REPORT

NEURODEVELOPMENT

Aging and neurodegeneration are associated with increased mutations in single human neurons

Michael A. Lodato,[1,2,3]* Rachel E. Rodin,[1,2,3,4]* Craig L. Bohrson,[5]* Michael E. Coulter,[1,2,3,4]* Alison R. Barton,[5]* Minseok Kwon,[5]* Maxwell A. Sherman,[5] Carl M. Vitzthum,[5] Lovelace J. Luquette,[5] Chandri N. Yandava,[6] Pengwei Yang,[6] Thomas W. Chittenden,[6,7,8] Nicole E. Hatem,[1,2,3] Steven C. Ryu,[1,2,3] Mollie B. Woodworth,[1,2,3]† Peter J. Park,[5,9]‡ Christopher A. Walsh[1,2,3]‡

It has long been hypothesized that aging and neurodegeneration are associated with somatic mutation in neurons; however, methodological hurdles have prevented testing this hypothesis directly. We used single-cell whole-genome sequencing to perform genome-wide somatic single-nucleotide variant (sSNV) identification on DNA from 161 single neurons from the prefrontal cortex and hippocampus of 15 normal individuals (aged 4 months to 82 years), as well as 9 individuals affected by early-onset neurodegeneration due to genetic disorders of DNA repair (Cockayne syndrome and xeroderma pigmentosum). sSNVs increased approximately linearly with age in both areas (with a higher rate in hippocampus) and were more abundant in neurodegenerative disease. The accumulation of somatic mutations with age—which we term genosenium—shows age-related, region-related, and disease-related molecular signatures and may be important in other human age-associated conditions.

Aging in humans brings increased incidence of nearly all diseases, including neurodegenerative diseases (*1*). Markers of DNA damage increase in the brain with age (*2*), and genetic progeroid diseases such as Cockayne syndrome (CS) and xeroderma pigmentosum (XP), both caused by defects in DNA damage repair (DDR), are associated with neurodegeneration and premature aging (*3*). Mouse models of aging, CS, and XP have shown inconsistent relationships between these conditions and the accumulation of permanent somatic mutations in brain and nonbrain tissue (*4*–*7*). Although analysis of human bulk brain DNA, composed of multiple proliferative and nonproliferative cell types, revealed an accumulation of mutations during aging in the human brain (*8*), it is not known whether permanent somatic mutations accumulate with age in mature neurons of the human brain. Here, we quantitatively examined whether aging or disorders of defective DDR result in more somatic mutations in single postmitotic human neurons.

Somatic mutations that occur in postmitotic neurons are specific to each cell and thus can only be comprehensively assayed by comparing the genomes of single cells (*9*). Therefore, we analyzed human neurons by single-cell whole-genome sequencing (WGS). Because alterations of the prefrontal cortex (PFC) have been linked to age-related cognitive decline and neurodegenerative disease (*10*), we analyzed 93 neurons from PFC of 15 neurologically normal individuals (Table 1 and tables S1 and S2) from ages 4 months to 82 years. We further examined 26 neurons from the hippocampal dentate gyrus (DG) of 6 of these individuals because the DG is a focal point for other age-related degenerative conditions such as Alzheimer's disease. Moreover, the DG is one of the few parts of the brain that appears to undergo neurogenesis after birth (*11*), which might create regional differences in number and type of somatic mutations. Finally, to determine whether defective DDR in early-onset neurodegenerative diseases is associated with increased somatic mutations, we analyzed 42 PFC neurons from 9 individuals diagnosed with CS or XP (Table 1 and table S3).

We isolated single neuronal nuclei using flow cytometry, lysed the nuclei on ice in alkaline conditions as previously performed (*12*, *13*) (see the supplementary materials) to minimize lysis-induced artifacts, amplified their genomes using multiple displacement amplification (MDA), and subjected the amplified DNA to 45X WGS (Fig. 1A). To identify somatic single-nucleotide variants (sSNVs) with high confidence, we developed a bioinformatic pipeline called linked-read analysis (LiRA) (*14*) to delineate true double-stranded sSNVs from single-stranded variants and artifacts. This method employs read-based linkage of candidate sSNVs with nearby germline SNPs and performs a model-based extrapolation of the genome-wide mutational frequency based on the ~20% of sSNVs that are sufficiently close to germline SNPs (Fig. 1B and supplementary materials). sSNVs determined by our algorithm (tables S4 and S5) showed alternate allele frequency distribution strikingly matching that of the germline SNVs (fig. S1). sSNV counts were not systematically influenced by technical metrics, such as postmortem interval, time in storage, and coverage uniformity (fig. S2).

Across all normal neurons, genome-wide sSNV counts correlated with age (Fig. 2A and fig. S3) ($P = 2 \times 10^{-5}$, mixed effects model) (see the supplementary materials), despite some within-individual and within–age group heterogeneity. To explore potential variation in different brain regions, matched DG and PFC neurons were sequenced in six cases (Fig. 2A and fig. S3). Our analysis uncovered region-specific sSNV accumulation with age in both PFC ($P = 4 \times 10^{-5}$) and DG single neurons ($P = 2 \times 10^{-7}$), suggesting an almost twofold increase in the rate of accumulation in DG (~40 sSNVs/year) relative to PFC (~23 sSNVs/year) ($P = 8 \times 10^{-4}$). Among the six cases, two had significant increases in DG, three had nominally increased counts in DG compared to PFC, and one had a nominally higher count in PFC (Fig. 2B).

Neurons from postmortem brains of CS individuals showed a ~2.3-fold excess of sSNVs relative to the expected age-adjusted normal PFC rate, and XP neurons showed a ~2.5-fold increase ($P = 0.006$ for both) (Fig. 2C). Progeroid neurons showed a similar number of sSNVs as neurons from aged normal PFC, suggesting that defective nucleotide excision repair accelerates aging through sSNV accumulation.

Molecular patterns of sSNVs also evolved with age. We previously reported that cytosine deamination influences patterns of human neuron sSNVs, resulting in abundant C>T mutations (*13*). C>T sSNVs accounted for most variants in the youngest PFC samples, but this fraction decreased with age (Fig. 2D and figs. S4 and S5). C>T mutations, although common in many biological contexts (*15*–*18*), are also a known artifact of MDA (*19*). Systematic differences in C>T burden during aging suggest that C>T variants are largely biological and not technical in nature. T>C variants increased in the PFC with age (Fig. 2E and figs. S4 and S5), possibly representing DNA damage linked to fatty-acid oxidation (*20*). As demonstrated previously (*13*), neuronal sSNVs in normal PFC were enriched in coding exons (fig. S6 and table S6) and displayed a transcriptional strand bias (fig. S7), and genes involved in neural function were enriched for neuronal sSNVs (fig. S8 and table S7). Coincident probability modeling suggested that the linear accumulation of sSNVs in our data set would result in an exponential

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease, and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA. [2]Departments of Neurology and Pediatrics, Harvard Medical School, Boston, MA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Program in Neuroscience and Harvard/MIT MD-PHD Program, Harvard Medical School, Boston, MA, USA. [5]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [6]Computational Statistics and Bioinformatics Group, Advanced Artificial Intelligence Research Laboratory, WuXi NextCODE, Cambridge, MA, USA. [7]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [9]Division of Genetics, Brigham and Women's Hospital, Boston, MA, USA.
*These authors contributed equally to this work. †Present address: Cañada College, Redwood City, CA, USA.
‡Corresponding author. Email: peter_park@hms.harvard.edu (P.J.P.); christopher.walsh@childrens.harvard.edu (C.A.W.)

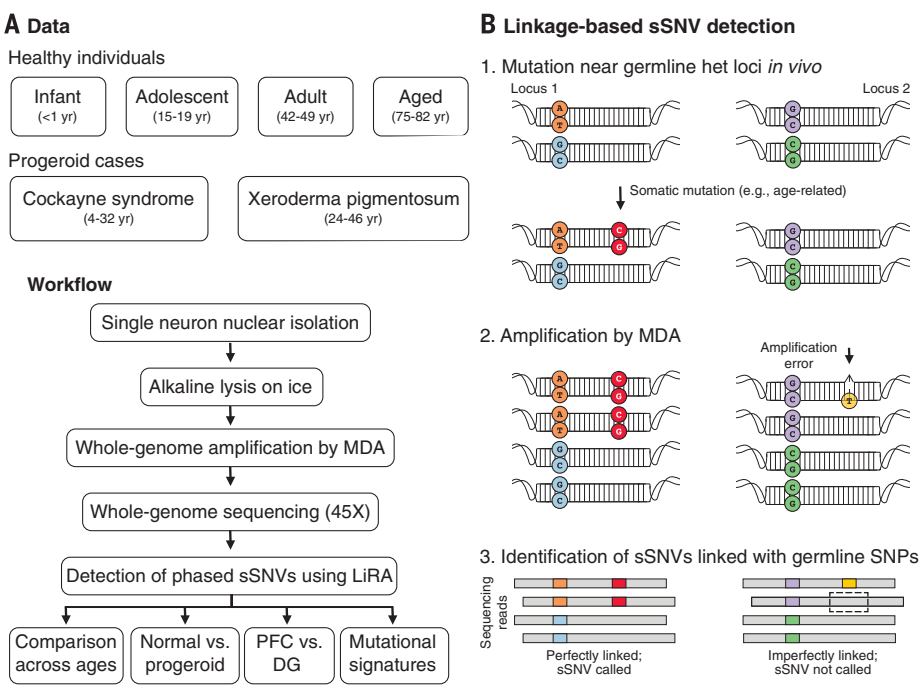

Fig. 1. Detection of sSNVs across individuals and brain regions using single-neuron WGS and linkage-based analysis. (**A**) Experimental outline. (**B**) Schematic of linkage-based mutation calling. A somatic mutation (red) may occur on one allele of a locus (Locus 1), potentially in close proximity to a germline heterozygous site (blue and orange), while other loci, such as Locus 2, remain unmutated. Later amplification errors could create a mismatch (yellow) on one strand of one allele of Locus 2 near a germline variant (purple). For Locus 1, any WGS read that covers both sites and contains the germline (orange) variant should also contain the somatic (red) variant; thus, these variants are perfectly linked. In contrast, at Locus 2, some but not all reads that cover the relevant germline variant (purple) will contain the somatic "variant" candidate (yellow), generating two classes of reads, some with the somatic variant on that allele, some without. Only perfectly linked sSNV candidates were considered in this study.

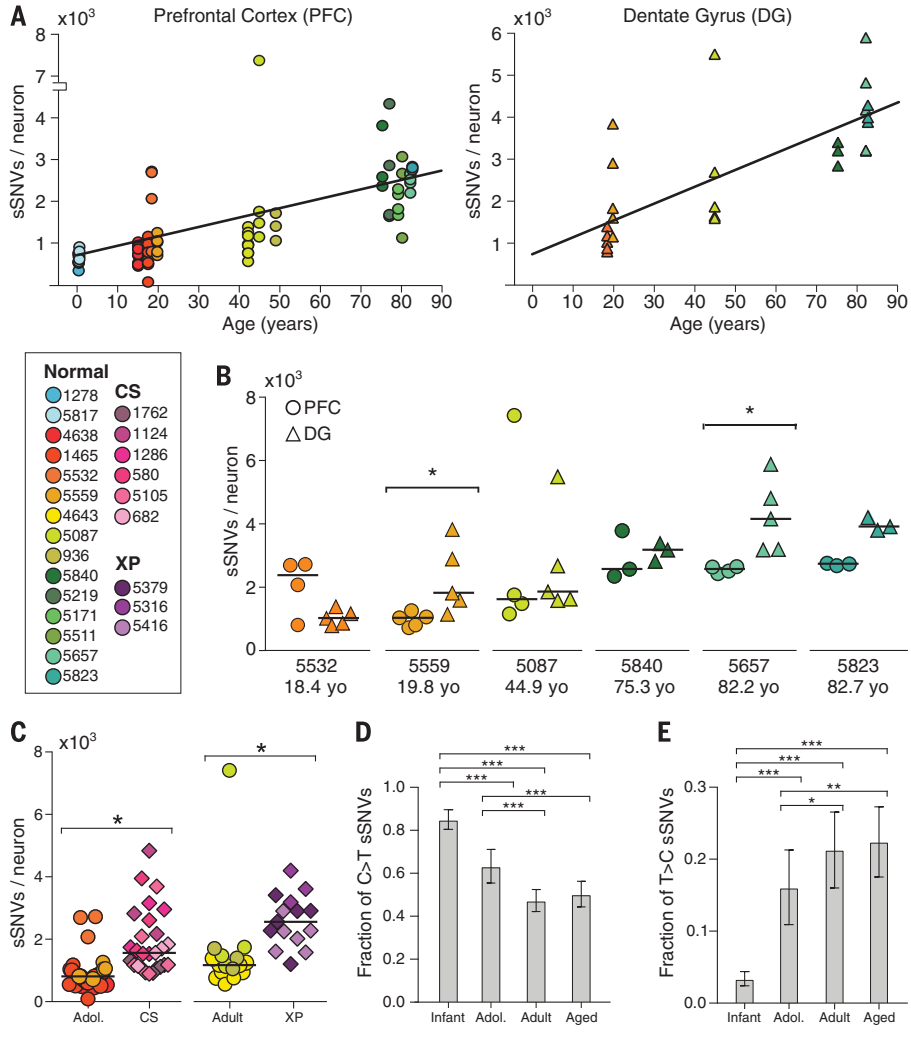

Fig. 2. sSNVs increase with age in the PFC and DG and are elevated in CS and XP. (**A**) sSNV counts plotted against age for neurons derived from PFC (circles) and DG (triangles), with linear regression lines. There is a strong correlation with age in both cases, with the rate of accumulation being nearly twofold higher in the DG than in the PFC. (**B**) Comparison of sSNV counts in matched PFC and DG within brains. (**C**) CS and XP patient neurons display elevated sSNV counts. * denotes $P < 0.05$ for (B) and (C). (**D** and **E**) Fraction of sSNVs composed of C>T (D) and T>C (E). *, **, and *** denote $P \leq 0.05$, 0.001, and 0.0001, respectively, using two-way analysis of variance with Sidak's correction.

accumulation of biallelic deleterious coding mutations, in agreement with classical hypotheses regarding the relationship between mutation and aging (fig. S9) (*21*), exacerbating differences in sSNV load in aging, across brain regions, and in disease.

Mutational signature analysis (*22*) revealed three signatures driving single-neuron mutational spectra (Fig. 3, A and B, and figs. S10 and S11). Signature A was composed mainly of C>T and T>C mutations and was the only signature to increase with age ($P = 9 \times 10^{-12}$), independent of brain region or disease status (Fig. 3, C to E). Signature A resembled a "clocklike" signature found in nearly all samples in a large-scale cancer genome analysis (Signature 5) (*22*) (fig. S10). Our data show

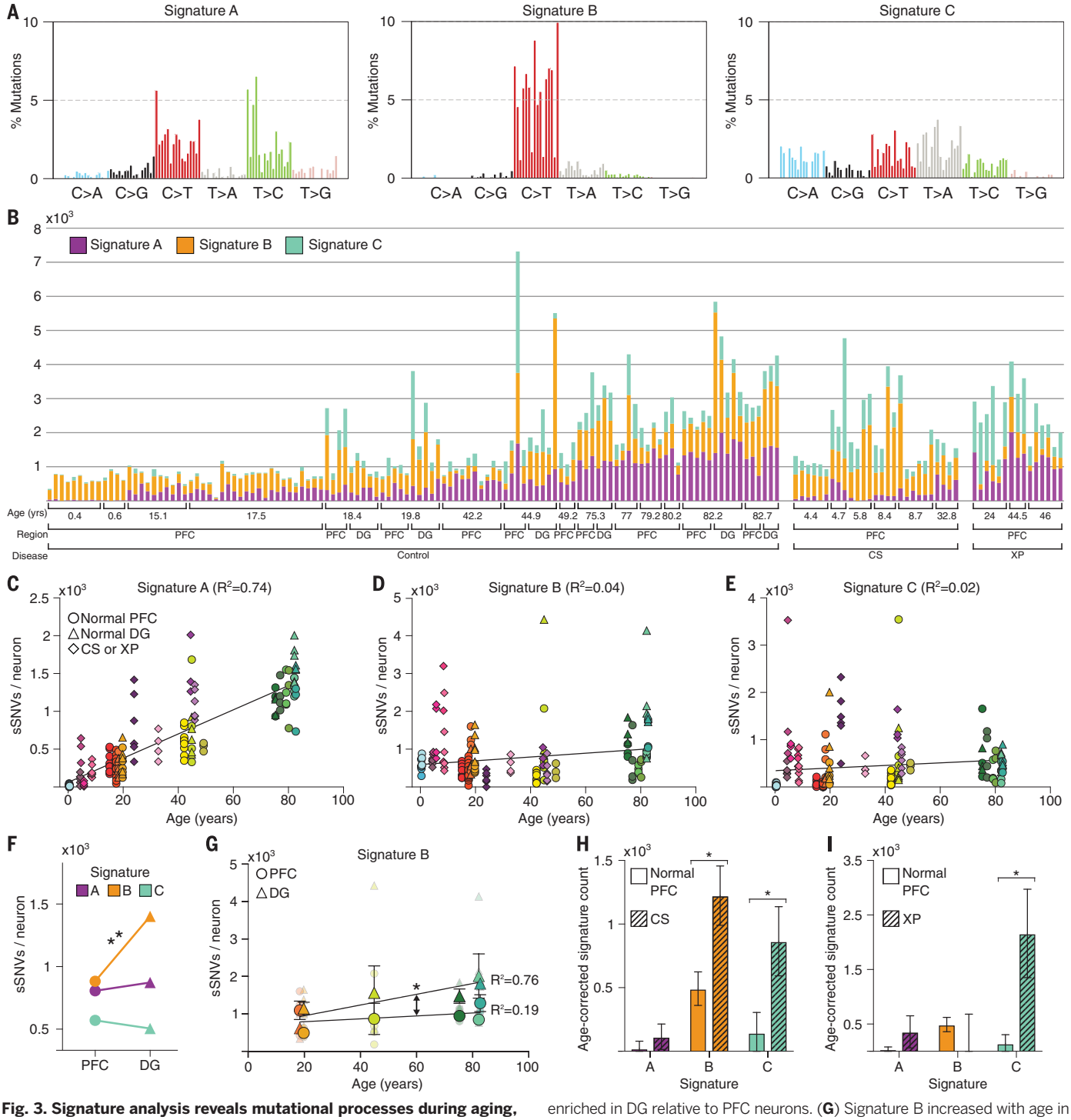

Fig. 3. Signature analysis reveals mutational processes during aging, across brain regions, and in disease. (**A**) Three mutational signatures identified by non-negative matrix factorization (each substitution is classified by its trinucleotide context). (**B**) Number of variants from Signatures A, B, and C in each of the 161 neurons in the data set. (**C** to **E**) Signature A strongly correlates with age, regardless of disease status or brain region, whereas Signatures B and C do not. (**F**) Signature B is enriched in DG relative to PFC neurons. (**G**) Signature B increased with age in DG neurons, but not in matched PFC neurons, revealing a DG-specific aging signature. Solid shapes represent regional means, and transparent shapes represent individual neurons. (**H** and **I**) Comparison of age-corrected estimate of sSNVs per signature in CS and XP compared with PFC controls revealed enrichment in Signature C in both CS and XP. * and ** denote $P \leq 0.05$ and $P \leq 0.001$, respectively; mixed linear model.

that a similar clocklike signature is also active in postmitotic cells and hence independent of DNA replication.

Signature B consisted primarily of C>T mutations and did not correlate with age (Fig. 3D), suggesting a mutational mechanism active at very young ages, perhaps prenatally. Signature B may include technical artifacts, which are primarily C>T, but bona fide clonal sSNVs are also predominantly C>T (*8*, *12*). This signature was enriched in DG compared with PFC ($P = 2 \times 10^{-4}$) (Fig. 3F) and increased with age in DG, but not in PFC ($P = 0.04$, difference in slopes) (Fig. 3G). The observable difference in Signature B between these brain regions, and its correlation with age in DG alone, suggest that it is dominated by a biological mechanism, and these PFC-DG differences strikingly mirror differences in neurogenesis.

A third signature, Signature C, was distinguished from Signatures A and B by the presence of C>A variants, the mutation class most closely associated with oxidative DNA damage (*20*). Indeed, CS and XP neurons, defective in DDR, were enriched for Signature C ($P = 0.016$ and 0.023, respectively) (Fig. 3, H and I), whereas Signature C also increased modestly with age in normal neurons ($P = 0.03$). An outlier 5087 PFC neuron with the highest sSNV rate in our data set had a high proportion of Signature C mutations relative to other normal neurons, highlighting that even within a normal brain some neurons may be subject to catastrophic oxidative damage.

Our analysis revealed that sSNVs accumulated slowly but inexorably with age in the normal human brain, a phenomenon we term genosenium, and more rapidly still in progeroid neurodegeneration. Within 1 year of birth, postmitotic neurons already have ~300 to 900 sSNVs, strikingly dovetailing with the 200 to 400 sSNVs estimated to be already present in human progenitor cells at 20 weeks of gestation or earlier (*23*). Three signatures were associated with mutational processes in human neurons: a postmitotic, clocklike signature of aging, a possibly developmental signature that varied across brain regions, and a disease- and age-specific signature of oxidation and defective DNA damage repair. These associations were present even when data were reanalyzed after removal of all C>T mutations (fig. S11), demonstrating that any artifactual C>T mutations that may have escaped our filtering do not affect our main conclusions. The increase of oxidative mutations in aging and in disease presents a potential target for therapeutic intervention. Further, elucidating the mechanistic basis of the clocklike accumulation of mutations across brain regions and other tissues would increase our knowledge of age-related disease and cognitive decline. CS and XP cause neurodegeneration associated with higher rates of sSNVs, and it will be important to define how other, more common causes of neurodegeneration may influence genosenium as well.

Table 1. Case information and number of neurons analyzed in this study.

Case ID	Sex	Age (years)	Diagnosis	PFC neurons	DG neurons
Infant					
1278	M	0.4	Normal	9	–
5817	M	0.6	Normal	4	–
				13	–
Adolescent					
4638	F	15.1	Normal	10	–
1465	M	17.5	Normal	22	–
5532	M	18.4	Normal	4	5
5559	F	19.8	Normal	5	5
				41	**10**
Adult					
4643	F	42.2	Normal	10	–
5087	M	44.9	Normal	4	5
936	F	49.2	Normal	3	–
				17	**5**
Aged					
5840	M	75.3	Normal	3	3
5219	F	77	Normal	4	–
5171	M	79.2	Normal	4	–
5511	F	80.2	Normal	3	–
5657	M	82	Normal	5	5
5823	F	82.7	Normal	3	3
				22	**11**
Cockayne syndrome					
1762	F	4.4	CS (*CSB*)	6	–
1124	F	4.7	CS (*CSB*)	3	–
1286	M	5.8	CS (*CSB*)	4	–
580	F	8.4	CS (*CSB*)	4	–
5105	M	8.7	CS (*CSB*)	6	–
682	M	32.8	CS (*CSB*)	4	–
				27	–
Xeroderma pigmentosum					
5379	F	24	XP (*XPA*)	6	–
5316	F	44.5	XP (*XPA*)	3	–
5416	F	46	XP (*XPD*)	6	–
				15	**–**
Total	24 cases			**135 PFC neurons**	**26 DG neurons**
				161 neurons	

REFERENCES AND NOTES

1. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, *Cell* **153**, 1194–1217 (2013).
2. T. Lu *et al.*, *Nature* **429**, 883–891 (2004).
3. J. A. Marteijn, H. Lans, W. Vermeulen, J. H. Hoeijmakers, *Nat. Rev. Mol. Cell Biol.* **15**, 465–481 (2014).
4. H. Giese, M. E. Dollé, A. Hezel, H. van Steeg, J. Vijg, *Oncogene* **18**, 1257–1260 (1999).
5. M. E. Dollé, W. K. Snyder, D. B. Dunson, J. Vijg, *Nucleic Acids Res.* **30**, 545–549 (2002).
6. M. E. Dollé *et al.*, *Mutat. Res.* **596**, 22–35 (2006).
7. R. A. Busuttil *et al.*, *PLOS ONE* **2**, e876 (2007).
8. M. L. Hoang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9846–9851 (2016).
9. J. P. Dumanski, A. Piotrowski, *Methods Mol. Biol.* **838**, 249–272 (2012).
10. S. J. van Veluw *et al.*, *Brain Struct. Funct.* **217**, 797–808 (2012).
11. P. S. Eriksson *et al.*, *Nat. Med.* **4**, 1313–1317 (1998).
12. G. D. Evrony *et al.*, *Cell* **151**, 483–496 (2012).
13. M. A. Lodato *et al.*, *Science* **350**, 94–98 (2015).
14. C. L. Bohrson *et al.*, bioRxiv 211169 [Preprint] 30 October 2017.
15. G. Genovese *et al.*, *N. Engl. J. Med.* **371**, 2477–2487 (2014).
16. J. L. Hazen *et al.*, *Neuron* **89**, 1223–1236 (2016).
17. I. Martincorena *et al.*, *Science* **348**, 880–886 (2015).
18. E. T. Lim *et al.*, *Nat. Neurosci.* **20**, 1217–1224 (2017).
19. Y. Hou *et al.*, *Cell* **148**, 873–885 (2012).
20. R. De Bont, N. van Larebeke, *Mutagenesis* **19**, 169–185 (2004).
21. L. Szilard, *Proc. Natl. Acad. Sci. U.S.A.* **45**, 30–45 (1959).
22. L. B. Alexandrov *et al.*, *Nat. Genet.* **47**, 1402–1407 (2015).
23. T. Bae *et al.*, *Science* **359**, 550–555 (2018).

ACKNOWLEDGMENTS

We thank R. S. Hill, M. P. Anderson, J. Gulcher, Z. Chen, I. Cortez, the Dana Farber Cancer Institute Flow Cytometry Core, and the Research Computing group at Harvard Medical School for assistance. Human tissue was obtained from the NIH NeuroBioBank at the University of Maryland, and we thank the donors and their families for their invaluable donations for the advancement of science. This work was supported by K99 AG054749 01, F30 MH102909, 1S10RR028832-01, T32HG002295, U01MH106883, P50MH106933, R01 NS032457 U01 MH106883, the Harvard/MIT MD-PHD program, the Stuart H. Q. and Victoria Quan Fellowship in Neurobiology, and the Paul G. Allen Family Foundation. C.A.W. is an investigator of the Howard Hughes Medical Institute. See the supplementary materials for full acknowledgments. The supplementary materials also contain additional data.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6375/555/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S7
References (*24*–*52*)

21 July 2017; accepted 22 November 2017
Published online 7 December 2017
10.1126/science.aao4426

Science

Aging and neurodegeneration are associated with increased mutations in single human neurons

Michael A. Lodato, Rachel E. Rodin, Craig L. Bohrson, Michael E. Coulter, Alison R. Barton, Minseok Kwon, Maxwell A. Sherman, Carl M. Vitzthum, Lovelace J. Luquette, Chandri N. Yandava, Pengwei Yang, Thomas W. Chittenden, Nicole E. Hatem, Steven C. Ryu, Mollie B. Woodworth, Peter J. Park and Christopher A. Walsh

Science **359** (6375), 555-559.
DOI: 10.1126/science.aao4426originally published online December 7, 2017

Brain mutations, young and old

Most neurons that make up the human brain are postmitotic, living and functioning for a very long time without renewal (see the Perspective by Lee). Bae *et al.* examined the genomes of single neurons from the prenatal developing human brain. Both the type of mutation and the rates of accumulation changed between gastrulation and neurogenesis. These early mutations could be generating useful neuronal diversity or could predispose individuals to later dysfunction. Lodato *et al.* also found that neurons take on somatic mutations as they age by sequencing single neurons from subjects aged 4 months to 82 years. Somatic mutations accumulated with increasing age and accumulated faster in individuals affected by inborn errors in DNA repair. Postmitotic mutations might only affect one neuron, but the accumulated divergence of genomes across the brain could affect function.

Science, this issue p. 550, p. 555; see also p. 521