Title

Analysis of somatic mutations in 131 human brains reveals aging-associated hypermutability

Journal

Science, 377(6605)

ISSN

0036-8075

Authors

Bae, Taejeong
Fasching, Liana
Wang, Yifan
et al.

Publication Date

2022-07-29

DOI

10.1126/science.abm6222

Peer reviewed

HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2023 January 29.

Published in final edited form as:

Science. 2022 July 29; 377(6605): 511–517. doi:10.1126/science.abm6222.

Analysis of somatic mutations in 131 human brains reveals aging associated hypermutability

Taejeong Bae[1,†], **Liana Fasching**[2,†], **Yifan Wang**[1,†], **Joo Heon Shin**[3,4,†], **Milovan Suvakov**[1], **Yeongjun Jang**[1], **Scott Norton**[2], **Caroline Dias**[5,6], **Jessica Mariani**[2], **Alexandre Jourdon**[2], **Feinan Wu**[2], **Arijit Panda**[1], **Reenal Pattni**[7], **Yasmine Chahine**[5,6], **Rebecca Yeh**[5,6], **Rosalinda C. Roberts**[8], **Anita Huttner**[9], **Joel E. Kleinman**[3,10], **Thomas M. Hyde**[3,4,10], **Richard E. Straub**[3], **Christopher A. Walsh**[5,6],
Brain Somatic Mosaicism Network[‡],
Alexander E. Urban[7], **James F. Leckman**[2], **Daniel R. Weinberger**[3,4,10,11,12,*], **Flora M. Vaccarino**[2,13,*], **Alexej Abyzov**[1,*]

[1]Department of Quantitative Health Sciences, Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905

[2]Child Study Center, Yale University, New Haven, CT 06520

[3]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD

[4]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD

[5]Division of Genetics and Genomics and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA

[6]Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA, USA

[7]Department of Psychiatry and Behavioral Sciences, Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305

[*]Corresponding authors. drweinberger@libd.org (D. R. W.); flora.vaccarino@yale.edu (F.M.V.); Abyzov.Alexej@mayo.edu (A.A.).
[†]These authors contributed equally to this work.
[‡]The Brain Somatic Mosaicism Network authors and affiliations are listed in the supplementary materials

Author contributions: J.F.L. performed retrospective clinical phenotyping of the TS cases by medical record and interviews with family members. L.F., R.C.R., A.H., J.E.K., T.M.H. processed brains. L.F., J.H.S., J.M., A.J. performed experiments and generated the data. T.B., Y.W., J.H.S., M.S., Y.J., S.N., F.W., A.P. performed data analyses. L.F., C.D., R.P., Y.C. conducted validation of mutations. T.B., L.F., Y.W., M.S. prepared display items. T.B., L.F., Y.W., M.S., S.N., C.D., R.E.S., D.R.W., F.M.V., A.A. wrote the initial draft of the manuscript. T.M.H., R.E.S., C.A.W, A.E.U., D.R.W., F.M.V., A.A. oversaw experiments and analyses. All authors participated in discussions of results and manuscript editing.

Competing interests: The authors declare no competing interests.

Ethics Statement:
The analyzed samples and generated data were derived from de-identified postmortem individuals. Brains with Tourette Syndrome sequenced by Yale University were obtained from the Harvard's Brain and Tissue Resource Center (HBTRC). Most normal brains sequenced at Yale University were obtained from NIH NeuroBioBank.
For normal brains from Yale University autopsy authorization, consent to utilize tissues for research and consent to publish was provided by the patient's next-of-kin according to Connecticut state law and approved protocols of the Yale University BioBank. The research study was approved by the Yale Alzheimer Disease Research Center (ADRC) and was reviewed and deemed exempt by the Yale University Institutional Review Board. The BioBank protocols are in accordance with the ethical standards of Yale University. Handling of brains originating from and sequenced by the Lieber Institute was conducted under the approved WCG IRB protocol 20111080 titled "Collection of Postmortem Human Brain, Blood and Scalp Samples for Neuropsychiatric Research".
Data for brains sequenced by Harvard University were publicly available from NIMH Data Archive.

[8]Department of Psychiatry and Behavioral Neurobiology, University of Alabama at Birmingham, Birmingham Al, 35294

[9]Department of Pathology, Yale University, New Haven, CT 06520

[10]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD

[11]Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD

[12]Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD

[13]Department of Neuroscience, Yale University, New Haven, CT 06520

Abstract

We analyzed 131 human brains (44 neurotypical, 19 with Tourette syndrome, 9 with schizophrenia, and 59 with autism) for somatic mutations after whole genome sequencing to over 200X depth. Typically, brains had 20 to 60 detectable single nucleotide mutations but ~6% of brains harbored hundreds of somatic mutations. Hypermutability was associated with age and damaging mutations in genes implicated in cancers and in some brains reflected in vivo clonal expansions. Somatic duplications, likely arising during development, were found in ~5% of normal and diseased brains, reflecting background mutagenesis. Brains with autism were associated with mutations creating putative transcription factor binding motifs in enhancer-like regions in the developing brain. The top-ranked affected motifs corresponded to MEIS transcription factors, suggesting a potential link between their involvement in gene regulation and autism.

One-Sentence Summary:

A survey of brain genomes revealed the extent of somatic mutations with normal age and with certain neuropsychiatric disorders.

Somatic mutations naturally occur in proliferative and post-mitotic cells throughout human development and during aging, starting from the first cleavage of the zygote (1–4). How frequent are somatic mutations in the population, and are they a contributing factor to the etiology of neuropsychiatric disorders (5)? Mutations analyzed in bulk tissues are typically present in >1% of cells, often arise during early development, and may be enriched for variants exerting a stronger phenotypic effect. We investigated somatic mosaicism in normal human brains (n = 44) and in brains of those affected by 3 neuropsychiatric diseases: Tourette Syndrome (TS, n = 19); schizophrenia (SCZ, n = 9); or autism spectrum disorder (ASD, n = 59).

Results

Mutation discovery and counts across cohorts

For each frozen post-mortem brain, 1 to 2 regions (cortex, striatum, or hippocampus) were extracted (~1 cm^3) in one of three institutions: Yale University, the Lieber Institute for Brain Development, and Harvard University (the data was reported previously (6)). DNA extracted

from bulk samples, from FACS-sorted neuronal and glial cell fractions, or from both was analyzed by whole genome sequencing (WGS) on an Illumina platform at various depths, resulting in a combined coverage above 200X per brain for 92% of the samples, and as high as 620X (Fig. S1, Table S1). The data were uniformly processed – reads were aligned and somatic point mutations were called in all samples using a previously verified workflow (7). Somatic mutations were distinguished from inherited variations based on their frequency and based on no overlap with the catalogues of known germline variations in the human population.

In brains with bulk samples sequenced to at least 200X coverage (tier 1 set), we discovered 20 to 60 somatic single nucleotide mutations per brain with typical allele frequencies that ranged from 1% to 10% across all cohorts (Figs. 1A & S2B; Table S2). Additional mutations (included in tier 2 set) were discovered when considering additional sequencing coverage including data for cell fractions (**Methods**), but for an unbiased comparison across cohorts we used tier 1 set unless otherwise noted. In every cohort there was at least one brain with an outlier high count (i.e., above upper outer fence of 101) somatic mutations, which we term a hypermutable brain (Fig. 1A, **Methods**). The mutation calls in those brains had the same substitution spectrum as for other brains (Figs. 1B,C & S2E). From read-backed phasing of calls to personal DNA haplotypes, the proportions of supported (i.e., assigned to a single haplotype) and unsupported (i.e., with evidence of being at two haplotypes) calls in the hypermutable brains was the same as in other brains (Fig. 1D), indicating the same quality of calls across all brains. Thus, we concluded that calls in hypermutable brains represent *bona fide* somatic mutations. We further conducted mutation validation in two hypermutable brains, NC7 and LIBD82. In NC7 at least 87.5% of mutation calls were validated using single cell sequencing (Fig. 2). Similarly, we validated at least 80% of mutations in transcriptome in brain LIBD82 by RNA sequencing (**Methods**).

When we excluded the hypermutable brains, there were no significant differences in somatic mutation burden between diseased and normal brains, the allele frequency of the somatic mutations, or the somatic mutation spectrum (Fig. 1C,E). The combined mutation spectrum was dominated by C>T transitions in CpG motifs and matched the spectrum of mutations arising during development (Figs. 1B,C and S2E,G) (1, 7, 8). Brains in each cohort had different age distributions, but the mutation burden in non-hypermutable brains did not correlate with age (Fig. S2F), also consistent with somatic mutations being of early developmental origin. Contrary to that, the proportion of hypermutable brains rose with age (p-value = $8.2\text{x}10^{-3}$ by χ^2 test for trend), suggesting that those brains have mutations arising both in development and during aging (Fig. 1E). Specifically, hypermutable brains constituted 16% of brains over 60 years old, while only 2% of those under 40 years old. One of the hypermutable schizophrenia brains with the highest mutation burden, LIBD82, had genomic aneuploidies in the hippocampus, including a duplication of chromosome 7 and a deletion of chromosome 10 (Fig. 1F,G), which is a signature of glioblastoma (9). Since clinical characterization of the brain did not provide any evidence of the disease, the presence of aneuploidies in this brain is consistent with the idea that driver mutations often precede cancer diagnosis by years, if not decades (10). Thus, hypermutability in this brain may reflect incipient glioblastoma.

Origin of hypermutability

Hypermutable brains were found in patients and controls and in every cohort of samples generated by each lab. Thus, it is unlikely that the large mutation load in these brains is associated with disease status or arises from cohort- or lab-specific conditions used for brain storage, handling, or sample processing. We hypothesize that the hypermutability arises from individual-specific biological causes. One possibility is that cells of such brains are intrinsically hypermutable. However, the frequency spectrum of mutations in hypermutable as compared to non-hypermutable brains was depleted for higher frequency mutations (>6% VAF), overrepresented for intermediate frequencies (3%-6%), and almost the same for low frequency (<3% VAF) mutations (Figs. S2B–D). These data suggest that intrinsic hypermutability, if it exists, is not innate (that is, present in all cell divisions throughout development), but occurred sometime during the lifetime of that individual.

Out of 6 hypermutable brains, two, TS9 and NC7, had the same known missense mutation (chr1:115,258,747 C>T) in the *NRAS* oncogene. This mutation has been previously identified in multiple cancers and is deemed to be pathogenic in the COSMIC database (ID: COSV54736383). Furthermore, 3 hypermutable brains (NC7, TS9, and AN05983) carried 11 mutations predicted to have a damaging effect in genes previously implicated in cancers (Table 1), such as *MTOR*, *TET2*, *DNMT3A*, and *IDH2*. Overall, hypermutable brains were enriched for damaging mutations in genes implicated in cancers (12, 13) as compared to non-hypermutable brains (p-value = $2.4x10^{-3}$ by Fischer's exact test; Fig. S3A).

Given the above data, we hypothesized that the cell lineage carrying one or more of these mutations underwent a clonal expansion, which allowed other preceding lineage-specific mutations to arise to a detectable frequency. In both TS9 and NC7 brains, virtually all of the mutations were present in both bulk cortex and striatum at an allele frequency of about 2%, as well as in multiple other cell fractions at various frequencies (Figs. 2A,B and S3B,C). This could be explained by clonal hematopoiesis coupled with blood cell infiltration into brain, so that mutations from blood are present in different brain regions. Hence, we investigated the expression of blood-specific genes (i.e., markers) in the studied brains. Several markers consistently indicated that only brain NC7 had a higher fraction of blood cells as compared to all other brains (Fig. S4). However, in the same brain NC7, all the mutations were present in sorted interneurons from the striatum (STR-INT) at a much higher frequency (15-25% VAF), suggesting that the expanded lineage originated not in the blood, but rather in the embryonic basal ganglia, where interneurons are generated, and then populated cortex and striatum by cell migration.

We tested the hypothesis of lineage expansion (either of interneurons or of hematopoietic cells) by sequencing 16 single nuclei isolated from a different sample of the STR-INT fraction from the NC7 brain. Of all discovered somatic mutations in bulk NC7 brain, 346 (94%) were genotyped in eight out of 16 nuclei, with 154 to 197 mutations genotyped in each nucleus (Fig. 2C). The genotyping rate, 42% to 53% across nuclei, was comparable to the genotyping sensitivity (34% to 46%, see **Methods**) allowing us to extrapolate that each of the eight nuclei contained almost all the mutations discovered in the bulk cortex and striatum, and proving that the mutations belonged to the same cell lineage.

Differently than NC7 and TS9, mutations in other hypermutable brains with data available for multiple regions (LIBD82, LIBD96, LIBD98, and LIBD107) were restricted to one brain region (Fig. S3D–F). For brain LIBD82, the frequencies (~15% in hippocampus) of point mutations and of genomic aneuploidies were matching, consistent with both mutation types being present in expanded lineages (Fig. 1E,F). Altogether, the above evidence suggests that hypermutability in some brains is caused by lineage expansion, either in development or later in life.

Somatic Mutations revealed uneven cell lineage distribution in brain

For each brain processed at Yale University (19 TS and 21 neurotypical brains), we obtained sequencing data for bulk cortex, striatum, and for up to 8 cell fractions per brain region (see **Methods**). In both the cortex and striatum, we successfully fractionated neurons and non-neuronal cells (NeuN+/NeuN−); in the striatum, we also fractionated medium spiny neurons (STR-MSN) and interneurons (STR-INT) (Fig. S1B and **Methods**). From each brain processed at the Lieber Institute for Brain Development, we obtained sequenced data for bulk cortex and hippocampus. The frequency of mosaic mutations in multiple regions and fractions allowed us to explore the cell lineage distribution in the brain.

To investigate whether the above result (Fig. 2A,B) could reflect a more general phenomenon, we tested for the difference in mutation VAFs between the cortex and striatum. If the early lineage distribution is the same across these two regions, there should be, on average, no difference in VAFs between them (**Methods**). Out of 22 brains of the Yale cohort for which we had a complete dataset for the bulks and fractions, eight (six non-hypermutable) showed a statistically significant difference between the cortex and striatum (Figs. 2D and S5A,B). Except for the hypermutable brain NC7, a higher VAF of somatic mutations was always observed in the cortex. Even in NC7, the VAFs were higher in cortex when excluding neuronal fractions (Fig. 2A, S5C). Similarly, in 5 out of 13 non-hypermutable brains processed by the Lieber Institute (6 SCZ and 7 neurotypical brains), we also observed a higher VAF in cortex as compared to the hippocampus (Figs. 2D and S5D,E). Combining the brains with biased VAFs from the two datasets, we observed a significant overrepresentation of brains with higher VAF values in the cortex (p-value = $5x10^{-4}$ by binomial test). A possible explanation of this observation is that the founder population of the cortex is allocated with fewer earlier lineages, so that the frequency of each lineage is, on average, higher. Alternatively, the cortex may have a higher propensity for expansion in some lineages, when compared to the striatum and hippocampus.

Relevance of somatic mutations to genome function

Given the mostly random distribution of somatic mutations in the human genome, the majority (~60%) of the mutations were outside genes with roughly the same distribution across the genome in each cohort (Figs. 3A and S6A). Approximately 2% to 3% of mutations were in the coding part of the genome and another 6% had the potential to affect regulatory regions. Of them, six putative deleterious mutations in genes previously associated with neuropsychiatric diseases were found in ASD and SCZ brains (Table 1). For example, in the UMB4231 ASD brain, we detected and validated a missense mutation in *PCDH15* – a gene encoding a protein mediating calcium-dependent cell-cell adhesion,

which was previously associated with ASD (18). Furthermore, in the AN05983 ASD brain, we detected and validated a splice mutation in *MTOR*, a gene which is part of the IGF1/PI3K signaling pathway implicated in ASD (19).

We did not detect deleterious somatic mutations in TS associated genes but a missense mutation in the *ARHGEF6* gene (chrX:135,761,734 G>T) was found in brain TS9 at a relatively high VAF in both the cortex (4.2% VAF) and striatum (1.6% VAF). The *ARHGEF6* gene has been associated with X-linked intellectual disability and dendrite orientation/cell polarity (20). Indeed, knockout of *ARHGEF6* in mouse models results in hippocampal dendrite/synaptic abnormalities and deficits in learning (21). Deficits in cell polarity have been implicated in Tourette syndrome from analyses of rare de novo mutations (22).

We next investigated the possible functional impact of the discovered mutations in non-hypermutable brains on enhancer-like elements active in the fetal human brain (23) and in an organoid model of developing human brain (**Methods**). For each set of enhancer-like elements, analysis of the ASD cohort revealed an excess, as compared to controls, of somatic mutations creating putative transcription factor binding motifs (p-value $< 10^{-4}$ by binomial test) (Fig. 3B and S6C). Top-ranked mutations in ASD brains created 16 putative binding motifs for the MEIS1, MEIS2 and MEIS3 transcription factors (Fig. 3C); in normal brains only 4 of such mutations were detected (not shown). Those TFs were also among the top-ranked when simulating random distribution of mutations in ASD brains across the genome (see **Methods**).

Somatic structural mutations

To call structural mutations, we used an approach implemented in CNVpytor (Fig. 4). The tool segments the genome, considering both depth of coverage and split in allele frequencies for single nucleotide polymorphisms (SNPs), allowing us to call deletions, duplications, and copy number neutral losses of heterozygosity (CNN-LOH). The called mutations can be either inherited or somatic and we distinguished between these two types by their cell frequency (**Methods**). We excluded brain LIBD82 as having aneuploidies and few other brains based on coverage and data quality. We focused on a total of 9 high confidence calls encompassing an entire chromosome or for which we could resolve breakpoints (Tables S4 and S5; Figs. S7–S21). For five mutations, we attempted validation in brain tissue by PCR across the breakpoints, which confirmed all of them (Figs. 4B and S7). Two more were validated previously (24).

In total, we detected 9 rearrangements in 9 (6%) non-hypermutable brains ranging from 70 kbp to 16.5 mbp in length, with the largest mutation representing the loss of chromosome Y. That loss of the chromosome and another deletion were detected only in ASD brains. Seven somatic structural mutations were duplications, which were found in about 5% normal and diseased brain. The two duplications discovered in the Tourette cohort (brain TS1 and NC6, the only brains for which more than one region was available) were present in all bulk samples and fractions analyzed (Figs. 4A and S7A). This finding is consistent with the idea that all these duplications occurred during development, which is also supported by their frequency in at least 10% of cells. Consistent with a developmental origin, sequence

microhomologies (1-4 bps) were found in most of the resolved breakpoints (Fig. S22), suggesting that the corresponding mutations were generated by replicative mechanisms, as previously observed for mosaic CNA in the fetal human brain (25). In many aspects, duplications resemble *de novo* CNVs in the CNV Mutator Phenotype (26) – in that they are mostly tandem duplications of early developmental origin and have sequence microhomologies at breakpoints – although the mutations detected here were smaller than reported in previous studies. Thus, we hypothesize that somatic duplications represent background mutations that arise in early development, with the CNV mutator phenotype representing the extreme manifestation of this phenomenon leading to a disease phenotype.

Discussion

Our study revealed a complex etiology of somatic mutations in brain and their possible relationship with phenotype. While most of the brains had just a few dozen mutations of likely developmental origin in proliferating cells, ~6% of brains were classified as hypermutable. Hypermutability was not related to disease, but increased with age, reaching at least a 3% population frequency (95% confidence interval) for brains over 40 years old. The exact cause of hypermutability is to be determined, but besides possible higher mutation rate, in some brains it originates from expansion of a cell lineage. Based on the association of hypermutability with age and its frequent localization in one brain region, we hypothesize that it could be related to local clonal gliogenesis in adult brains, consistent with previous reports of increased glial fraction with age (27, 28). Such gliogenesis could be relevant to cognitive aging, such as Parkinson's and Alzheimer's Disease. Alternatively, hypermutability could be a remnant of a brain tumor escaped earlier in life.

Half of the hypermutable brains carried mutations in genes that have been involved in clonal hematopoiesis in aging adult (29), consistent with previous observations from the analysis of a panel of genes (30). Thus, it is tempting to speculate that increases in clonal hematopoiesis with age (29), coupled with a leakier blood-brain barrier with age and extravasation of blood in brain tissue (31), could explain the higher mutation burden in those brains. Analysis of expression data in brain NC7 is consistent with an increased fraction of blood cells in that brain. However, analysis of copy number for loci of immunoglobulin receptors in B and T cells does not reveal an obvious presence of immune cells in the clonally expanded lineage (Fig. S23). Furthermore, in that brain all mutations were present in two regions and were enriched in post-mitotic neuronal cell fractions, giving support to yet another possibility, that of a neural lineage clonal expansion during development. In support of this scenario, a spectrum of *de novo* variants in *DNMT3A*, which overlaps with those found in hematologic malignancies, causes Tatton-Brown-Rahman syndrome, a developmental disorder with macrocephaly (32). It was also recently shown that human interneurons, differently than other neurons, continue to expand until birth (33). Nonetheless, establishing the cause(s) and extent of the phenomenon of hypermutability requires analysis of larger sets of brains, and additional analyses at the single cell level.

Somatic structural mutations were detected in ~7% of brains, but since most of them (~5%) were duplications, they are unlikely to have functional consequences. The ASD cohort was the only one where deletions were detected. Additional CNV calls (without resolved

breakpoint but with confirmed haplotype imbalance) were also made only in the ASD cohort (Table S4), perhaps pointing to the relevance of somatic deletions for the ASD phenotype. Though, further studies of larger ASD and other disease cohorts are necessary to make definitive conclusions on this observation.

In ASD brains, somatic point mutations within noncoding regions created putative TFs binding motifs within enhancer-like elements that are active in the developing human brain. We hypothesize that some of the newly created motifs result in binding of their corresponding TFs, thereby affecting TF protein dosage and dysregulating gene regulatory networks. The top-ranked affected motifs are putative binding sites for MEIS genes, which are homeodomain-containing transcriptional activators that promote chromatin decompaction (34). MEIS TFs have been implicated in intellectual disabilities, act as cofactors of HOX genes and are regulators of proliferation, growth, neurogenesis and patterning during development (34). MEIS2 is highly expressed in cortical development and was defined as a marker of a subpopulation of cortical interneurons that populates the white matter (35). We previously proposed that deleterious de novo ASD variants are enriched in homeodomain TF binding motifs within developmental enhancer-like elements (36). Similarly, enhancer-like elements associated with ASD genes from the SFARI collection are enriched for homeobox binding motifs (37). Therefore, we suggest that functional studies of mutations within homeodomain binding motifs during human development might aid efforts in understanding the etiology of ASD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

The authors thank the Tourette's Association of America for collecting and providing brain tissue. We also thank the donors and their families. We acknowledge the NIH Neurobiobank and its Director, Dr. Sabina Berretta, for providing postmortem tissue from TS and NC individuals. We are grateful to Geoff Lyon of the Yale Flow Cytometry Facility for help with FACS sorting of brain postmortem nuclei. Some tissues were obtained from Autism BrainNet, a resource of the Simons Foundation Autism Research Initiative (SFARI). Autism BrainNet also manages the Autism Tissue Program (ATP) collection, previously funded by Autism Speaks. We are grateful and indebted to the families who donated tissue for research purposes to Autism BrainNet and the ATP. We thank member of BSMN consortium for useful discussions and suggestions during the study and preparation of the manuscript. Members of the BSMN consortiums are listed in the supplementary file. We express our gratitude to the NDA and Synapse team for providing collaborative data storage space and support for data transfer needed for the successful completion of this study.

Funding:

National Institutes of Health grant U24CA220242

National Institutes of Health grant U01MH106892

National Institutes of Health grant U01MH106876

National Institutes of Health grant U01MH106883

National Institutes of Health grant U01MH106874

National Institutes of Health grant U01MH106893

National Institutes of Health grant U01MH106882

National Institutes of Health grant U01MH108898

National Institutes of Health grant U01MH106891

National Institutes of Health grant U01MH106884

National Institutes of Health grant R01MH109648

Simons Foundation Autism Research Initiative grant 399558

The Urban laboratory receives funding through the Jaswa Innovator Award and from B. Blackie and W. Mclvor.

Data and materials availability:

WGS data and call sets have been deposited in the NIMH Data Archive (NDA Study ID 814; DOI: 10.15154/1506068) and can be accessed as part of the NIMH Data Archive permission groups: https://nda.nih.gov/user/dashboard/data_permissions.html. Data for brain organoids are available at NIMH Data Archive (NDA collection ID 2424; https://nda.nih.gov/edit_collection.html?id=2424). The BSMN mutation calling workflow is available on GitHub (https://github.com/bsmn/bsmn-pipeline) and Zenodo (doi: https://doi.org/10.5281/zenodo.4321679). CNVpytor is available on GitHub (https://github.com/abyzovlab/CNVpytor) and on bio.tools (ID: cnvpytor). The code for data analysis is freely available on Zenodo (doi: https://doi.org/10.5281/zenodo.6374179).

References and Notes

1. Bae T et al., Science (New York, NY). 359, 550–555 (2018).
2. Fasching L et al., Science. 371, 1245–1248 (2021). [PubMed: 33737484]
3. Lodato MA et al., Science (New York, NY). 359, 555–559 (2018).
4. Abascal F et al., Nature. 593, 405–410 (2021). [PubMed: 33911282]
5. McConnell MJ et al., Science (New York, NY). 356, eaal1641 (2017).
6. Rodin RE et al., Nat Neurosci. 24, 176–185 (2021). [PubMed: 33432195]
7. Network BSM et al., Genome Biol. 22, 92 (2021). [PubMed: 33781308]
8. Bizzotto S et al., Science. 371, 1249–1253 (2021). [PubMed: 33737485]
9. Crespo I et al., J Mol Diagnostics. 13, 634–647 (2011).
10. Gerstung M et al., Nature. 578, 122–128 (2020). [PubMed: 32025013]
11. Tate JG et al., Nucleic Acids Res. 47, gky1015- (2018).
12. Bailey MH et al., Cell. 173, 371–385.e18 (2018). [PubMed: 29625053]
13. Sondka Z et al., Nat Rev Cancer. 18, 696–705 (2018). [PubMed: 30293088]
14. Rees E et al., Nat Neurosci. 23, 179–184 (2020). [PubMed: 31932766]
15. Howrigan DP et al., Nat Neurosci. 23, 185–193 (2020). [PubMed: 31932770]
16. Pardiñas AF et al., Nat Genet. 50, 381–389 (2018). [PubMed: 29483656]
17. Ripke S et al., Nature. 511, 421–427 (2014). [PubMed: 25056061]
18. Anney R et al., Hum Mol Genet. 21, 4781–4792 (2012). [PubMed: 22843504]
19. Zhang W et al., Gene Dev. 34, 580–597 (2020). [PubMed: 32115408]
20. Meseke M, Rosenberger G, Förster E, Eur J Neurosci. 37, 1404–1412 (2013). [PubMed: 23406282]
21. Ramakers GJA et al., Hum Mol Genet. 21, 268–286 (2012). [PubMed: 21989057]
22. Wang S et al., Cell Reports. 24, 3441–3454.e12 (2018). [PubMed: 30257206]

23. Boix CA, James BT, Park YP, Meuleman W, Kellis M, Nature. 590, 300–307 (2021). [PubMed: 33536621]
24. Sherman MA et al., Nat Neurosci. 24, 197–203 (2021). [PubMed: 33432194]
25. Sekar S et al., Genome Res. 30, 1695–1704 (2020). [PubMed: 33122304]
26. Liu P et al., Cell. 168, 830–842.e7 (2017). [PubMed: 28235197]
27. Conde JR, Streit WJ, J Neuropathology Exp Neurology. 65, 199–203 (2006).
28. Beach TG, Walker R, McGeer EG, Glia. 2, 420–436 (1989). [PubMed: 2531723]
29. Zink F et al., Blood. 130, 742–752 (2017). [PubMed: 28483762]
30. Keogh MJ et al., Nat Commun. 9, 4257 (2018). [PubMed: 30323172]
31. Banks WA, Reed MJ, Logsdon AF, Rhea EM, Erickson MA, Nat Aging. 1, 243–254 (2021). [PubMed: 34368785]
32. Shen W et al., Am J Med Genet A. 173, 3022–3028 (2017). [PubMed: 28941052]
33. Paredes MF et al., Science. 375, eabk2346 (2022). [PubMed: 35084970]
34. Schulte D, Geerts D, Development. 146, dev174706 (2019). [PubMed: 31416930]
35. Frazer S et al., Nat Commun. 8, 14219 (2017). [PubMed: 28134272]
36. Amiri A et al., Science (New York, NY). 362, eaat6720 (2018).
37. Trevino AE et al., Science (New York, NY). 367, eaay1645 (2020).
38. Vonsattel JPG, del Amaya MP, Keller CE, Acta Neuropathol. 115, 509–532 (2008). [PubMed: 17985145]
39. Matevossian A, Akbarian S, Journal of visualized experiments : JoVE, e914–e914 (2008).
40. Arlotta P, Molyneaux BJ, Jabaudon D, Yoshida Y, Macklis JD, J Neurosci. 28, 622–632 (2008). [PubMed: 18199763]
41. Dou Y et al., Nature biotechnology. 14, 307–6 (2020).
42. Gori K, Baez-Ortega A, Biorxiv, 372896 (2020).
43. McLaren W et al., Genome biology. 17, 122 (2016). [PubMed: 27268795]
44. Yu D et al., Am J Psychiat. 176, 217–227 (2019). [PubMed: 30818990]
45. Lennington JB et al., Biol Psychiat. 79, 372–382 (2016). [PubMed: 25199956]
46. Collado-Torres L et al., Neuron. 103, 203–216.e8 (2019). [PubMed: 31174959]
47. Bray NL, Pimentel H, Melsted P, Pachter L, Nat Biotechnol. 34, 525–527 (2016). [PubMed: 27043002]
48. Soneson C, Love MI, Robinson MD, F1000research. 4, 1521 (2016).
49. Xie X et al., Natl Sci Rev. 8, nwaa180 (2020). [PubMed: 34691592]
50. Winkler EA et al., Science, eabi7377 (2022). [PubMed: 35084939]
51. Ernst J, Kellis M, Nature methods. 9, 215–216 (2012). [PubMed: 22373907]
52. Consortium RE et al., Nature. 518, 317–330 (2015). [PubMed: 25693563]
53. Visel A et al., Cell. 152, 895–908 (2013). [PubMed: 23375746]
54. Fornes O et al., Nucleic Acids Res. 48, D87–D92 (2019).
55. Suvakov M, Panda A, Diesh C, Holmes I, Abyzov A, Gigascience. 10, giab074 (2021). [PubMed: 34817058]
56. Chen X et al., Bioinformatics (Oxford, England). 32, 1220–1222 (2016).
57. Consortium EP et al., Nature. 489, 57–74 (2012). [PubMed: 22955616]

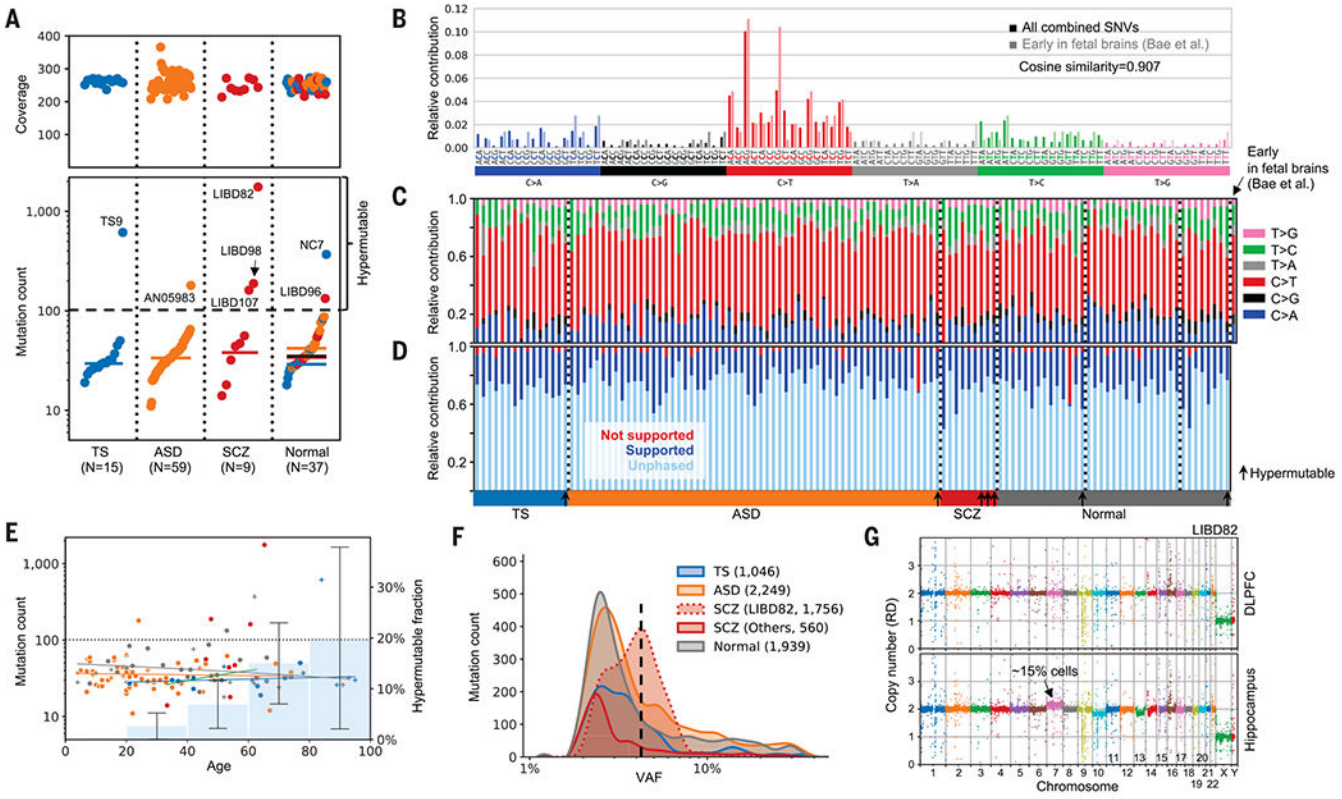

Fig. 1. Mutations discovered across cohorts of brains.

A) Summary of coverage and detected somatic point mutations across cohorts using bulk brain samples. Mutation burden across cohorts was comparable when excluding hypermutable brains with more than 101 of mutations (named in the plots). Three colors represent institutions that processed the brain samples: Yale (blue), Harvard (orange), Lieber (red). **B)** The substitution spectra of detected mutations across all brains are comparable. **C)** Fractions of supported and not supported calls by assigning to haplotypes are comparable across brains, indicating the same accuracy of calls for hypermutable (indicated by arrows) and non-hypermutable brains. Due to the short DNA fragments (~450 bps), only ~20% of calls can be assigned to a haplotype using nearby heterozygous SNP; other calls are indicated as unphased. **D)** The combined mutation spectrum is dominated by C>T transitions in CpG motifs and matches the spectrum of developmental mutations (Fig. S2). **E)** When excluding hypermutable samples, the mutation burden does not correlate with age. However, there are more hypermutable brains (those are above the dotted line) in older brains (blue histogram with error bars in the background). Each data point is an individual brain with colors representing phenotypes: TS (blue), ASD (orange), SCZ (red), and normal (grey). Circles are males, crosses are females. **F)** Distributions of mutation counts and allele frequencies across cohorts. Numbers in parenthesis list the mutation count. Brain LIBD82 is an exception and its mutations had higher frequencies than in other brains. The VAF value of the vertical dashed line corresponds to ~15% of aneuploid cells in the hippocampus of that brain (which corresponds to displayed average VAF of ~4% between

cortex and hippocampus). **G)** Brain LIBD82 had apparent aneuploidies in ~15% of cells in the hippocampus, which may be a signature of incipient glioma or glioblastoma.

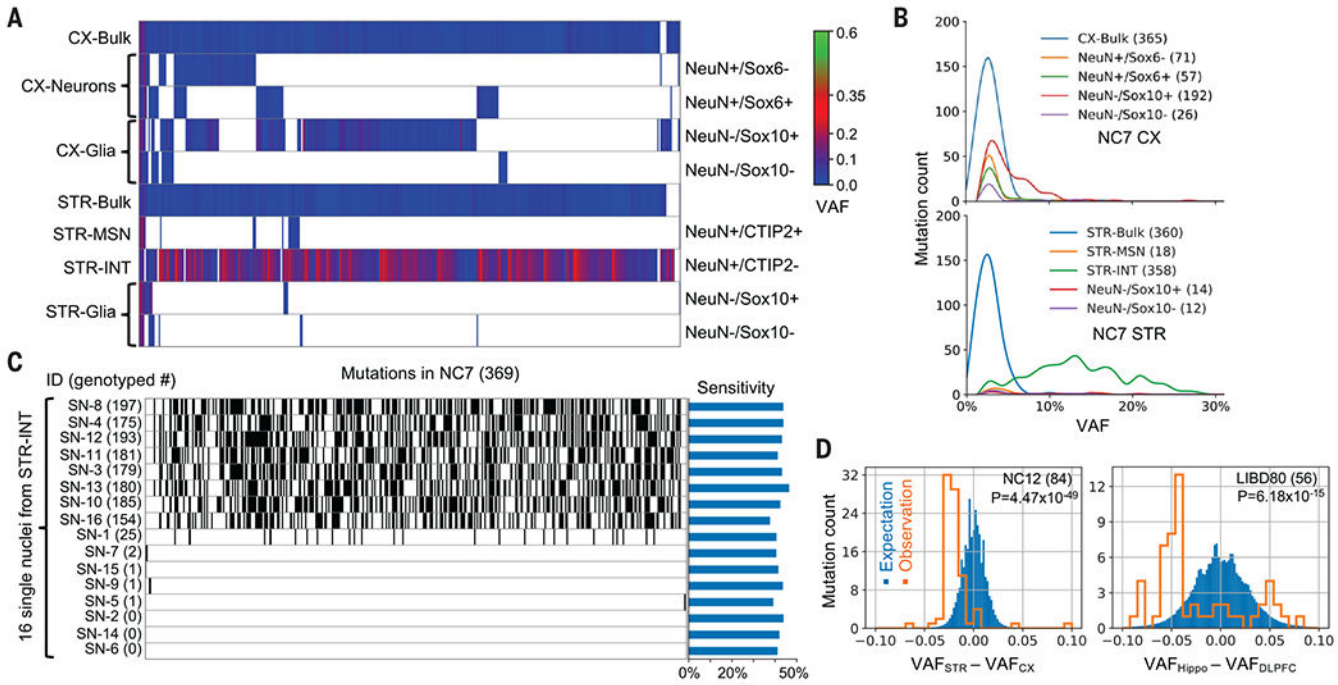

Fig. 2. Uneven cell lineage distribution of somatic mutations in brain.

A) Mutation allele frequencies in samples from brain NC7. Almost all mutations discovered in NC7 are present in the striatal interneuron fraction (STR-INT) at high VAFs. **B)** Distribution of mutation allele frequencies across samples in NC7. **C)** Genotyping of mutations in 16 single nuclei originated from the NC7 STR-INT fraction by sequencing at 5X depth. Black bars represent genotyped mutations. **D)** Examples of brains where frequencies of mutations are significantly biased toward cortex as compared to striatum (NC12) or as compared to hippocampus (LIBD80).

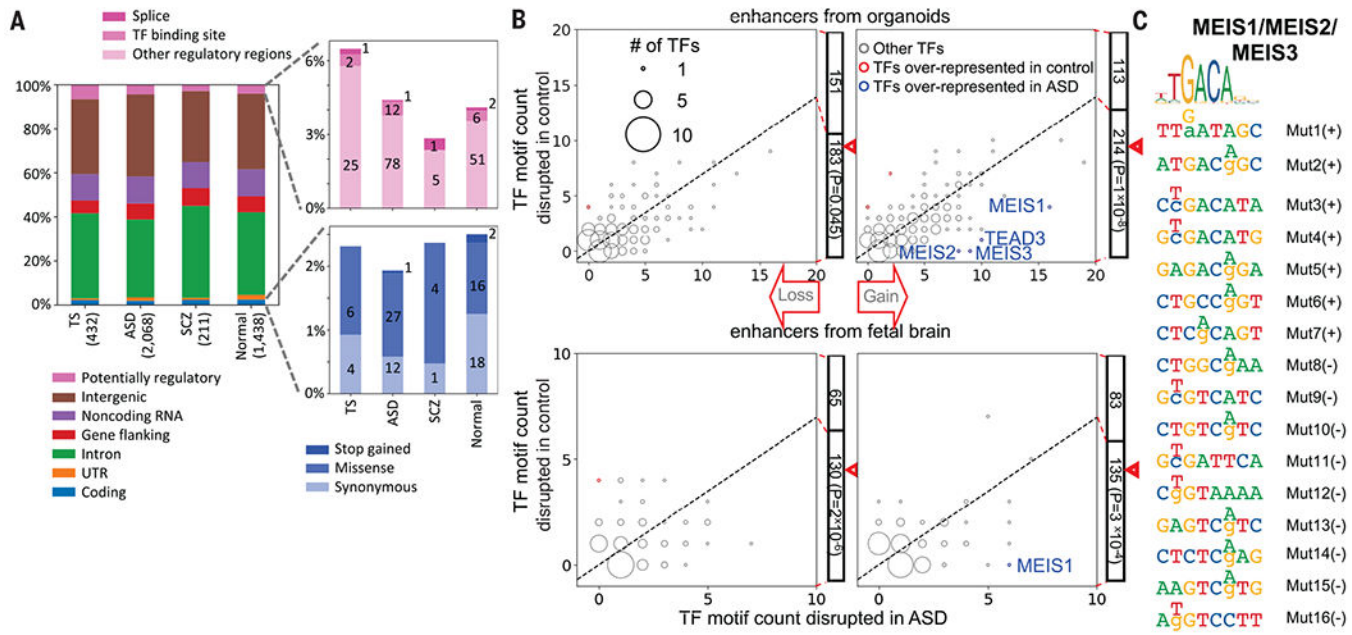

Fig. 3. Relevance of somatic mutations to genome function.

A) In each cohort mutations from non-hypermutable brains have roughly the same distribution across genomes and similar predicted functional impacts by VEP. **B)** Counts of putative TF motifs disrupted by somatic mutations within enhancer-like elements from organoids (upper plots) or fetal brain (lower plots), showing that mutations significantly disrupt more putative TF binding sites in ASD vs controls. Each circle represents the number of TFs with x (count of mutations in ASD) and y coordinates (count of mutations in controls) of the circle. **C)** Somatic mutations in the ASD cohort predicted to lead to gain of binding sites for MEIS1, MEIS2, and MEIS3 in enhancer-like elements. The consensus motif is on top. Mutations on positive (+) or negative (−) strand are enumerated on the right with coordinates listed in Table S3. For each mutation, a reference base is show in small letter and the mutation base is noted by a capital letter above.

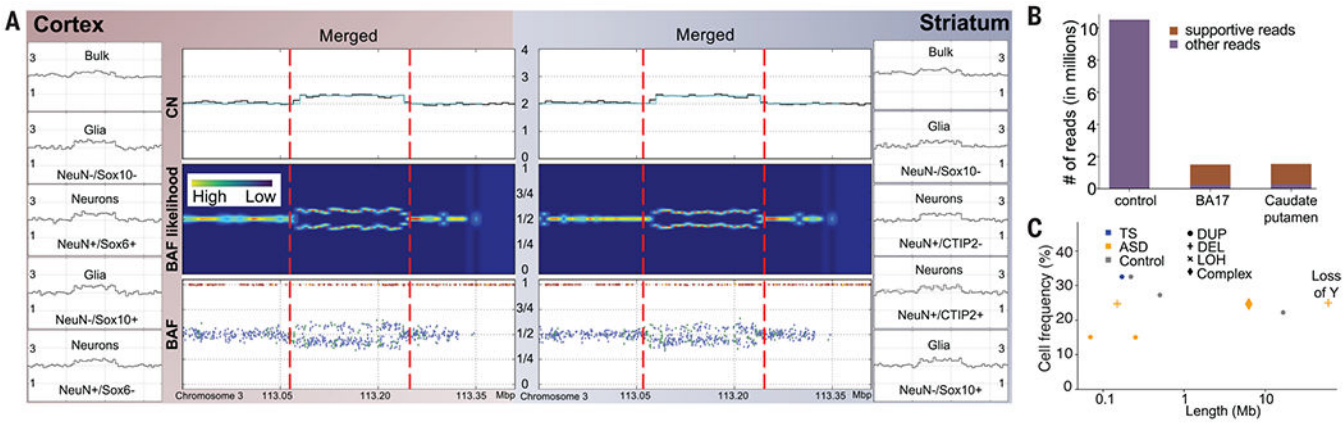

Fig. 4. Detection and validation of somatic large structural mutations.

A) Example of a duplication on chromosome 3 (chr3:113,067,261-113,233,476) identified in brain TS1 in eight cell fractions and two bulks: cortex and striatum. Red vertical lines outline the region of the duplication. The two merged plots combine data for bulk and fractions for cortex (left) and striatum (right) and present an increase in the read depth corresponding to the cell frequency of ~32% (top), split in BAF likelihood function (brighter color corresponds to higher values) matching to the same cell frequency (middle) and BAF for individual SNPs (bottom). SNPs in the accessible genome (P-bases) are in blue, other SNPs are in green. **B)** Amplicon-seq validation of the duplication. Bar plot shows the number of reads mapping to the junction of the duplication in the cortical region BA17 and caudate putamen of TS1 and in control sample. **C)** Summary of size and cell frequency of all structural mutations detected from all analyzed individuals.

Table 1.

List of detected somatic mutations that putatively affect genes previously implicated in cancers and genes previously associated with neuropsychiatric diseases.

Some mutations are also annotated as pathogenetic in the COSMIC database (11). A known mutation in *NRAS* oncogene is present in two brains.

Mutation	Brain	Phenotype	Hypermutable	VAF	Validation**	Consequence (VEP)	Gene	Gene annotation***	Genomic Mutation ID (COSMIC)	Prec (COS
In genes implicated in cancer										
chr1:115,258,747 C>T	NC7	Normal	Yes	1.84%	A, S	Missense	*NRAS*	B, C	COSV54736383	Path
	TS9	TS	Yes	1.10%	A					
chr2:25,466,766 C>G	NC7	Normal	Yes	1.91%	A, S	Splice	*DNMT3A*	B, C	-	
chr3:81,695,552 G>T	TS9	TS	Yes	1.26%	n/a	Missense	*GBE1*	-	COSV101450116	Path
chr4:106,156,747 C>T*	TS9	TS	Yes	1.05%	A	Stop	*TET2*	B, C	COSV54395664	
chr4:106,197,392 G>T			Yes	1.75%	A	Stop			-	
chr4:183,267,873 C>T	TS9	TS	Yes	1.27%	n/a	Missense	*TENM3*	-	COSV69301741	Path
chr12:48,107,157 C>G	TS9	TS	Yes	1.16%	n/a	Missense	*ENDOU*	-	COSV57465821	Path
chr15:90,631,934 C>T*	NC7	Normal	Yes	3.55%	S	Missense	*IDH2*	B, C	COSV57468751	Path
chrX:129,168,460 C>T	NC7	Normal	Yes	2.99%	S	Stop	*BCORL1*	C	COSV54390206	
chr1:11,300,357 T>C	AN05983	ASD	Yes	2.63%	A	Splice	*MTOR*	B, C	-	
chr10:32,326,503 A>T*	LIBD99	Normal	No	2.62%	n/a	Missense	*KIF5B*	C	-	
chr6:168,315,889 C>T*	TS1	TS	No	0.95%	A	Stop	*MLLT4*	C	COSV60045383	Path
chr4:1,977,081 G>A	UMB797	ASD	No	2.78%	n/a	Missense	*WHSC1*	B, C	COSV56389769	Path
chr7:148,495,674 A>G	UMB5297	ASD	No	2.99%	n/a	Missense	*CUL1*	B	-	
In disease-implicated genes in brains with diseases										
chr1:11,300,357 T>C	AN05983	ASD	Yes	2.63%	A	Splice	*MTOR*	ASD	-	
chr2:179,612,387 T>A	UMB1638	ASD	No	2.09%	n/a	Stop	*TTN*	ASD	-	
chr8:146,033,199 C>T	AN13654	ASD	No	4.15%	n/a	Missense	*ZNF517*	ASD	-	
chr10:55,600,099 C>T	UMB4231	ASD	No	4.41%	A	Missense	*PCDH15*	ASD	COSV57387920	Path
chr16:20,976,490 C>T	AN09412	ASD	No	12.09%	R	Missense	*DNAH3*	ASD	-	

Author Manuscript

Author Manuscript

Author Manuscript

Mutation	Brain	Phenotype	Hypermutable	VAF	Validation**	Consequence (VEP)	Gene	Gene annotation***	Genomic Mutation ID (COSMIC)	Prec (COS
chr2:73,800,041 T>G*	LIBD100	SCZ	No	4.46%	n/a	Missense	*ALMS1*	SCZ	-	

* tier 2 mutation (see **Methods**); Validation:

** A – by amplicon, R – previously reported and validated in Rodin et al. (6), S – by single cells; Gene annotation:

*** B – cancer implicated genes in Bailey et al. (12), C – cancer gene census (13), ASD – SFARI genes, SCZ – compiled from previous SCZ studies (14–17).