matter, white matter, striatum, and thalamus. *AJNR Am J Neuroradiol* 32, 682-687.
906 10.3174/ajnr.A2372.
- 907 15. Salat, D.H., Greve, D.N., Pacheco, J.L., Quinn, B.T., Helmer, K.G., Buckner, R.L., and Fischl,
908 B. (2009). Regional white matter volume differences in nondemented aging and
909 Alzheimer's disease. *Neuroimage* 44, 1247-1258. 10.1016/j.neuroimage.2008.10.030.
- 910 16. Koshiyama, D., Fukunaga, M., Okada, N., Morita, K., Nemoto, K., Usui, K., Yamamori, H.,
911 Yasuda, Y., Fujimoto, M., Kudo, N., et al. (2020). White matter microstructural
912 alterations across four major psychiatric disorders: mega-analysis study in 2937
913 individuals. *Mol Psychiatry* 25, 883-895. 10.1038/s41380-019-0553-7.
- 914 17. Laug, D., Glasgow, S.M., and Deneen, B. (2018). A glial blueprint for gliomagenesis. *Nat*
915 *Rev Neurosci* 19, 393-403. 10.1038/s41583-018-0014-3.
- 916 18. Ganz, J., Maury, E.A., Becerra, B., Bizzotto, S., Doan, R.N., Kenny, C.J., Shin, T., Kim, J.,
917 Zhou, Z., Ligon, K.L., et al. (2021). Rates and patterns of clonal oncogenic mutations in
918 the normal human brain. *Cancer Discov.* 10.1158/2159-8290.CD-21-0245.
- 919 19. Behrendt, G., Baer, K., Buffo, A., Curtis, M.A., Faull, R.L., Rees, M.I., Gotz, M., and Dimou,
920 L. (2013). Dynamic changes in myelin aberrations and oligodendrocyte generation in
921 chronic amyloidosis in mice and men. *Glia* 61, 273-286. 10.1002/glia.22432.
- 922 20. Jellinger, K.A. (2018). Multiple System Atrophy: An Oligodendroglioneural
923 Synucleinopathy1. *J Alzheimers Dis* 62, 1141-1179. 10.3233/JAD-170397.
- 924 21. Tkachev, D., Mimmack, M.L., Ryan, M.M., Wayland, M., Freeman, T., Jones, P.B.,
925 Starkey, M., Webster, M.J., Yolken, R.H., and Bahn, S. (2003). Oligodendrocyte
926 dysfunction in schizophrenia and bipolar disorder. *Lancet* 362, 798-805. 10.1016/S0140-
927 6736(03)14289-4.
- 928 22. Nagy, C., Maitra, M., Tanti, A., Suderman, M., Theroux, J.F., Davoli, M.A., Perlman, K.,
929 Yerko, V., Wang, Y.C., Tripathy, S.J., et al. (2020). Single-nucleus transcriptomics of the
930 prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor
931 cells and excitatory neurons. *Nat Neurosci* 23, 771-781. 10.1038/s41593-020-0621-y.
- 932 23. Yeung, M.S.Y., Djelloul, M., Steiner, E., Bernard, S., Salehpour, M., Possnert, G., Brundin,
933 L., and Frisen, J. (2019). Dynamics of oligodendrocyte generation in multiple sclerosis.
934 *Nature* 566, 538-542. 10.1038/s41586-018-0842-3.
- 935 24. Yeung, M.S., Zdunek, S., Bergmann, O., Bernard, S., Salehpour, M., Alkass, K., Perl, S.,
936 Tisdale, J., Possnert, G., Brundin, L., et al. (2014). Dynamics of oligodendrocyte
937 generation and myelination in the human brain. *Cell* 159, 766-774.
938 10.1016/j.cell.2014.10.011.

- 939 25. Rivkin, M.J., Flax, J., Mozell, R., Osathanondh, R., Volpe, J.J., and Villa-Komaroff, L.
940 (1995). Oligodendroglial development in human fetal cerebrum. *Ann Neurol* 38, 92-101.
941 [10.1002/ana.410380116](https://doi.org/10.1002/ana.410380116).
- 942 26. Barnabe-Heider, F., Goritz, C., Sabelstrom, H., Takebayashi, H., Pfrieger, F.W., Meletis,
943 K., and Frisen, J. (2010). Origin of new glial cells in intact and injured adult spinal cord.
944 *Cell Stem Cell* 7, 470-482. [10.1016/j.stem.2010.07.014](https://doi.org/10.1016/j.stem.2010.07.014).
- 945 27. Sim, F.J., Zhao, C., Penderis, J., and Franklin, R.J. (2002). The age-related decrease in CNS
946 remyelination efficiency is attributable to an impairment of both oligodendrocyte
947 progenitor recruitment and differentiation. *J Neurosci* 22, 2451-2459. [20026217](https://doi.org/10.1523/JNEUROSCI.2002-02.2002).
- 948 28. Zong, H., Parada, L.F., and Baker, S.J. (2015). Cell of origin for malignant gliomas and its
949 implication in therapeutic development. *Cold Spring Harb Perspect Biol* 7.
950 [10.1101/cshperspect.a020610](https://doi.org/10.1101/cshperspect.a020610).
- 951 29. Liu, C., Sage, J.C., Miller, M.R., Verhaak, R.G., Hippenmeyer, S., Vogel, H., Foreman, O.,
952 Bronson, R.T., Nishiyama, A., Luo, L., and Zong, H. (2011). Mosaic analysis with double
953 markers reveals tumor cell of origin in glioma. *Cell* 146, 209-221.
954 [10.1016/j.cell.2011.06.014](https://doi.org/10.1016/j.cell.2011.06.014).
- 955 30. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R.,
956 Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An Integrative Model of Cellular
957 States, Plasticity, and Genetics for Glioblastoma. *Cell* 178, 835-849 e821.
958 [10.1016/j.cell.2019.06.024](https://doi.org/10.1016/j.cell.2019.06.024).
- 959 31. Filbin, M.G., Tirosh, I., Hovestadt, V., Shaw, M.L., Escalante, L.E., Mathewson, N.D.,
960 Neftel, C., Frank, N., Pelton, K., Hebert, C.M., et al. (2018). Developmental and
961 oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* 360,
962 331-335. [10.1126/science.aao4750](https://doi.org/10.1126/science.aao4750).
- 963 32. Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies,
964 H., Stratton, M.R., and Campbell, P.J. (2017). Universal Patterns of Selection in Cancer
965 and Somatic Tissues. *Cell* 171, 1029-1041 e1021. [10.1016/j.cell.2017.09.042](https://doi.org/10.1016/j.cell.2017.09.042).
- 966 33. Gonzalez-Pena, V., Natarajan, S., Xia, Y., Klein, D., Carter, R., Pang, Y., Shaner, B., Annu,
967 K., Putnam, D., Chen, W., et al. (2021). Accurate genomic variant detection in single cells
968 with primary template-directed amplification. *Proc Natl Acad Sci U S A* 118.
969 [10.1073/pnas.2024176118](https://doi.org/10.1073/pnas.2024176118).
- 970 34. Osorio, F.G., Rosendahl Huber, A., Oka, R., Verheul, M., Patel, S.H., Hasaart, K., de la
971 Fonteijne, L., Varela, I., Camargo, F.D., and van Boxtel, R. (2018). Somatic Mutations
972 Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis.
973 *Cell Rep* 25, 2308-2316 e2304. [10.1016/j.celrep.2018.11.014](https://doi.org/10.1016/j.celrep.2018.11.014).
- 974 35. Montgomery, S.B., Goode, D.L., Kvikstad, E., Albers, C.A., Zhang, Z.D., Mu, X.J., Ananda,
975 G., Howie, B., Karczewski, K.J., Smith, K.S., et al. (2013). The origin, evolution, and
976 functional impact of short insertion-deletion variants identified in 179 human genomes.
977 *Genome Res* 23, 749-761. [10.1101/gr.148718.112](https://doi.org/10.1101/gr.148718.112).
- 978 36. Cingolani, P. (2022). Variant Annotation and Functional Prediction: SnpEff. *Methods Mol*
979 *Biol* 2493, 289-314. [10.1007/978-1-0716-2293-3_19](https://doi.org/10.1007/978-1-0716-2293-3_19).
- 980 37. Lee-Six, H., Obro, N.F., Shepherd, M.S., Grossmann, S., Dawson, K., Belmonte, M.,
981 Osborne, R.J., Huntly, B.J.P., Martincorena, I., Anderson, E., et al. (2018). Population

- 982 dynamics of normal human blood inferred from somatic mutations. *Nature* 561, 473-
983 478. 10.1038/s41586-018-0497-0.
- 984 38. Machado, H.E., Mitchell, E., Obro, N.F., Kubler, K., Davies, M., Leongamornlert, D., Cull,
985 A., Maura, F., Sanders, M.A., Cagan, A.T.J., et al. (2022). Diverse mutational landscapes
986 in human lymphocytes. *Nature* 608, 724-732. 10.1038/s41586-022-05072-7.
- 987 39. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A.,
988 Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of
989 mutational signatures in human cancer. *Nature* 578, 94-101. 10.1038/s41586-020-1943-
990 3.
- 991 40. Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and
992 Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. *Nat Genet*
993 47, 1402-1407. 10.1038/ng.3441.
- 994 41. Reijns, M.A.M., Parry, D.A., Williams, T.C., Nadeu, F., Hindshaw, R.L., Rios Szwed, D.O.,
995 Nicholson, M.D., Carroll, P., Boyle, S., Royo, R., et al. (2022). Signatures of TOP1
996 transcription-associated mutagenesis in cancer and germline. *Nature* 602, 623-631.
997 10.1038/s41586-022-04403-y.
- 998 42. Consortium, I.T.P.-C.A.o.W.G. (2020). Pan-cancer analysis of whole genomes. *Nature*
999 578, 82-93. 10.1038/s41586-020-1969-6.
- 1000 43. Consortium, G.T. (2020). The GTEx Consortium atlas of genetic regulatory effects across
1001 human tissues. *Science* 369, 1318-1330. 10.1126/science.aaz1776.
- 1002 44. Marchal, C., Sasaki, T., Vera, D., Wilson, K., Sima, J., Rivera-Mulia, J.C., Trevilla-Garcia, C.,
1003 Nogues, C., Nafie, E., and Gilbert, D.M. (2018). Genome-wide analysis of replication
1004 timing by next-generation sequencing with E/L Repli-seq. *Nat Protoc* 13, 819-839.
1005 10.1038/nprot.2017.148.
- 1006 45. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A.,
1007 Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative
1008 analysis of 111 reference human epigenomes. *Nature* 518, 317-330.
1009 10.1038/nature14248.
- 1010 46. Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with
1011 ChromHMM. *Nat Protoc* 12, 2478-2492. 10.1038/nprot.2017.124.
- 1012 47. Nott, A., Holtman, I.R., Coufal, N.G., Schlachetzki, J.C.M., Yu, M., Hu, R., Han, C.Z., Pena,
1013 M., Xiao, J., Wu, Y., et al. (2019). Brain cell type-specific enhancer-promoter interactome
1014 maps and disease-risk association. *Science* 366, 1134-1139. 10.1126/science.aay0793.
- 1015 48. Polak, P., Karlic, R., Koren, A., Thurman, R., Sandstrom, R., Lawrence, M., Reynolds, A.,
1016 Rynes, E., Vlahovicek, K., Stamatoyannopoulos, J.A., and Sunyaev, S.R. (2015). Cell-of-
1017 origin chromatin organization shapes the mutational landscape of cancer. *Nature* 518,
1018 360-364. 10.1038/nature14221.
- 1019 49. Huang, W., Bhaduri, A., Velmeshev, D., Wang, S., Wang, L., Rottkamp, C.A., Alvarez-
1020 Buylla, A., Rowitch, D.H., and Kriegstein, A.R. (2020). Origins and Proliferative States of
1021 Human Oligodendrocyte Precursor Cells. *Cell* 182, 594-608 e511.
1022 10.1016/j.cell.2020.06.027.
- 1023 50. Hussmann, J.A., Ling, J., Ravisankar, P., Yan, J., Cirincione, A., Xu, A., Simpson, D., Yang,
1024 D., Bothmer, A., Cotta-Ramusino, C., et al. (2021). Mapping the genetic landscape of
1025 DNA double-strand break repair. *Cell* 184, 5653-5669 e5625. 10.1016/j.cell.2021.10.002.

- 1026 51. Reid, D.A., Reed, P.J., Schlachetzki, J.C.M., Nitulescu, I., Chou, G., Tsui, E.C., Jones, J.R.,
1027 Chandran, S., Lu, A.T., McClain, C.A., et al. (2021). Incorporation of a nucleoside analog
1028 maps genome repair sites in postmitotic human neurons. *Science* 372, 91-94.
1029 [10.1126/science.abb9032](https://doi.org/10.1126/science.abb9032).
- 1030 52. Wu, W., Hill, S.E., Nathan, W.J., Paiano, J., Callen, E., Wang, D., Shinoda, K., van
1031 Wietmarschen, N., Colon-Mercado, J.M., Zong, D., et al. (2021). Neuronal enhancers are
1032 hotspots for DNA single-strand break repair. *Nature* 593, 440-444. [10.1038/s41586-021-](https://doi.org/10.1038/s41586-021-03468-5)
1033 [03468-5](https://doi.org/10.1038/s41586-021-03468-5).
- 1034 53. Lodato, M.A., Woodworth, M.B., Lee, S., Evrony, G.D., Mehta, B.K., Karger, A., Lee, S.,
1035 Chittenden, T.W., D'Gama, A.M., Cai, X., et al. (2015). Somatic mutation in single human
1036 neurons tracks developmental and transcriptional history. *Science* 350, 94-98.
1037 [10.1126/science.aab1785](https://doi.org/10.1126/science.aab1785).
- 1038 54. Evrony, G.D., Cai, X., Lee, E., Hills, L.B., Elhosary, P.C., Lehmann, H.S., Parker, J.J., Atabay,
1039 K.D., Gilmore, E.C., Poduri, A., et al. (2012). Single-neuron sequencing analysis of L1
1040 retrotransposition and somatic mutation in the human brain. *Cell* 151, 483-496.
1041 [10.1016/j.cell.2012.09.035](https://doi.org/10.1016/j.cell.2012.09.035).
- 1042 55. von Bartheld, C.S., Bahney, J., and Herculano-Houzel, S. (2016). The search for true
1043 numbers of neurons and glial cells in the human brain: A review of 150 years of cell
1044 counting. *J Comp Neurol* 524, 3865-3895. [10.1002/cne.24040](https://doi.org/10.1002/cne.24040).
- 1045 56. Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L.,
1046 Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent
1047 features in human versus mouse cortex. *Nature* 573, 61-68. [10.1038/s41586-019-1506-](https://doi.org/10.1038/s41586-019-1506-7)
1048 [7](https://doi.org/10.1038/s41586-019-1506-7).
- 1049 57. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao,
1050 Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-
1051 Cell Data. *Cell* 177, 1888-1902 e1821. [10.1016/j.cell.2019.05.031](https://doi.org/10.1016/j.cell.2019.05.031).
- 1052 58. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet
1053 Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell*
1054 *Syst* 8, 329-337 e324. [10.1016/j.cels.2019.03.003](https://doi.org/10.1016/j.cels.2019.03.003).
- 1055

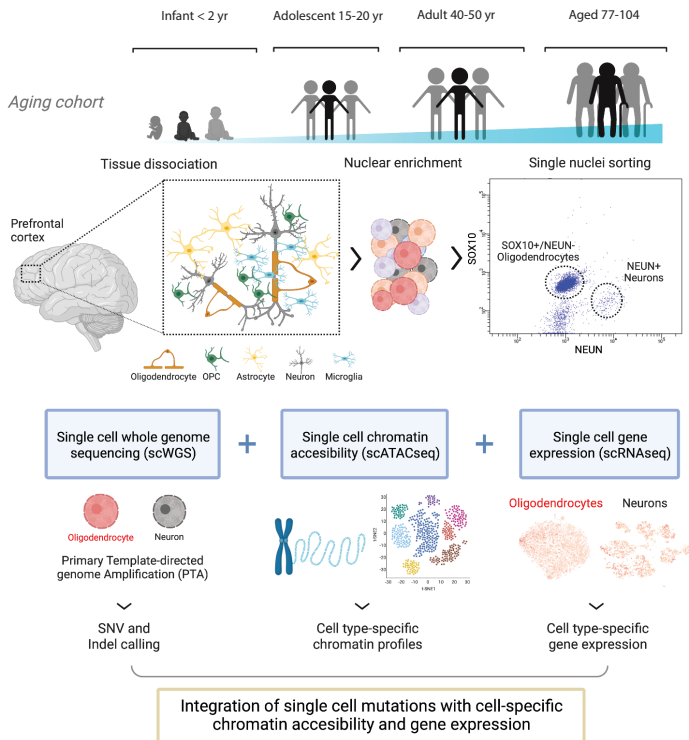
1056 Supplementary Information is available for this paper.

1057

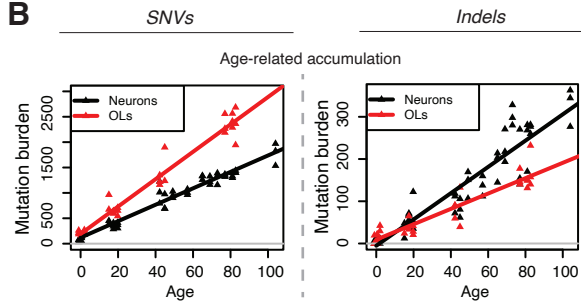
1058 Correspondence and requests for materials should be addressed to CAW and PJP.

Figure 1

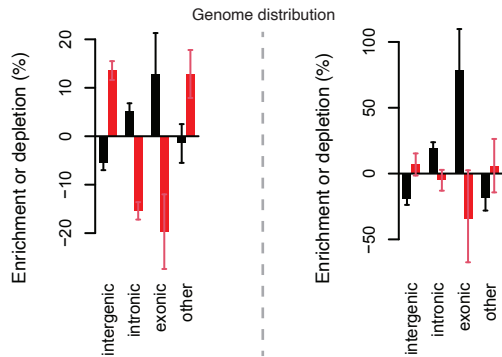
A



B



C



D

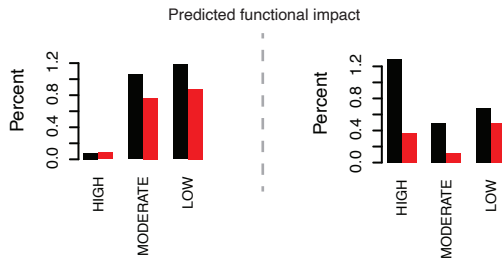


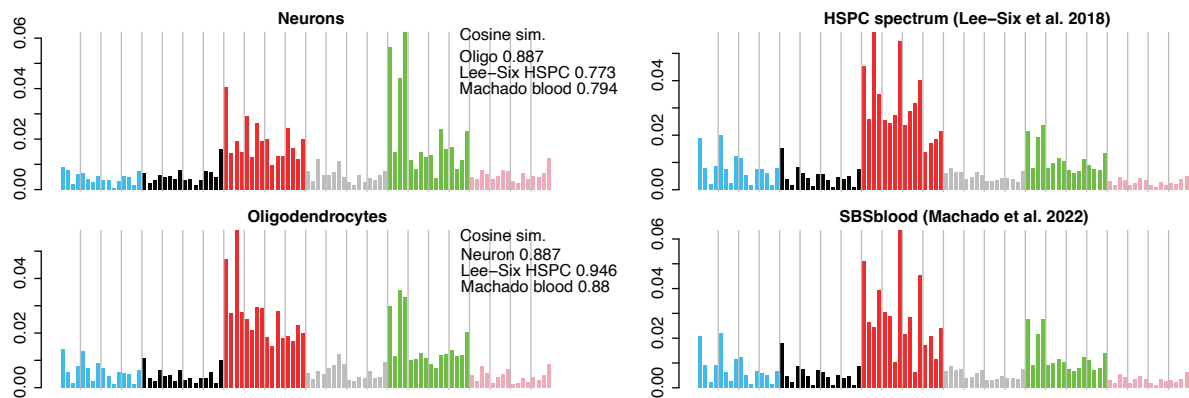
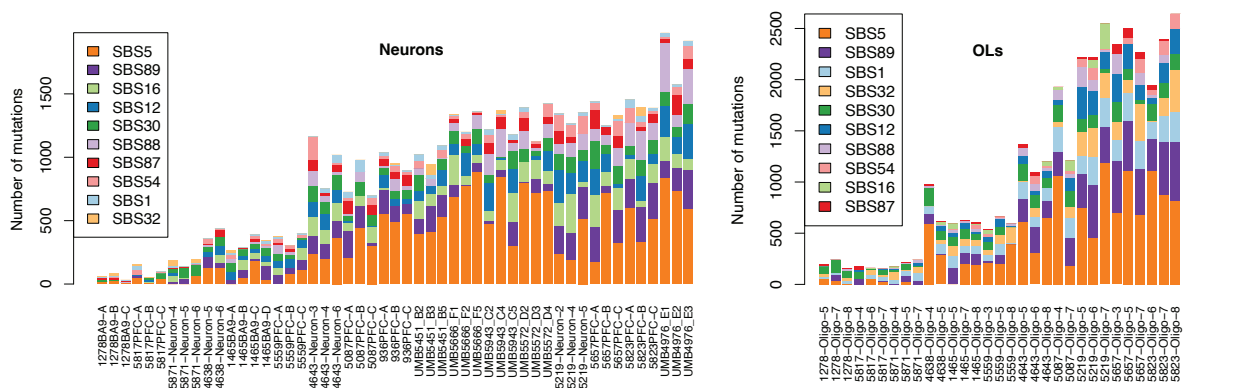
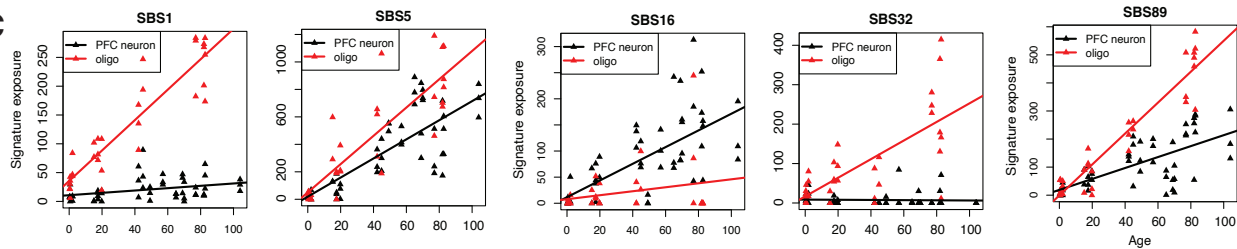
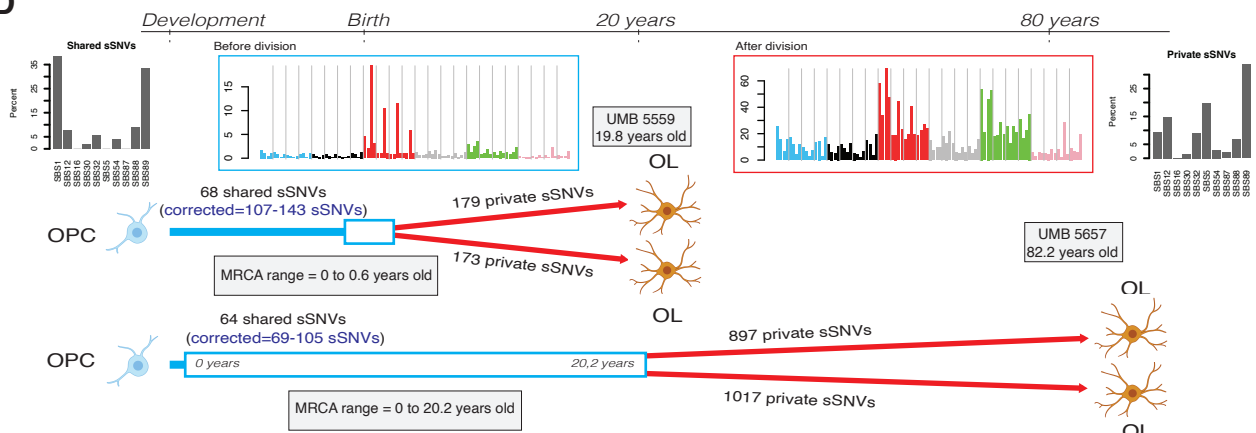
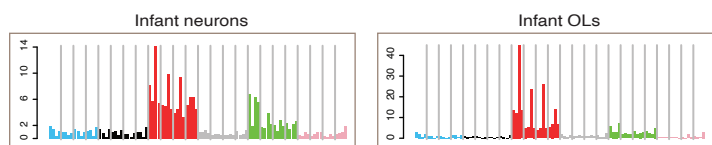
Figure 2**A****B****C****D****E**

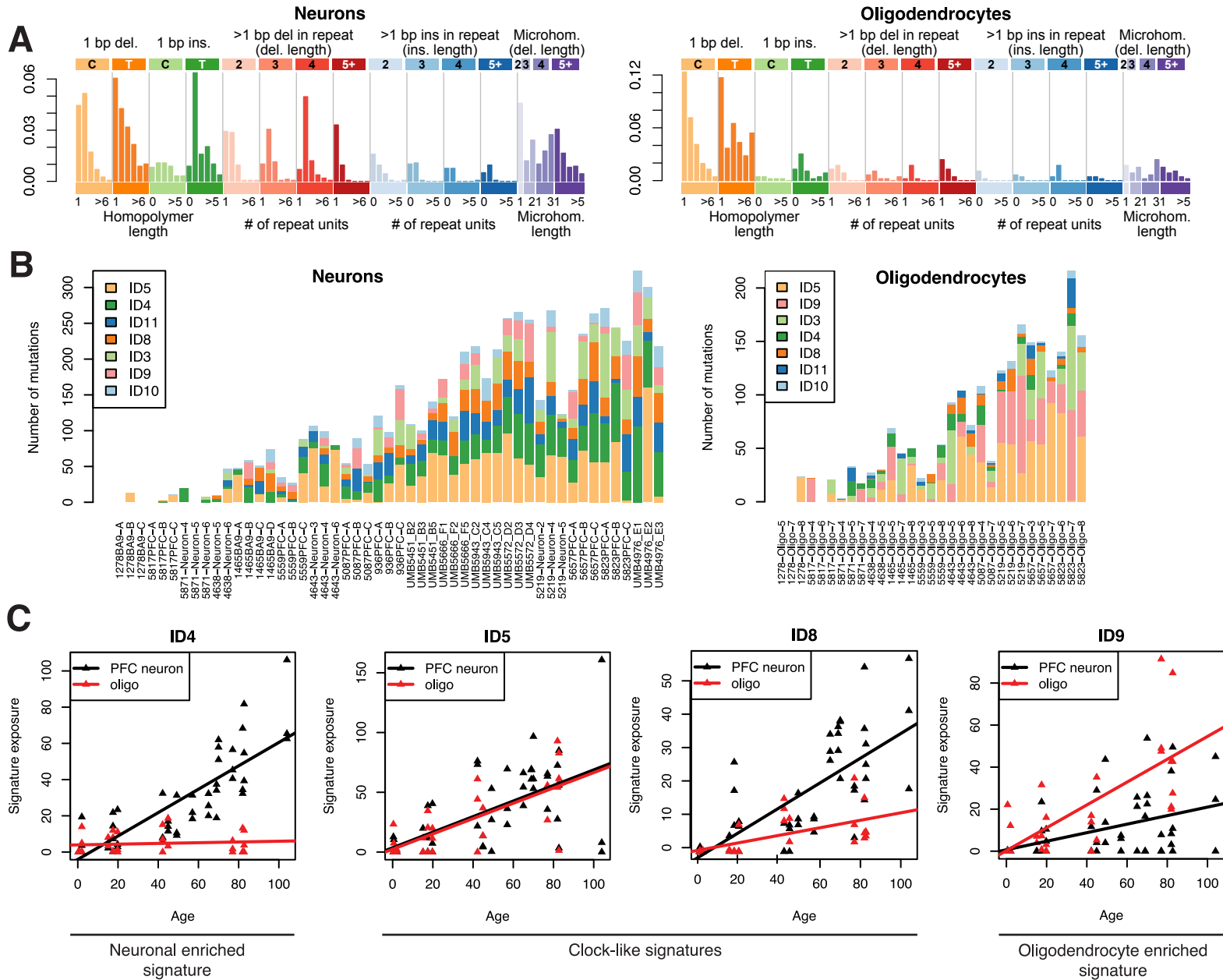
Figure 3

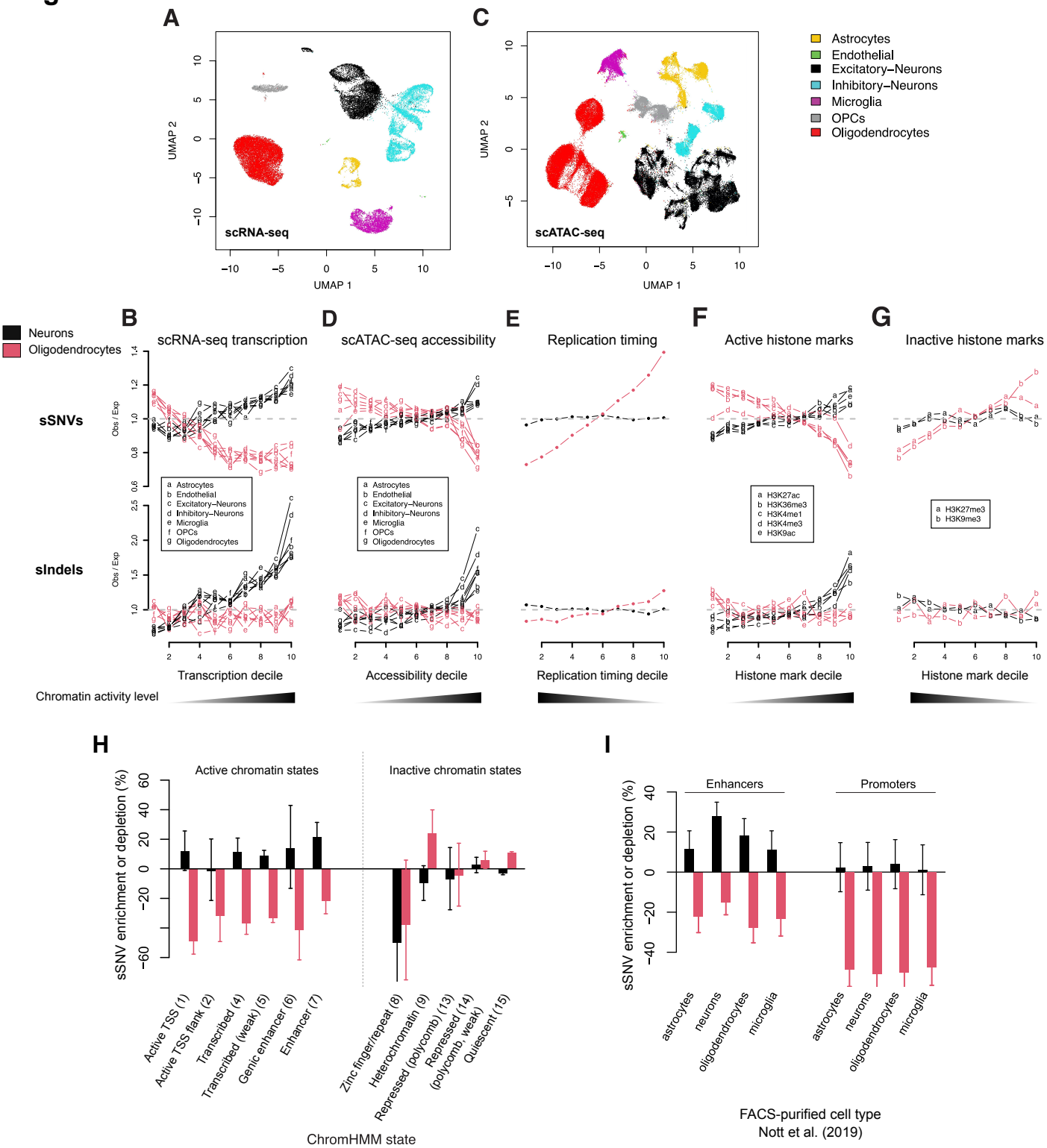
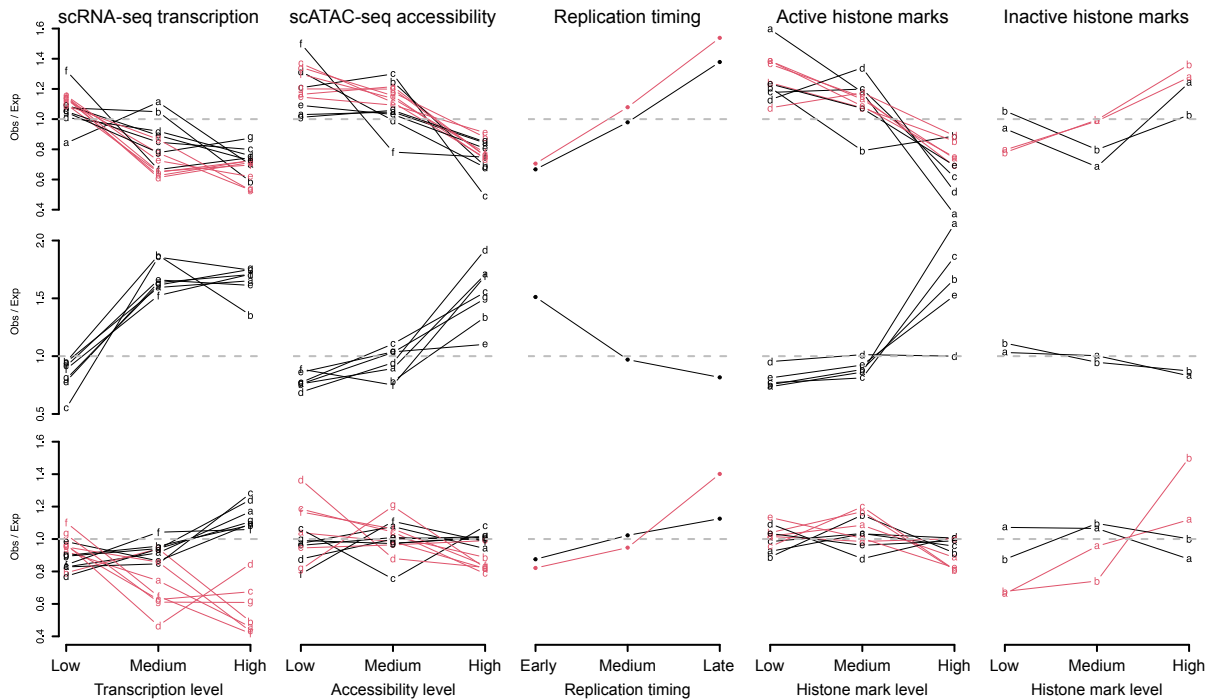
Figure 4

Figure 5

A

SBS1
Neurons
Oligodendrocytes



B

SBS16
Neurons

C

SBS5
Neurons
Oligodendrocytes

- a Astrocytes
- b Endothelial
- c Excitatory-Neurons
- d Inhibitory-Neurons
- e Microglia
- f OPCs
- g Oligodendrocytes

- a Astrocytes
- b Endothelial
- c Excitatory-Neurons
- d Inhibitory-Neurons
- e Microglia
- f OPCs
- g Oligodendrocytes

- a H3K27ac
- b H3K36me3
- c H3K4me1
- d H3K4me3
- e H3K9ac

- a H3K27me3
- b H3K9me3

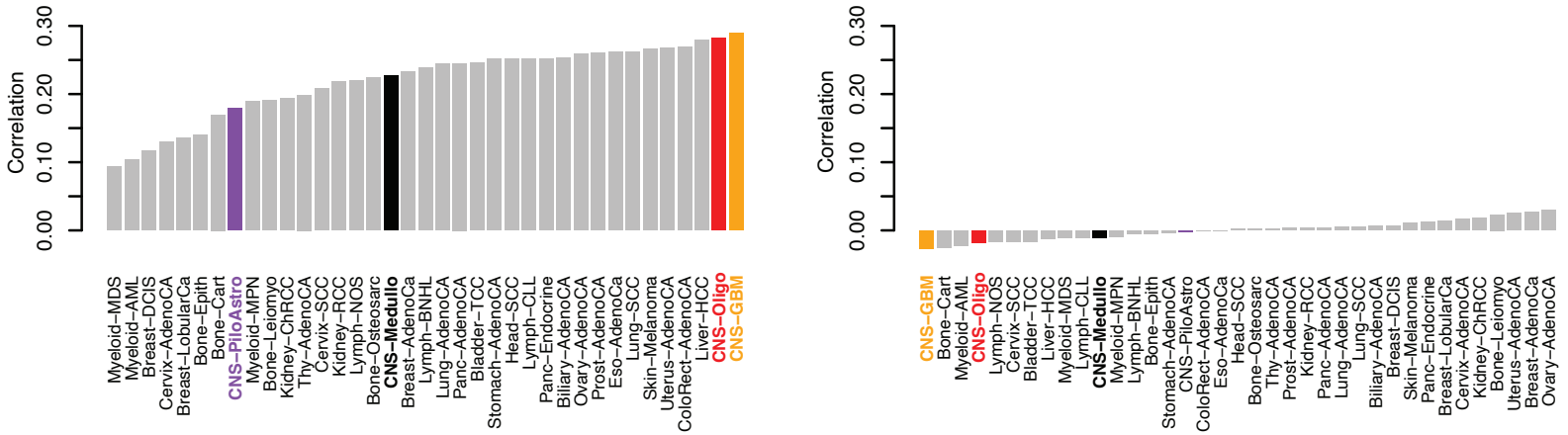
Figure 6

bioRxiv preprint doi: <https://doi.org/10.1101/2023.01.14.523958>; this version posted January 14, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A

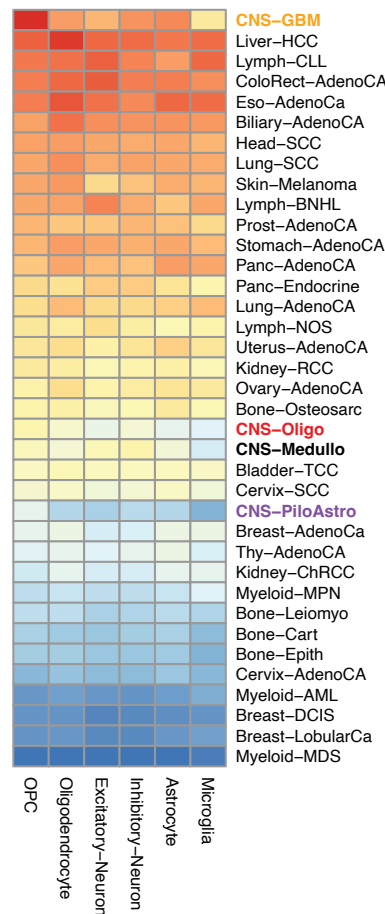
Oligodendrocyte-SNVs vs tumor SNVs

Neuronal-SNVs vs tumor SNVs



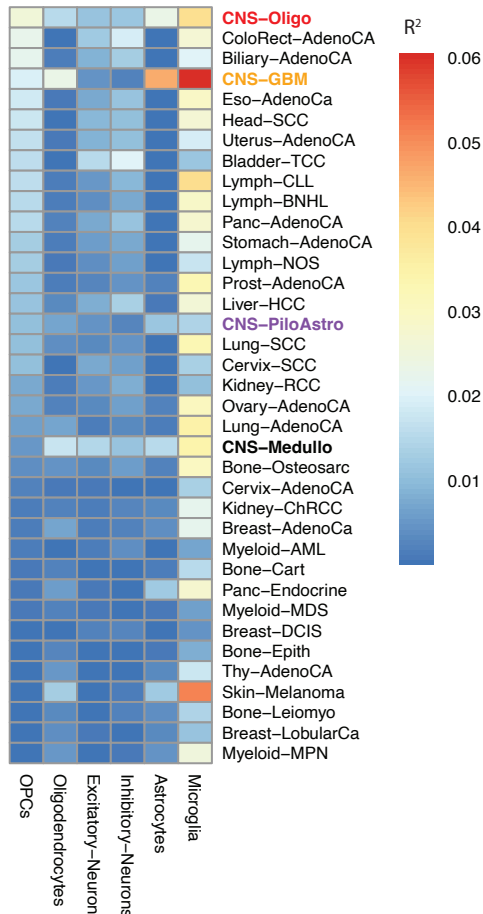
B

scATACseq



C

scrNAseq



D

Odds ratio
Odds of OL mutation in top 100 mutated gene
Odds of Neuron mutation in top 100 mutated gene

