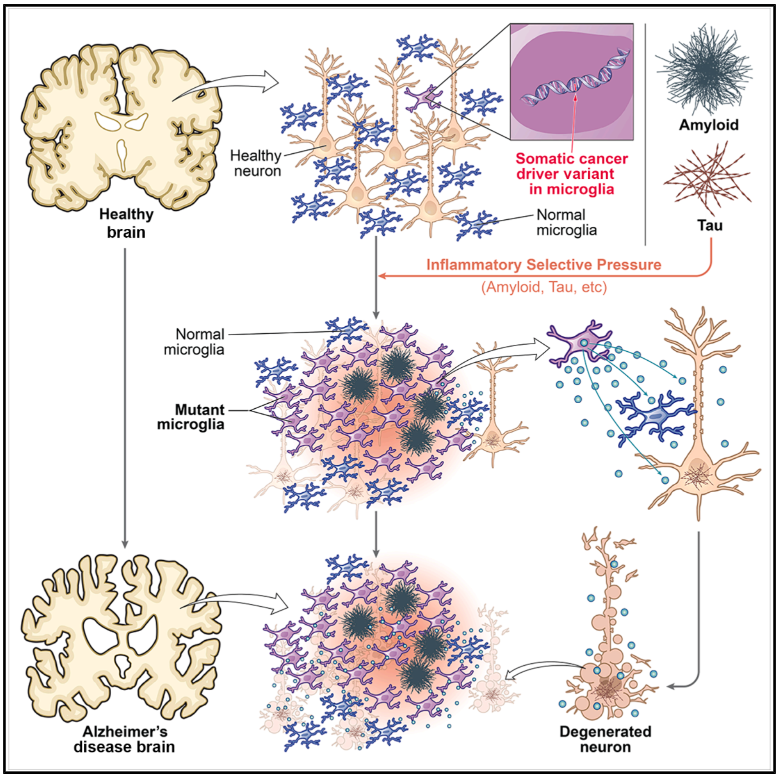
Cell

Article

Somatic cancer variants enriched in Alzheimer’s disease microglia-like cells drive inflammatory and proliferative states

Graphical abstract



Highlights

- Somatic variants linked to clonal hematopoiesis are elevated in Alzheimer’s brains
- These driver variants are carried by microglia-like macrophages across brain regions
- Mutant microglia-like macrophages display disease-related signatures *in vivo*
- iPSC-derived microglia-like cells with driver mutations recapitulate these signatures

Authors

August Yue Huang, Zinan Zhou, Maya Talukdar, ..., Eirini P. Papapetrou, Eunjung Alice Lee, Christopher A. Walsh

Correspondence

yue.huang@childrens.harvard.edu (A.Y.H.), samuele.marro@mssm.edu (S.G.M.), eirini.papapetrou@mssm.edu (E.P.P.), ealice.lee@childrens.harvard.edu (E.A.L.), christopher.walsh@childrens.harvard.edu (C.A.W.)

In brief

Cancer-associated somatic variants are enriched in the brain immune cells of Alzheimer’s disease (AD) patients, driving them toward inflammation and proliferation, which may contribute to neuroinflammation and neurodegeneration during AD progression.

Huang et al., 2026, Cell *189*, 1–17
June 11, 2026 © 2026 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2026.03.040

CellPress

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

Cell

CellPress
OPEN ACCESS

Article

Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states

August Yue Huang,[1,2,3,22,*] Zinan Zhou,[1,2,3,22] Maya Talukdar,[1,2,3,4,22] Liz Enyenihi,[1,2,3,4,22] Michael B. Miller,[1,2,3,5] Brian Chhouk,[1] Ila Rosen,[1] Mengyue Zheng,[1,2,3] Minye Zhou,[1] Averill Yang,[1] Edward Stronge,[1,2,3,4] Madel Durens,[6] Minh Nguyen,[7,8,9,10] Jaejoon Choi,[1,2,3] Boxun Zhao,[1,2,3] Sattar Khoshkhoo,[1,2,3,11] Junho Kim,[1,2,3,12] Rebecca Andersen,[1,2,3] Zheming An,[1,2,3] Yuchen Cheng,[1,2] Javier Ganz,[1,2,3] Levan Mekerishvili,[13,14,15] Kyle J. Travaglini,[16] Mariano I. Gabitto,[16] Rebecca D. Hodge,[16] Eitan S. Kaplan,[16] Julia A. Belk,[17] Dan Landau,[13,14] Ed S. Lein,[16] Philip L. De Jager,[18] David A. Bennett,[19] Samuele G. Marro,[6,8,10,*] Eirini P. Papapetrou,[7,8,9,10,*] Eunjung Alice Lee,[1,2,3,*] and Christopher A. Walsh[1,2,3,20,21,23,*]

[1]Division of Genetics and Genomics and Manton Center for Orphan Diseases, Boston Children's Hospital, Boston, MA, USA
[2]Department of Pediatrics, Harvard Medical School, Boston, MA, USA
[3]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[4]Harvard-MIT MD/PhD Program, Boston, MA, USA
[5]Division of Neuropathology, Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA
[6]Nash Family Department of Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[7]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[8]Center for Advancement of Blood Cancer Therapies, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[9]Tisch Cancer Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[10]Department of Stem Cell Biology and Regenerative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[11]Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA
[12]Department of Biological Sciences, Sungkyunkwan University, Suwon, South Korea
[13]New York Genome Center, New York, NY, USA
[14]Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA
[15]Physiology, Biophysics and Systems Biology Graduate Program, Weill Cornell Medicine, New York, NY, USA
[16]Allen Institute for Brain Science, Seattle, WA, USA
[17]Department of Computer Science, Stanford University, Stanford, CA, USA
[18]Center for Translational and Computational Neuroimmunology, Department of Neurology and The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Center, New York, NY, USA
[19]Rush Alzheimer's Disease Center, Rush University Medical College, Chicago, IL, USA
[20]Howard Hughes Medical Institute, Boston, MA, USA
[21]Department of Neurology, Harvard Medical School, Boston, MA, USA
[22]These authors contributed equally
[23]Lead contact
*Correspondence: yue.huang@childrens.harvard.edu (A.Y.H.), samuele.marro@mssm.edu (S.G.M.), eirini.papapetrou@mssm.edu (E.P.P.), ealice.lee@childrens.harvard.edu (E.A.L.), christopher.walsh@childrens.harvard.edu (C.A.W.)
https://doi.org/10.1016/j.cell.2026.03.040

SUMMARY

Alzheimer's disease (AD) is a neurodegenerative condition characterized by microglia-mediated neuroinflammation. Deep (>1,000×) panel sequencing of 311 brain samples revealed enrichment of somatic single-nucleotide variants (sSNVs) in cancer driver genes in AD brains, especially in genes associated with clonal hematopoiesis (CH). These sSNVs were associated with clonal expansion and carried by both microglia-like brain macrophages (MLBMs) in multiple brain regions as well as paired blood, suggesting a likely hematopoietic origin. Single-nucleus RNA sequencing data from 62 additional AD and control brains revealed increased somatic copy number variants (sCNVs) associated with CH in AD MLBMs, whereas single-cell multi-omic analyses demonstrated that sSNV- and sCNV-carrying MLBMs exhibited inflammatory and proliferative transcriptional signatures characteristic of disease-associated microglia. These signatures were recapitulated in induced pluripotent stem cell-derived microglia-like cells with *TET2*, *ASXL1*, and *DNMT3A* variants. These findings suggest that clonal somatic driver variants in MLBMs are enriched in AD, potentially promoting neuroinflammation and neurodegeneration.

Cell *189*, 1–17, June 11, 2026 © 2026 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

INTRODUCTION

The importance of microglia in Alzheimer's disease (AD) pathogenesis has been demonstrated by large-scale genetic association studies, which have identified AD risk variants in a growing list of microglia-related genes.[1–4] As the primary macrophages in the central nervous system (CNS), microglia originate from erythro-myeloid progenitors in the yolk sac during early embryogenesis.[5] Despite comprising only about 5% of cells in the cortex,[6] they play critical roles in brain development, injury response, pathogen defense, and modulating cellular responses involved in aging and neurodegeneration.[7–10] Once abnormally reactive in AD, microglia can promote synaptic and neuronal loss while exacerbating tau proteinopathy.[11,12] Moreover, recent studies have reported that during aging or neurodegeneration, bone marrow-derived blood monocytes can infiltrate the brain through a disrupted blood-brain barrier (BBB) and differentiate into microglia-like cells, exhibiting transcriptional and functional similarity to brain-resident microglia.[13–15]

Somatic variants accumulate in all cell types that have been studied, both during normal development and during aging.[16–18] Clonal expansion, driven by somatic variants in genes regulating cell proliferation and fitness, is considered the major cause of cancer, but has also been recently reported in various non-cancerous cells, often in the absence of visible pathology.[19] Clonal expansion of blood cells, called clonal hematopoiesis (CH), increases in prevalence with age and is associated with increased risk of hematologic malignancies, but also unexpectedly with increased risk of cardiovascular disease,[20,21] likely through inflammatory effects of mutant cells.[22] Clonal hematopoiesis of indeterminate potential (CHIP, denoting somatic single-nucleotide variants [sSNVs] and somatic insertions and deletions [sIndels] in cancer driver genes) and mosaic chromosomal alterations (mCAs, denoting chromosome-level somatic copy number variants [sCNVs] enriched in CH) represent two primary categories of somatic variants that propel the progression of CH.[23]

Two recent studies investigating CHIP variants primarily using whole-exome sequencing (WES) data from blood cells found no effect[24] or a protective effect[25] on AD risk. However, these studies were limited in their ability to detect somatic variants with low variant allele fractions (VAFs) due to their low sequencing depth (<70×). A more recent study utilizing deep panel sequencing revealed an increased burden of CHIP variants in AD blood samples compared with age-matched controls, primarily driven by low-VAF variants,[26] and this AD-related increase of CHIP variants was independently confirmed in whole-genome sequencing (WGS) of ~30,000 cases and controls.[26] A few small studies of primary brain samples have hinted that cancer-associated variants and clonal overgrowth might occur in brain microglia, which proliferate throughout life.[25,27,28] Moreover, the somatic *BRAF* p.V600E variant, a common cancer driver variant, in the microglial lineage has been causally implicated not in cancer, but instead in degeneration of neurons secondary to clonal expansion and activation of mutant microglia in both mouse models and humans.[29]

Here, we tested directly whether somatic cancer driver variants in the brain are associated with AD by two prospective and orthogonal approaches in 221 AD and 152 control brain samples of two AD cohorts (Figures 1A and 1B). We found consistent increases in somatic driver variants in AD brains compared with controls, as well as specific enrichments in genes previously implicated in CH and other pre-malignant conditions. These somatic variants were enriched in cells transcriptionally indistinguishable from microglia, which we refer to as microglia-like brain macrophages (MLBMs), compared with other brain cell types. Epigenetic and transcriptomic analyses of primary MLBMs harboring CH variants, as well as transcriptomics of *in vitro* induced pluripotent stem cell (iPSC)-derived microglia-like cells harboring CH variants, revealed that mutant cells exhibit pro-inflammatory and proliferative transcriptional signatures that have been previously associated with neurodegeneration (Figures 1B and 1C).

RESULTS

Somatic variants in cancer driver genes are enriched in AD brains

To examine the mutational burden of somatic variants in cancer driver genes in AD, we designed a hybrid capture gene panel covering 149 cancer driver genes with unique molecular identifier (UMI) barcoding (Table S1). We sequenced DNA from the prefrontal cortex (PFC) samples of 190 AD patients and 121 matched controls from the Religious Orders Study/Memory and Aging Project (ROSMAP) cohort at an average sequencing depth of >1,000× after UMI collapsing (Figures 1A, S1A, and S1B; Table S2). By generating a consensus sequence from multiple reads derived from the same original DNA molecule, UMI-based panel sequencing exponentially reduces base-calling errors and detects somatic variants with VAFs as low as 0.1% (Figures S1C and S1D), with much higher sensitivity and precision than previous methods not employing consensus error correction.[30] Using our customized computational pipeline, we identified 901 sSNVs and 18 sIndels from AD and control PFC samples ("sensitive" list; Table S3). We randomly selected 22 sSNVs with a range of VAFs for validation using amplicon sequencing, along with 17 potentially pathogenic sSNVs identified in AD brains that were predicted to be deleterious, and all 10 frameshift sIndels. Thirty-five of 39 (90%) tested sSNVs and 8 of 10 (80%) sIndels successfully validated in newly extracted DNA samples from the corresponding PFC samples, confirming the high accuracy of our somatic variant calling strategy even for those with VAFs as low as 0.1% (Figures S1E–S1G). The mutational spectrum of sSNVs is consistent with the cell division/mitotic clock signature SBS1 (Figure S2A; cosine similarity 0.92), suggesting that these variants predominantly occurred during cell division.

With similar sequencing depth, coverage, and genic distribution between AD and control PFC samples (Figures S1A, S1B, and S2B), we observed that AD brains harbored significantly more sSNVs among the 149 targeted genes than aged-matched controls (Figure 2A; $p = 0.001$, two-tailed proportion test; odds ratio [OR] = 1.3). When we only considered variants that were exclusively present in a single PFC sample ("stringent" list; Table S3), the OR of sSNV burden between AD and control became even more dramatic (Figure 2B; $p = 0.008$, two-tailed

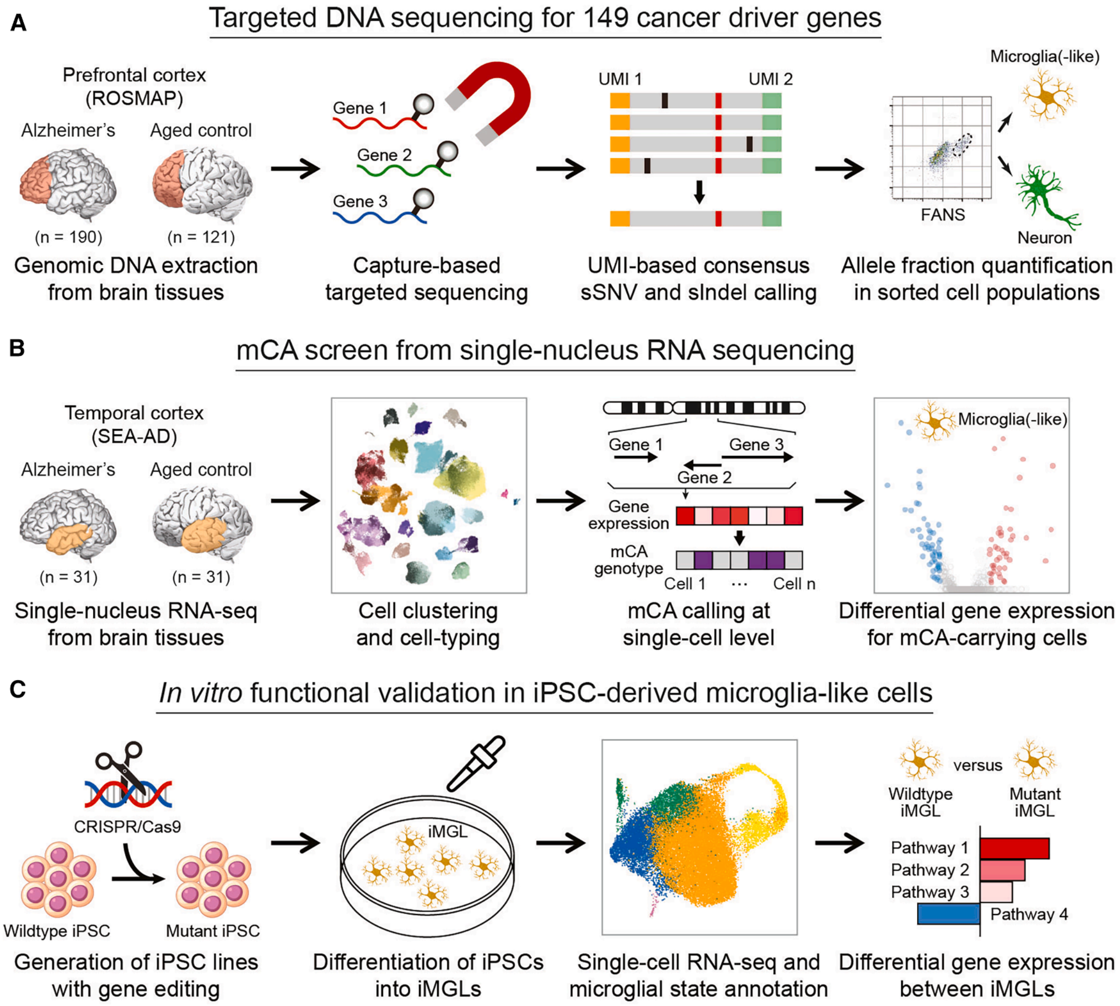


Figure 1. Overview of the experimental and analysis strategies

(A) Profiling sSNVs and sIndels in 311 AD and control PFC samples using deep molecular barcode sequencing with a panel of 149 cancer driver genes. Variant candidates were validated by amplicon sequencing, and their VAFs were measured in different FANS-sorted nuclei populations. FANS, fluorescence-activated nuclei sorting.

(B) Identification and transcriptomic profiling of microglia(-like cells) in AD and control brain snRNAseq samples carrying mCA.

(C) Functional validation of cancer driver variants using iPSC-derived microglia-like cells (iMGLs). Single-cell RNA sequencing (RNA-seq) was performed on mutant iMGLs carrying variants in CH-related genes, and microglial state composition and gene expression changes were compared against wild-type (WT) iMGLs. iPSC, induced pluripotent stem cell.

proportion test; OR = 1.6). This increased burden of sSNVs in our targeted genes in AD remained significant after controlling for confounding factors, including sex, age, sequencing coverage, and post-mortem interval (Figure 2C; $p = 0.03$, linear regression).

In addition to an increased sSNV burden in AD, we also found age was an independent factor positively correlated with sSNV burden (Figure 2C; $p = 0.002$, linear regression) and the proportion of sSNV carriers (Figure S2C), suggesting a likely age-associated accumulation of somatic variants in cancer driver genes in both normal and diseased brains. Previous studies have highlighted the age-related accumulation of low-VAF (<1%–5%) somatic variants in cancer driver genes in blood.[31] Our finding of age-related accumulation in the brain is also consistent with a recent study using deep WGS of a smaller sample size.[32]

When we divided cancer driver genes into (proto-)oncogenes and tumor suppressor genes (TSGs), we observed a greater sSNV burden in AD for TSGs but not for oncogenes (Figure 2D). TSGs canonically increase proliferation when they are inactivated by loss-of-function variants throughout the gene body, whereas oncogenes are usually only activated by specific, recurrent gain-of-function missense variants affecting critical domains. Thus, our results suggest that most sSNVs are associated with AD by a loss-of-function of TSGs. Besides sSNVs, we also observed more frameshift sIndels in AD brains (5 in AD versus 2 in control; Table S3), though this enrichment did not reach significance in this small sample size.

Examination of the mutational burden at the level of individual genes revealed that somatic variants in the top 10 most

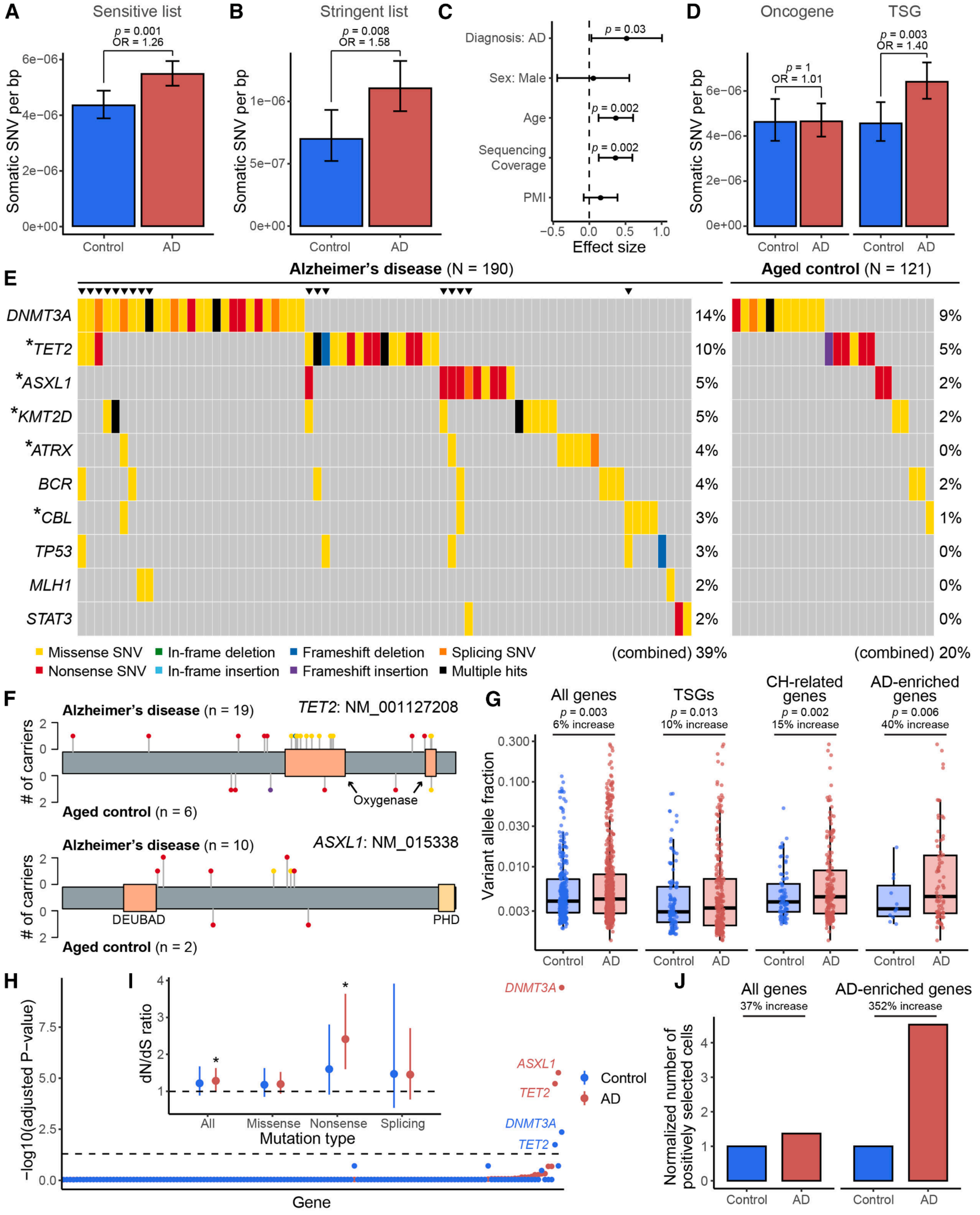


Figure 2. Elevated burdens of somatic variants in cancer driver genes in AD brains

(A and B) AD prefrontal cortex samples harbor significantly more sSNVs in 149 targeted cancer driver genes than matched controls, using both the sSNV list of sensitive (A) and stringent (B) identification pipelines. The sensitive pipeline allows recurrent sSNVs if they were exclusively present or specifically enriched (two-sample Z-test of proportion with $p < 0.05$) in the AD or control groups, whereas the stringent list removes all recurrent sSNVs that were shared by any samples.

(C) Linear regression modeling confirmed that the AD effect on greater sSNV burden remains significant ($p = 0.03$) after controlling for potential confounding factors. In addition to AD status, age is also positively correlated with the sSNV burden ($p = 0.002$). PMI, post-mortem interval.

(D) The significant increase of sSNV burden in AD brains is only observed for tumor suppressor genes (TSGs) but not for (proto-)oncogenes.

(E) Top 10 recurrently mutated genes in AD and control brains. Different types of protein-altering sSNV and sIndel are shown in various colors, where "multiple hits" (black) denotes multiple protein-altering variants in the same gene. Asterisks denote the five genes that contain significantly more somatic variants in AD patients than matched controls ($p < 0.05$, one-tailed proportion test). Triangles highlight individuals who carry variants in multiple genes.

(legend continued on next page)

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

commonly mutated genes were found in 39% of the AD patients, compared with only 20% of the aged controls (Figure 2E). Moreover, only AD brain samples carried variants in multiple genes ($p = 0.0002$, hypergeometric test). Five genes—*TET2*, *ASXL1*, *KMT2D*, *ATRX*, and *CBL*—harbored nominally more somatic variants in AD brains than controls (Figure 2E; $p < 0.05$, one-tailed proportion test), though these individual gene enrichments were not significant after multiple hypothesis testing correction across 149 assayed genes. All of these AD-enriched genes represent critical TSGs widely implicated in various cancers[33] and CH.[34] Most AD somatic variants in *ASXL1* were nonsense variants broadly distributed across the encoded protein, including two recurrent variants observed in multiple AD patients, similar to what is seen in *ASXL1* variants in CH events of blood[34] (Figure S2D). AD patients also showed missense variants in *TET2* that clustered in its critical oxygenase domains (Figure 2F), a similar mutational pattern to that seen in CH[34] (Figure S2D) that was not observed in our aged controls. In addition to individual genes, AD patients showed significantly more somatic variants in PI3K-PKB/Akt pathway genes than controls (Figure S2E; $p < 0.05$, one-tailed proportion test), a pathway that has been previously suggested to be enriched with somatic variants in AD brains.[35]

Somatic variants in AD brains also showed significantly higher VAFs than did variants in control brains, with the effect being more pronounced in TSG, CH-related genes, or AD-enriched genes (Figures 2G and S2F). This higher VAF implies that many somatic variants found in AD drive the clonal expansion of cells that carry them to a greater extent than in control brains. Direct examination of the signal of positive selection on these variants using the dN/dS ratio test[36] confirmed that somatic variants in AD brains experienced stronger positive selection than somatic variants in control brains (Figures 2H and 2I), with greater than a 350% increase in positively selected cells (Figure 2J).

We further confirmed the enrichment of somatic variants in AD when considering only CHIP variants previously identified from blood WGS datasets[25],[37] (Figures S3A–S3C). Additional analyses suggested that this enrichment could not be explained by differences in age (Figures S3D–S3H) or sequencing depth (Figures S3I and S3J) between AD and control groups. Furthermore, logistic regression analysis confirmed a significant positive association between somatic variants—especially CHIP variants—and AD risk, after controlling for potential confounders such as sex, age, sequencing coverage, post-mortem interval, and *APOE* e4 (APOE4) genotype (Figure S3K), suggesting that somatic variants could be an independent genetic factor in addition to the well-known APOE4 allele.

Overall, our panel sequencing results revealed more frequent somatic variants in cancer driver genes, particularly CHIP variants, and stronger clonal selection in favor of these variants in AD brains, highlighting their potential role in driving the clonal expansion of proliferating cell types during AD pathogenesis.

Somatic CHIP variants are primarily carried by MLBMs

The overlap between many specific driver genes mutated in AD and those implicated in clonal blood disorders, along with the increased burden of CHIP variants in AD blood samples,[26] led us to hypothesize that microglia—or microglia-like cells—carried these variants in AD brains. To test this, we developed a fluorescence-activated nuclei sorting (FANS) method to selectively enrich microglial nuclei from frozen post-mortem brain tissues using an antibody targeting CSF1R (Figure S4A), a well-known cell-surface marker for microglia whose nuclear localization and function have been recently reported.[38] Our subsequent single-nucleus RNA sequencing (snRNAseq) confirmed that >75% of sorted nuclei in both AD and control brains showed high expression of microglial marker genes, including *CX3CR1*, *TMEM119*, and *P2RY12* (Figures 3A and S4B). Interestingly, another 4%–9% of the nuclei were classified as CNS-associated macrophages (CAMs; Figures 3A and S4B), a recently identified class of brain-resident macrophages with high expression of *MS4A7* and *MRC1*,[39] whereas the remaining cells represented various other brain cell types.

We selected 15 sSNVs and 4 sIndels identified from AD and control brains, all of which were predicted to be deleterious to critical oncogenes or TSGs, and found a marked enrichment of these variants in the sorted microglia-enriched brain macrophage fraction. We measured the VAF of each somatic variant using amplicon sequencing in sorted CSF1R+ (microglia-enriched brain macrophages) and NeuN+ (neurons) cell populations, as well as in NeuN- (glia and other nonneuronal cells) and DAPI+ (all brain cells) populations. All 10 AD somatic variants and 7 of 8 control variants in TSGs were enriched (2- to 438-fold) in the CSF1R+ cells compared with neurons from the same PFC samples (Figures 3B, S4C, and S4D). AD variants trended toward higher VAF in CSF1R+ nuclei than controls (Figure 3C),

---

(F) Distribution of somatic variants in two AD-enriched genes, *TET2* and *ASXL1*. The color and height of each lollipop denote the variant type and the number of carrying individuals.

(G) Somatic variants in AD brains show significantly higher allele fractions than controls (two-tailed *t* test), with a larger increase when only considering TSGs, CH-related genes, or AD-enriched genes, suggesting the clonal expansion of cells that carry the somatic variants. The increase in allele fraction was calculated using the ratio of medians between the AD and control groups. Boxplots show median and the first and third quartiles, with whiskers denoting 1.5× interquartile range (IQR) from hinges.

(H), Positive selection of individual genes in AD and control somatic variants. *Y*-axis denotes *p* value for testing if the gene's dN/dS ratio is higher than 1, with Benjamini-Hochberg's multiple hypothesis testing correction. *DNMT3A*, *ASXL1*, and *TET2* show significant positive selection in AD brains, stronger than in control brains.

(I) dN/dS ratios across all 149 targeted genes, in which the rates of all protein-altering variants, missense variants, nonsense variants, and splicing variants are compared with the background neutral rate estimated by synonymous variants. Asterisks denote p value < 0.05.

(J) AD brains harbor more positively selected cells than control brains, especially when we only consider somatic variants in AD-enriched genes. The number of positively selected cells was inferred based on the gene-specific dN/dS ratio, the count of somatic variants per sample, and the average VAF (see details in the STAR Methods). (A)–(D) and (I), error bar, 95% CI.

See also Figures S1–S3 and Tables S1, S2, and S3.

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

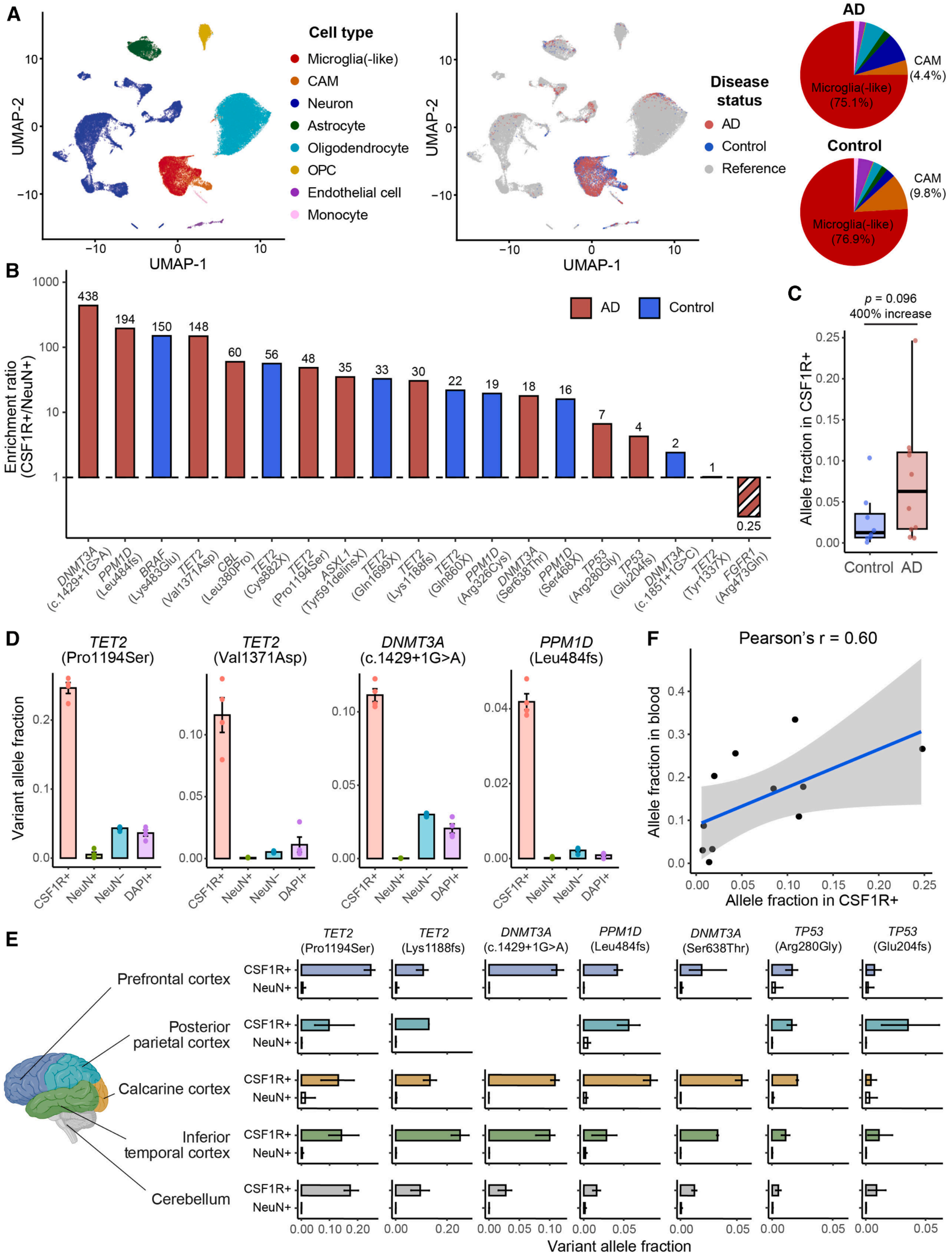


Figure 3. Somatic CHIP variants are enriched in MLBMs of AD brains

(A) 10× snRNAseq confirms the high purity and unbiased representation of microglia-enriched brain macrophages in CSF1R+ nuclei sorted from AD and control PFC samples. Clustering results suggest that about 80% of the sorted nuclei transcriptionally match microglia (red), whereas another 3%–9% are CNS-associated macrophages (CAMs, orange). Minimal blood cell contamination is confirmed with up to 1% monocytes and the absence of B cells, T cells, and red blood cells. OPC, oligodendrocyte progenitor cell.

(legend continued on next page)

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

consistent with our VAF findings in bulk brain tissues (Figure 2G). For a splicing sSNV in *DNMT3A* (c.1429+1G>A) and two deleterious missense sSNVs in *TET2* (p.Pro1194Ser and p.Val1371Asp) identified from three AD brains, we observed >10% VAFs in brain CSF1R+ cells, significantly higher than VAFs in neurons and other mixed cell populations (Figure 3D; $p < 0.05$, two-tailed Wilcoxon test) suggesting that mutant cells constitute >20% of all CSF1R+ cells in the sample. The last tested variant, in the oncogene *FGFR1* (p.Arg506Gln), is a non-recurrent variant predicted to cause decreased activation of this oncogene and was not enriched in CSF1R+ cells. Interestingly, this same AD sample harbored a variant in a TSG gene (*DNMT3A* (c.1429+1G>A)) that was almost exclusively present in CSF1R+ cells, suggesting that these two variants originated in different lineages (Figure S4C).

Regional analysis of these somatic CHIP variants showed that all seven profiled AD variants were globally present in CSF1R+ cells sorted from four distinct cerebral cortex regions and the cerebellum, but absent in neuronal populations across all brain regions (Figures 3E and S4E). Analysis of matched blood DNA showed that all 10 AD variants enriched in CSF1R+ cells were also detected in blood, with a trend toward a positive correlation between VAFs in brain CSF1R+ nuclei and blood (Figures 3F and S4F; $p = 0.052$, Pearson correlation). We confirmed minimal blood contamination in both bulk brain tissue and sorted CSF1R+ nuclei (Figures 3A and S4B). The enrichment of CHIP variants, the widespread distribution of mutant CSF1R+ cells across brain regions, and the presence of the same variants in matched blood samples, all suggest that these mutant CSF1R+ cells are more likely to have arisen from blood-infiltrating monocytes[13–15] that differentiate into microglia-like cells in the brain, rather than originating from the yolk sac-derived microglial lineage; however, definitive analyses would require specific cell lineage studies. We therefore refer to these variant-enriched cells as MLBMs.

Somatic CHIP variants lead to disease-associated microglia states *in vivo*

We next sought to understand the functional impact of these somatic cancer driver variants on MLBMs. To do so, we first utilized the recently developed genotyping of targeted loci with single-cell chromatin accessibility (GoT-ChA)[40] assay to perform simultaneous single-nucleus ATAC (assay for transposase-accessible chromatin) sequencing (snATACseq) and targeted genotyping on PFC samples from five AD patients and two controls; all five AD patients and one control individual carried somatic cancer driver variants in CH-related genes (Figure 4A; Table S4). Samples were enriched for MLBMs using CSF1R+ sorting, but deliberately sorted less stringently to preserve significant numbers of neurons and other diverse cell types. Quality control and cell-type annotation (Figures S5A–S5C) yielded 14,928 high-quality MLBMs and 8,252 high-quality neurons, of which driver variants were successfully genotyped in 8,387 and 5,861 cells, respectively (Figures 4B–4D). Consistent with our amplicon experiments, we found a strong and significant enrichment of mutant cells in MLBMs versus neurons (Figure 4E; Table S4; OR = 16.21, $p < 2.2 \times 10^{-16}$, hypergeometric test).

Analysis of the epigenetic state of mutant MLBMs versus wild-type (WT) MLBMs showed that mutant MLBMs in AD brains were enriched in cluster C8 (adjusted $p = 2.12 \times 10^{-81}$, hypergeometric test), which exhibited high activity of genes associated with the disease-associated microglia (DAM) state[41] (Figures 4F and 4G; Table S4), a state previously reported to be enriched in AD brains and to modulate the neuroinflammatory response during neurodegeneration.[7,42] In contrast, mutant MLBMs from control brains were enriched in clusters C4 and C7 (Figure 4F). Whereas C7 did not show enrichment for a particular microglial state, C4 displayed strong activity of a proliferation-related microglial state gene module (Figure S5D). This suggested that MLBMs carrying similar driver variants may adopt distinct cellular states in AD and control brains, with mutant MLBMs in AD brains more involved in neuroinflammation and neurodegeneration.

We also performed simultaneous genotyping and transcriptomic analysis of sorted brain CSF1R+ cells from one AD patient carrying the p.Pro1194Ser *TET2* variant using genotyping of transcriptomes enhanced with nanopore sequencing (GO-TEN), a method that allows simultaneous genotyping and transcriptomic analysis of single cells using long mRNA reads from Nanopore sequencing.[43] Consistent with other approaches, we confirmed that MLBMs carried the mutant *TET2* allele (Figures S5E and S5F). We did not detect mutant alleles in any neural or neuroglial cell types, though the modest number of genotyped cells in this method limits our ability to fully interrogate the variant carrier status of these cell types or transcriptomic profiles of mutant and WT cells.

Transcriptional effects of mosaic chromosome alterations in AD snRNAseq data

To further explore the effects of somatic variants in MLBMs in AD, we utilized a recent high-quality snRNAseq dataset of middle temporal gyrus neocortex samples obtained from AD donors and age-matched controls, the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD). Whereas the high degree of transcriptional

---

(B) The enrichment ratio of VAFs between sorted brain macrophage (CSF1R+) nuclei versus neuronal (NeuN+) nuclei from PFC of the same AD and control brains, estimated by amplicon sequencing. Besides the non-CHIP *FGFR1* variant serving as a negative control, 17 of the 18 profiled somatic CHIP variants demonstrate at least a 2× enrichment in brain macrophages.

(C) AD brains exhibit a trend toward higher VAF in CSF1R+ nuclei compared with controls ($p = 0.096$, two-tailed t test). Boxplots show median and the first and third quartiles, with whiskers denoting 1.5× IQR from hinges.

(D) Four somatic CHIP variants as examples show significantly higher allele fractions in CSF1R+ nuclei than the fractions in the other three populations ($p < 0.05$, two-tailed Wilcoxon test), suggesting their distinct origins. Error bar, SE.

(E) Somatic CHIP variants are distributed across multiple brain regions of the same AD brains in CSF1R+ nuclei. Error bar, min to max in replicates. (B)–(E), Each nuclear population was commonly sorted four times from each brain sample to serve as replicates.

(F) All but the *FGFR1* variants are shared between PFC CSF1R+ nuclei and whole-blood samples of the same AD individuals, indicating a common origin of these somatic variants.

See also Figure S4.

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

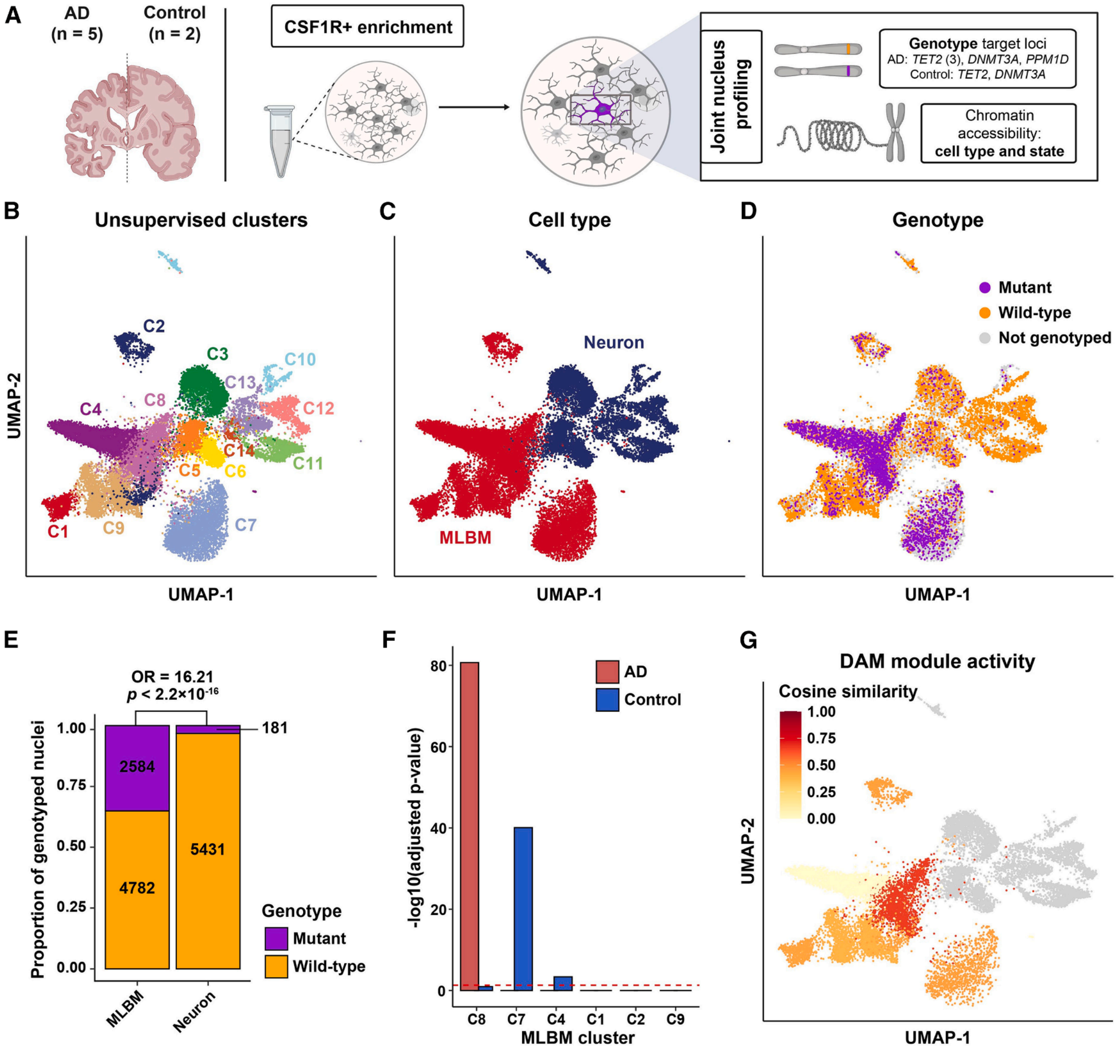


Figure 4. Simultaneous single-nucleus genotyping and ATAC sequencing display an enrichment of somatic CHIP variants in MLBMs, which exhibit high DAM activity in AD brains

(A) Schematic of experimental design for genotyping of targeted loci with single-cell chromatin accessibility (GoT-ChA) assay. Detailed information about the samples and variants we profiled is presented in Table S4.

(B–D) Uniform manifold approximation projection (UMAP) embeddings of snATACseq data, highlighting clusters derived from unsupervised clustering (B), cell type (C), and genotype (D).

(E) Mutant nuclei are significantly enriched in MLBMs compared with neurons ($p < 2.2 \times 10^{-16}$, hypergeometric test), confirming that they are the predominant cell type carrying somatic driver variants.

(F) Mutant MLBMs from AD brains are enriched in cluster C8, whereas mutant MLBMs from control brains are enriched in clusters C4 and C7 (adjusted $p < 0.05$, hypergeometric test).

(G) Activity of genes associated with the DAM microglial state is highest in cluster C8, suggesting that mutant MLBMs from AD brains are DAM-like.

See also Figure S5 and Table S4.

noise and sparsity in snRNAseq data prevent reliable sSNV calls without matched DNA-seq,[44] several methods have successfully identified mCAs from snRNAseq data.[45–47] Since recurrent mCA has also been associated with CH and other myeloid overgrowth syndromes,[48,49] generally disrupting specific genes also mutated by sSNV, we hypothesized that AD brains would also carry mCA in microglia or other brain macrophages, such as MLBM.

We extracted microglia(-like)-CAMs that were either annotated as microglia-perivascular macrophages (a subtype of CAMs) in SEA-AD or identified as microglia-CAMs through automatic cell-typing with scType (Figures S6A and S6B; Table S5). From all SEA-AD individuals with a consensus clinical diagnosis of AD (n = 31) or age-matched controls (n = 31) (Table S6), we called mCAs using CONICSmat[45] in microglia-CAMs as well as excitatory neurons (ExNs), astrocytes, oligodendrocytes, and oligodendrocyte precursor cells (OPCs), and retained only mCAs that were not called in any other cell types from the same donor and that passed several stringent filtering criteria

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

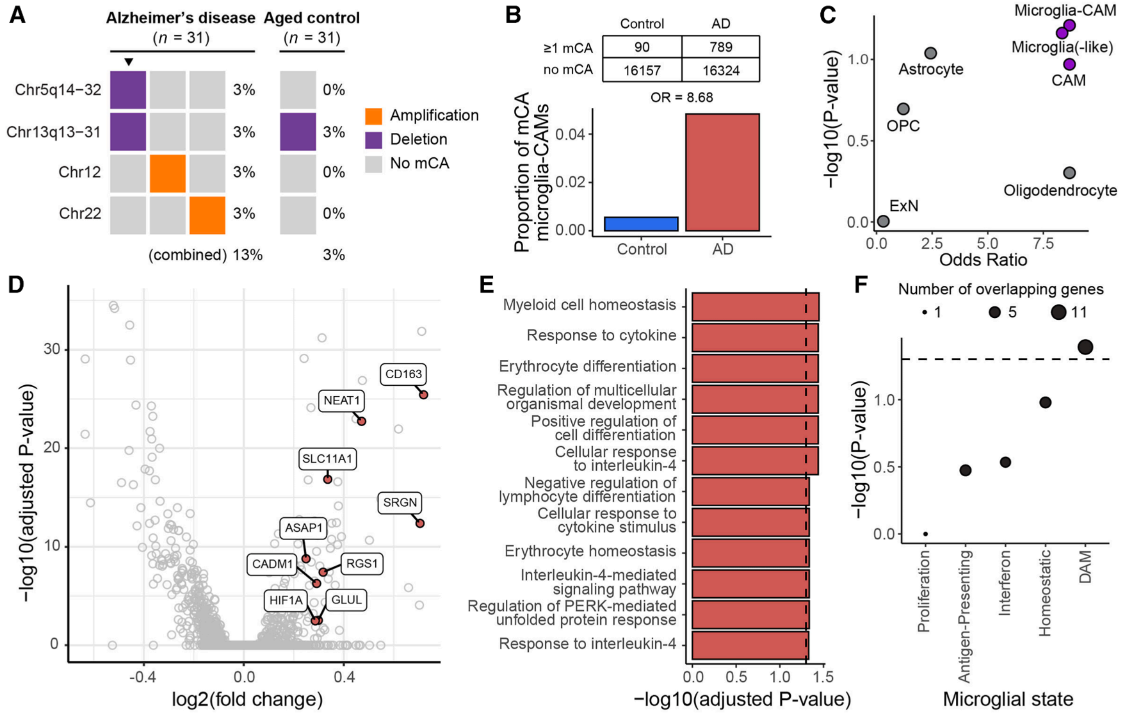


Figure 5. mCAs in AD microglia-CAMs are associated with a pro-inflammatory, disease-related signature

(A) Microglia-CAMs from AD brains contain nominally more mCAs associated with hematopoietic overgrowth syndromes compared with age-matched controls, even in this small sample ($n = 31$ each). Triangles highlight an individual with multiple mCAs.

(B) AD brains show a trend ($p = 0.06$, permutation test) toward a higher proportion of mCA-carrying microglia-CAMs than age-matched controls.

(C) Odds ratios of mCA-carrying cells between AD and control individuals across different cell types. Microglia-CAMs ($p = 0.06$) and microglia ($p = 0.07$) have the smallest nominal p values in the permutation test compared with CAMs ($p = 0.11$), astrocytes ($p = 0.09$), oligodendrocytes ($p = 0.50$), OPC ($p = 0.40$), and ExN ($p = 0.99$). OPC, oligodendrocyte progenitor cell; ExN, excitatory neuron.

(D) Volcano plot shows DEGs between AD donor microglia-CAMs with and without mCA. Positive fold-change indicates upregulation in microglia-CAMs with mCA. DAM-associated upregulated genes are colored red.

(E) Significantly enriched gene ontology terms for genes upregulated in microglia-CAMs with mCA (adjusted $p < 0.05$, hypergeometric test).

(F) Enrichment of microglial state modules[41] among genes upregulated in microglia-CAMs with mCA. Significant enrichments implicate inflammation and the DAM transcriptional state.

See also Figures S6 and S7 and Tables S5, S6, and S7.

(Figure S6C; STAR Methods). Using WGS of sorted microglia-CAMs and neurons, we validated the presence and specificity of three of the four mCAs in microglia-CAMs for which we could access primary brain tissue (Figure S6D).

AD brains harbored nominally more mCA events in microglia-CAMs (4 in AD versus 1 in control; Figure 5A) and nominally 8-fold more mCA-carrying cells (Figure 5B; $p = 0.06$, permutation test), though as expected, the SEA-AD sample size was too small for these differences to reach statistical significance. Microglia-CAMs showed the strongest AD enrichment for mCA events, compared with other profiled cell types (Figure 5C; Table S6). When we analyzed microglia and CAM clusters separately, we observed a stronger trend in microglia than in CAMs (Figure 5C; $p = 0.07$ and 0.11, permutation test).

Although the SEA-AD sample size is too small to demonstrate statistically significant enrichment of mCA in AD, the data enabled direct analysis of the transcriptional effects of mCA in microglia-CAMs. We created an integrated snRNAseq atlas of microglia-CAMs identified across AD cases and controls (Figure S7) and identified differentially expressed genes (DEGs) between mutant and WT microglia-CAMs from mCA-carrying AD brains (Figure 5D; Table S7). Using gene ontology enrichment analysis, we found that DEGs with increased expression in mutant microglia-CAMs were enriched (adjusted $p < 0.05$, hypergeometric test) for several terms related to immune activation and signaling, suggesting that mutant microglia-CAMs may upregulate pro-inflammatory pathways (Figure 5E; Table S7). We further characterized the microglial state associated with mCAs. Using a hypergeometric test for enrichment, we found marginally significant overlap between DEGs that are upregulated in mutant microglia-CAMs and genes associated with the DAM state, further suggesting that these mutant MLBMs, potentially also derived from blood immune cells, may contribute to AD pathology (Figure 5F; Table S7; $p = 0.04$).

CHIP variants induce inflammatory changes in iPSC-derived microglia-like cells

To further test the effects of these somatic variants in microglia, we generated isogenic iPSC lines harboring driver variants in the three genes most commonly mutated in AD cases (*DNMT3A*

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

R882H/WT, *TET2* DelE3-E11/WT, *ASXL1* E635fs*15/WT; Figure 6A) using CRISPR/Cas9-mediated gene editing of a parental iPSC line derived from an individual with the *APOE* e3/e3 genotype, considered neutral with regard to AD risk.[50] We differentiated the three mutant and WT isogenic iPSC lines to induced microglia-like cells (iMGLs) using a modified protocol of McQuade et al.[51] and performed scRNAseq (Figures 6B and S8A–S8C). We created an integrated scRNAseq atlas of these four genotypes and identified microglial states based on their transcriptional signatures (Figures 6C, S8D, and S8E). Mutant iMGLs were enriched within the DAM and proliferation microglial states, relative to WT iMGLs, depending on the variant type (Figure 6D; Table S8; $p < 0.05$, hypergeometric test), suggesting a cell-autonomous, causal relationship between these somatic variants and a more activated microglial state. We next performed differential expression analysis between WT and mutant iMGLs and found that DEGs with increased expression in mutant iMGLs included many canonical DAM genes. Using gene set enrichment analysis with the 50 Molecular Signature Databases Hallmark gene sets, we observed upregulation of pro-inflammatory and proliferation-associated pathways in mutant iMGLs (Figures 6E, 6F, and S8F; Table S8). We also observed several enriched metabolic terms, such as glycolysis. A shift from aerobic metabolism to glycolysis has been identified as critical to microglial activation in AD.[52] Taken together, our *in vivo* and *in vitro* results consistently suggest that recurrent AD-associated somatic CH variants are sufficient to lead to pro-inflammatory, microglia-activating transcriptional changes and a shift toward the AD-associated DAM microglial state.

DISCUSSION

Our results revealed a greater burden of somatic variants in AD cerebral cortex samples when compared with matched controls, suggesting an association between brain somatic variants and AD. These somatic variants were enriched in driver genes that have been widely implicated in cancer and pre-malignant conditions, notably in CH, with higher VAFs and stronger positive selection signals in AD brains, implying their roles in clonal expansion of mutant cells. We further found that most of the driver variants we profiled were specifically present in MLBMs, corresponding to parenchymal microglia and/or blood-derived monocyte lineage cells differentiating into microglia-like cells. Finally, using *in vivo* and *in vitro* single-cell analyses, we found that MLBMs carrying CH variants, as well as iMGLs carrying AD-associated variants, all showed pro-inflammatory and disease-associated transcriptional signatures compared with their WT counterparts. These transcriptional changes seen in somatically mutated MLBMs resemble effects of CH variants in blood myeloid cells, which are known to increase the risk of myocardial infarction and stroke while activating immune cascades, including IL1β and IL6.[53] These parallels suggest that somatic variant-driven changes in MLBMs in late-stage AD may be damaging to neurons.[54]

Two recent studies investigating the association between CHIP variants in blood with AD risk found no effect,[24] or a surprising protective effect of blood CHIP on AD.[25] Both studies primarily relied on blood WES at conventional depth (~50–70×; Bouzid et al. also analyzed ~3,000 blood WGS at ~30×), which could only reliably detect variants with VAFs >5%–10%,[24,37] although CH variants with lower VAFs are more typical in the blood.[31] In contrast, a more recent study[26] using both UMI-based deep panel sequencing (~800×) and a much larger cohort of ~30,000 WGS on blood samples both consistently identified an increased burden of CHIP variants in AD patients compared with age-matched controls, with this increase primarily attributed to variants with VAF <5%, which might explain why the two earlier studies failed to detect this association. This VAF-dependent pattern is further supported by a recent study on CHIP and adjudicated probable dementia, which reported a negative association when including only CHIP variants with VAF >8%, but a non-significant positive association when the VAF threshold was lowered to >2%.[55] Indeed, we confirmed that many variants with low VAF in the brain were undetectable in WES of matched blood samples[25] (Figure S9A), and in our data, the enrichment of somatic variants in AD was largely driven by such lower-VAF variants (Figures S9B and S9C).

In our study, the logistic regression model suggested that the burden of somatic driver variants is positively associated with AD, independent of the well-known germline genetic factor, the APOE4 allele (Figure S3K). This observation in the brain was further supported by several recent blood-based studies examining CHIP[26] or mCA[56] in AD cohorts, both of which reported a significant increase in CH variants in AD blood, particularly among non-APOE4 individuals. These findings highlight that somatic variants, especially CHIP variants, may contribute to AD risk by an *APOE*-independent mechanism. As APOE4 has been reported to increase neuroinflammation by disrupting lipid homeostasis,[57] our study and others underscore that somatic driver variants in MLBMs could represent an alternative pathway for inducing neuroinflammation in AD pathogenesis.

Our finding of shared somatic driver variants between brain and blood aligns with that of Bouzid et al.[25] as well as an earlier small study that also found cancer driver variants in AD brain.[27] The current consensus in the literature suggests that no progenitor capable of giving rise to both yolk sac-derived microglia and hematopoietic stem cells is known or believed to exist during embryonic development,[58] indicating that the shared somatic variants between brain and blood and the increase in AD are more likely to arise from another mechanism. Together with the blood studies,[26,56] our results—particularly the widespread distribution of mutant MLBMs across different regions—suggest a more likely scenario of brain infiltration by bone marrow-derived monocytes carrying these variants. Indeed, recent studies have shown that these blood monocytes can enter the brain when the BBB is compromised, an early feature of AD and brain aging,[59] and they can differentiate into microglia-like cells[60] and contribute to neurodegeneration.[14] Other reports suggest that monocytes may infiltrate the brain and adopt microglia-like characteristics even in the absence of BBB breakdown.[13,61] A recent study also reported infiltration of myeloid cells with CH variants in lung cancer and other tumors, which was associated with disease progression and higher mortality, raising the possibility that mutant monocytes/macrophages may have increased ability to infiltrate sites of inflammation.[62] Although our results favor the monocyte-derived origin of these mutant MLBMs, it cannot be excluded that these variants originate from yet unidentified

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

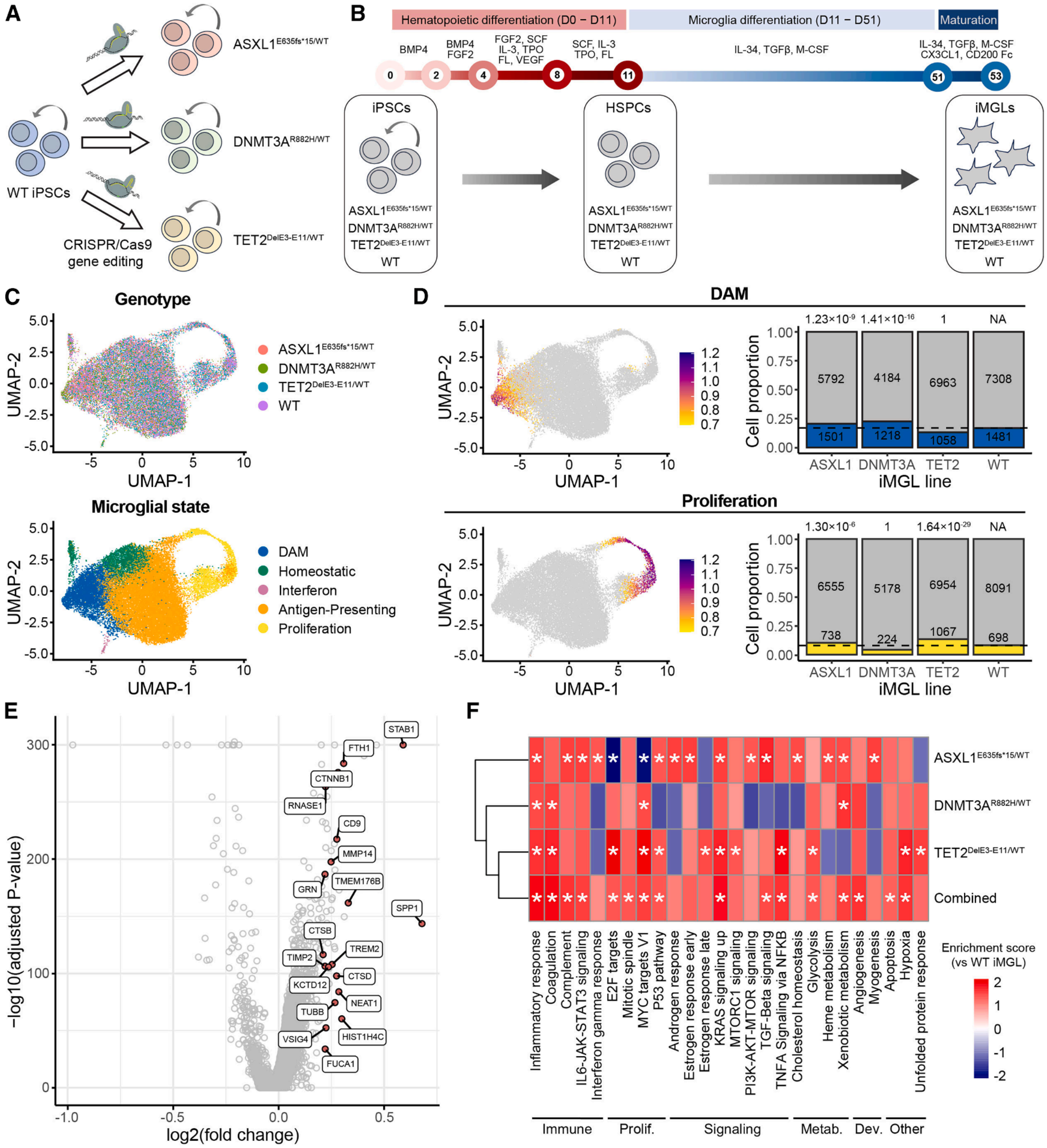


Figure 6. Transcriptomic changes in iMGLs carrying CHIP variants

(A) Schematic of the generation of isogenic induced pluripotent stem cell (iPSC) lines with driver variants in the three major CH-related genes that are recurrently mutated in AD brain, *DNMT3A*, *TET2*, and *ASXL1*. WT, wild-type.

(B) Schematic of two-step differentiation of iPSCs, first into hematopoietic stem/progenitor cells (HSPCs), through a hemogenic endothelium intermediate, and subsequently into iMGLs.

(C) Upper: UMAP representation of integrated scRNAseq data across all iPSC lines, colored by genotype. Lower: UMAP representation of the same data with annotation of distinct transcriptional microglial states.

(D) Expression of gene modules within the integrated scRNAseq atlas and relative proportion of iMGLs for DAM and proliferation states. iMGLs in the DAM state are significantly enriched in the *ASXL1*-mutant and *DNMT3A*-mutant lines, and iMGLs in the proliferative state are enriched in the *ASXL1*-mutant and *TET2*-mutant lines (adjusted $p < 0.05$, hypergeometric test).

(E) Volcano plot of DEGs between mutant and WT iMGLs. Genes that are upregulated in the mutant iMGLs include many canonical DAM genes (highlighted in red).

(F) Gene set enrichment analysis with Molecular Signatures Database Hallmark pathways for DEGs from each mutant genotype individually and all three mutant genotypes combined, compared with the WT iMGLs, highlighting several processes related to inflammation and proliferation. Significant pathways (adjusted $p < 0.05$) are indicated with asterisks. Prolif, proliferation; Metab, metabolic; Dev, development.

See also Figure S8 and Table S8.

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

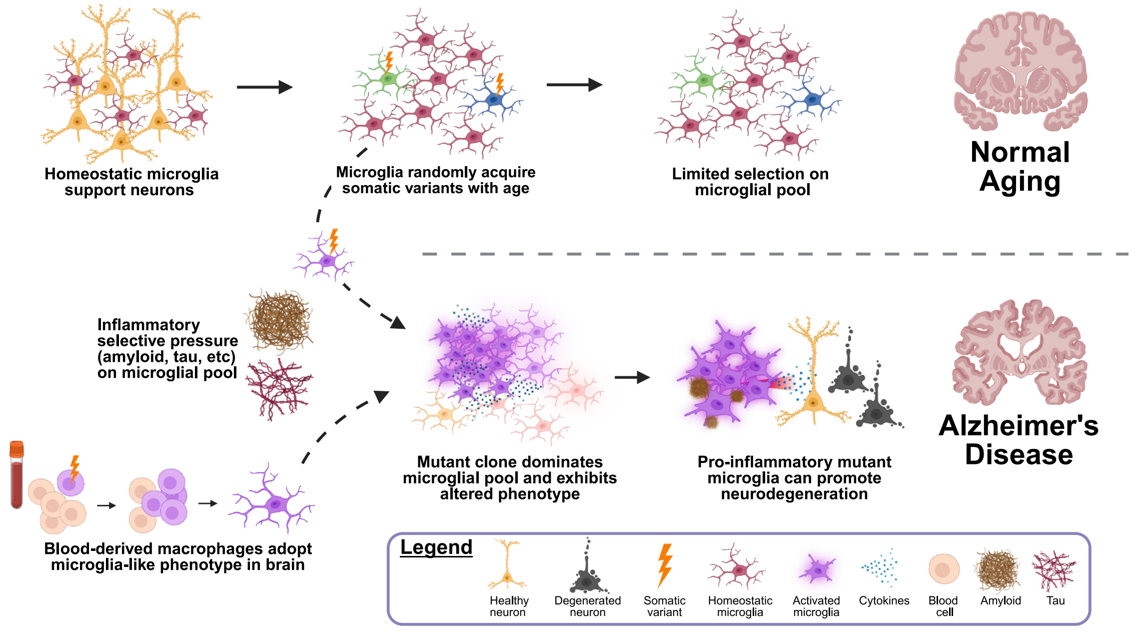


Figure 7. MLBMs with somatic driver variants exhibit clonal expansion with an altered phenotype capable of contributing to neurodegeneration

Under homeostatic conditions, microglia play a key role in supporting neuronal health. As with all cells, microglia accumulate somatic variants randomly with age. In normal aging, most of these variants are functionally neutral, and microglial clones are under limited natural selection. In contrast, certain somatic cancer driver variants can confer a fitness advantage, leading to the selective expansion of mutant microglial clones (denoted as purple) over time. In AD, pathological changes—such as the accumulation of amyloid-β and tau—may impose selective pressures that accelerate the expansion of these mutant immune cell populations. Given that infiltrating blood-derived monocytes can adopt a microglia-like phenotype within the brain, the presence of shared somatic driver variants between microglia and blood supports a potential hematopoietic origin for these clones. Clonal expansion could occur in the blood prior to infiltration, or the variant itself could promote enhanced infiltration compared with WT blood cells. In either scenario, the mutant clone dominates the native microglial pool. We further show that some of these driver variants promote pro-inflammatory transcriptional programs, which may contribute to the progression of neurodegeneration.

shared microglia-macrophage progenitors in early embryogenesis; thus, future lineage-tracing studies of microglia and blood cells will be needed to better resolve this question.

Although transcriptomic differences between monocyte-derived macrophages and brain-resident microglia have been reported in mouse models,[63] our snRNAseq data show that cells carrying CHIP variants express all classical microglial markers (Figure S4B) and are transcriptionally indistinguishable from brain-resident microglia. This is consistent with a recent study showing highly similar transcriptomes between yolk sac-derived and monocyte-derived microglia in aging mice and humans.[64] Furthermore, our results suggest that when MLBMs carry CH variants, they shift into pro-inflammatory and proliferative states, potentially contributing to neuronal loss during AD pathogenesis—paralleling another study showing that brain-resident microglia carrying driver variants may also lead to a neuroinflammatory phenotype in certain AD patients.[28]

Our data suggest a model (Figure 7) in which non-somatic drivers of AD—such as amyloid, tau, and APOE4—contribute to an inflammatory brain environment that exerts lifelong effects on clonal competition within the microglial population. Age-related disruption of the BBB may further promote clonal evolution by introducing MLBMs into the microglia pool. Chronic inflammation induces the proliferation of microglia, in which positive selection favors cells carrying cancer driver variants that provide fitness advantages over those without. Operation of this clonal selection over decades would be expected to lead to the high fraction of mutant cells that we observe, and as mutant clones accumulate additional variants over time, microglia carrying them would be expected to become increasingly abnormal in their transcriptomic patterns and their ability to sustain versus damage neighboring neurons. Therefore, somatic cancer driver variants could be a key driver of the switch in microglia from the homeostatic to the pro-inflammatory, disease-associated state that typifies late-stage disease.

Potential roles of somatic cancer driver variants in AD pathogenesis open up a whole new range of therapeutic avenues in AD, complementary to approaches emphasizing amyloid and tau. For example, our analysis highlighted five highly mutated genes as well as the PI3K-PKB/Akt pathway (including a *PIK3CA* p.His1047Leu activating variant and three loss-of-function variants in *TSC1/2*), which were recurrently disrupted by somatic variants in AD brains. Drugs targeting such genes are already widely used to treat cancer,[65,66] and thus they might serve as potential therapeutic agents to suppress somatic-variant-activated MLBMs and ultimately neurodegeneration in AD. The more commonly mutated genes in the CH-related pathways are now being actively targeted to treat their known roles in clonal blood diseases,[67] which may offer other approaches to AD. Since the role of DAM in neuronal loss and dysfunction may be a common feature shared across many neurodegenerative diseases, studying somatic variants in AD and brain aging may provide an important new approach to understanding the pathogenic mechanisms of dementia and other neurodegenerative conditions.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Christopher A. Walsh (christopher.walsh@childrens.harvard.edu).

Materials availability

ROSMAP resources can be requested at https://www.radc.rush.edu. Other materials are available from the authors upon reasonable request.

Data and code availability

The panel sequencing and single-cell multi-omic data generated in this study have been deposited in the AMP-AD Knowledge Portal (https://doi.org/10.7303/syn73904737) and are subject to controlled-use conditions in accordance with human privacy regulations. Custom bash and R scripts used in this study are publicly available at https://github.com/AugustHuang/AD-microglia-somatic-mutation.git.

ACKNOWLEDGMENTS

We thank S. Jaiswal for collaboration and discussions on the project; B. Stevens, B. Hyman, P. Loh, L. Harvey, and B. Yankner for constructive discussions and suggestions on the manuscript; and R.S. Hill, J.E. Neil, D. Gonzalez, M. Chin, and T. Dolbeare for their help with the experiments. R. Mathieu and T. Berisha from the Boston Children's Hospital Flow Cytometry Core and IDDRC Molecular Genetics Core helped with sorting. We thank the donors of postmortem tissues for their invaluable contributions to the advancement of science. This work was supported by R01 AG088082 (A.Y.H.); R56 AG079857 (A.Y.H., C.A.W., and E.A.L.); Alzheimer's Association Research Fellowship (A.Y.H.); PRMRP Discovery award W81XWH2010028 (Z.Z.); Edward R. and Anne G. Lefler Center Postdoctoral Fellowship (Z.Z.); T32 GM007753 (M.T. and L.E.); T32 GM144273 (M.T. and L.E.); T32 GM141745 (L.E.); Harvard Stem Cell Institute Fellowship (L.E.); K08 AG065502 (M.B.M.); T32 HL007627 (M.B.M.); Brigham and Women's Hospital Program for Interdisciplinary Neuroscience through a gift from L. and T. Rand (M.B.M.); Alzheimer's Disease Research Program of the BrightFocus Foundation A20201292F (M.B.M.); Doris Duke Charitable Foundation Clinical Scientist Development Award 2021183 (M.B.M.); the Manton Center Pilot Project Award and Rare Disease Research Fellowship (B.Z.); R25 NS065743 K08 NS128272 (S.K.); Physician Scientist Fellowship from the Doris Duke Charitable Foundation (S.K.); Career Award for Medical Scientists from the Burroughs Wellcome Fund (S.K.); U19 AG060909 (E.S.L.); R01 CA271418, CA260711, and CA271331 (E.P.P.); Edward P Evans Foundation Discovery Research Grant (E.P.P.); Suh Kyungbae Foundation (E.A.L.); DP2 AG072437 (E.A.L.); R01 NS032457-20S1 (C.A.W.); R01 AG070921 and AG078929 (C.A.W. and E.A.L.); F-Prime Foundation (C.A.W.); Allen Discovery Center program, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation (C.A.W. and E.A.L.); and UG3 NS132138 and UG3 NS132144 through the SMAHT Consortium. ROSMAP is supported by P30 AG10161, P30 AG72975, R01 AG15819, R01 AG17917, U01 AG46152, and U01 AG61356 (D.A.B.). C.A.W. is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

A.Y.H., Z.Z., M.B.M., E.A.L., and C.A.W. conceived and designed the study. Z.Z. designed and performed panel and amplicon sequencing for bulk brain and sorted cell populations, with the assistance of M.B.M., B.C., L.E., A.Y., I.R., E.S., S.K., and J.G. A.Y.H. performed somatic variants calling on panel sequencing and amplicon sequencing data, with the assistance of M. Zheng, M. Zhou, J.C., B.Z., J.K., Z.A., and Y.C. M.D., M.N., S.G.M., and E.P.P. generated iPSCs and iMGLs and performed scRNAseq on them. L.E. designed and performed GoT-ChA experiments, with the assistance of L.M. and D.L. M.T. performed somatic variants calling from SEA-AD snRNAseq data and GoT-ChA data, as well as downstream analyses for SEA-AD snRNAseq, iMGL scRNAseq, and GoT-ChA snATACseq data, under the guidance of A.Y.H. P.L.D.J. and D.A.B. provided brain tissues and genomic DNA for ROSMAP samples. K.J.T., M.I.G., R.D.H., E.S.K., and E.S.L. provided SEA-AD snRNAseq data. J.A.B. provided ADSP data. A.Y.H., E.A.L., and C.A.W. supervised the study. A.Y.H., Z.Z., M.T., L.E., S.G.M., E.P.P., E.A.L., and C.A.W. wrote the manuscript.

DECLARATION OF INTERESTS

C.A.W. is a founding scientific advisor of Mosaica Medicines (equity), on the clinical advisory board of Maze Therapeutics (equity), and on the scientific advisory board of Bioskryb Genomics (cash and equity). No research support is received. These companies did not fund and had no role in the conception or performance of this research project. Boston Children's Hospital has filed a provisional patent application based on this work, on which A.Y.H., Z.Z., E.A.L., and C.A.W. are co-inventors.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2026.03.040.

Received: April 18, 2024
Revised: May 21, 2025
Accepted: March 24, 2026

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

REFERENCES

1. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A., et al. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat. Genet. *51*, 414–430. https://doi.org/10.1038/s41588-019-0358-2.
2. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet. *51*, 404–413. https://doi.org/10.1038/s41588-018-0311-9.
3. Hardy, J., and Escott-Price, V. (2019). Genes, pathways and risk prediction in Alzheimer's disease. Hum. Mol. Genet. *28*, R235–R240. https://doi.org/10.1093/hmg/ddz163.
4. Bellenguez, C., Kucukali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat. Genet. *54*, 412–436. https://doi.org/10.1038/s41588-022-01024-z.
5. Saijo, K., and Glass, C.K. (2011). Microglial cell origin and phenotypes in health and disease. Nat. Rev. Immunol. *11*, 775–787. https://doi.org/10.1038/nri3086.
6. Ochocka, N., and Kaminska, B. (2021). Microglia Diversity in Healthy and Diseased Brain: Insights from Single-Cell Omics. Int. J. Mol. Sci. *22*, 3027. https://doi.org/10.3390/ijms22063027.
7. Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T.K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., et al. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell *169*, 1276–1290.e17. https://doi.org/10.1016/j.cell.2017.05.018.
8. Olah, M., Menon, V., Habib, N., Taga, M.F., Ma, Y., Yung, C.J., Cimpean, M., Khairallah, A., Coronas-Samano, G., Sankowski, R., et al. (2020). Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. Nat. Commun. *11*, 6129. https://doi.org/10.1038/s41467-020-19737-2.
9. Young, A.M.H., Kumasaka, N., Calvert, F., Hammond, T.R., Knights, A., Panousis, N., Park, J.S., Schwartzentruber, J., Liu, J., Kundu, K., et al. (2021). A map of transcriptional heterogeneity and regulatory variation in human microglia. Nat. Genet. *53*, 861–868. https://doi.org/10.1038/s41588-021-00875-2.
10. Colonna, M., and Butovsky, O. (2017). Microglia Function in the Central Nervous System During Health and Neurodegeneration. Annu. Rev. Immunol. *35*, 441–468. https://doi.org/10.1146/annurev-immunol-051116-052358.
11. Hickman, S., Izzy, S., Sen, P., Morsett, L., and El Khoury, J. (2018). Microglia in neurodegeneration. Nat. Neurosci. *21*, 1359–1369. https://doi.org/10.1038/s41593-018-0242-x.
12. Bohlen, C.J., Friedman, B.A., Dejanovic, B., and Sheng, M. (2019). Microglia in Brain Development, Homeostasis, and Neurodegeneration. Annu. Rev. Genet. *53*, 263–288. https://doi.org/10.1146/annurev-genet-112618-043515.
13. Lund, H., Pieber, M., Parsa, R., Han, J., Grommisch, D., Ewing, E., Kular, L., Needhamsen, M., Espinosa, A., Nilsson, E., et al. (2018). Competitive repopulation of an empty microglial niche yields functionally distinct subsets of microglia-like cells. Nat. Commun. *9*, 4845. https://doi.org/10.1038/s41467-018-07295-7.
14. Wilk, C.M., Cathomas, F., Török, O., Le Berichel, J., Park, M.D., Bigenwald, C., Heaton, G.R., Hamon, P., Troncoso, L., Scull, B.P., et al. (2023). Circulating senescent myeloid cells infiltrate the brain and cause neurodegeneration in histiocytic disorders. Immunity *56*, 2790–2802.e6. https://doi.org/10.1016/j.immuni.2023.11.011.
15. Silvin, A., Uderhardt, S., Piot, C., Da Mesquita, S., Yang, K., Geirsdottir, L., Mulder, K., Eyal, D., Liu, Z., Bridlance, C., et al. (2022). Dual ontogeny of disease-associated microglia and disease inflammatory macrophages in aging and neurodegeneration. Immunity *55*, 1448–1465.e6. https://doi.org/10.1016/j.immuni.2022.07.004.
16. Lodato, M.A., Rodin, R.E., Bohrson, C.L., Coulter, M.E., Barton, A.R., Kwon, M., Sherman, M.A., Vitzthum, C.M., Luquette, L.J., Yandava, C.N., et al. (2018). Aging and neurodegeneration are associated with increased mutations in single human neurons. Science *359*, 555–559. https://doi.org/10.1126/science.aao4426.
17. Li, R., Du, Y., Chen, Z., Xu, D., Lin, T., Jin, S., Wang, G., Liu, Z., Lu, M., Chen, X., et al. (2020). Macroscopic somatic clonal expansion in morphologically normal human urothelium. Science *370*, 82–89. https://doi.org/10.1126/science.aba7300.
18. Li, R., Di, L., Li, J., Fan, W., Liu, Y., Guo, W., Liu, W., Liu, L., Li, Q., Chen, L., et al. (2021). A body map of somatic mutagenesis in morphologically normal human tissues. Nature *597*, 398–403. https://doi.org/10.1038/s41586-021-03836-1.
19. Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. Science *349*, 1483–1489. https://doi.org/10.1126/science.aab4082.
20. Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. *371*, 2477–2487. https://doi.org/10.1056/NEJMoa1409405.
21. Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., et al. (2017). Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. N. Engl. J. Med. *377*, 111–121. https://doi.org/10.1056/NEJMoa1701719.
22. Avagyan, S., Henninger, J.E., Mannherz, W.P., Mistry, M., Yoon, J., Yang, S., Weber, M.C., Moore, J.L., and Zon, L.I. (2021). Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis. Science *374*, 768–772. https://doi.org/10.1126/science.aba9304.
23. Ahmad, H., Jahn, N., and Jaiswal, S. (2023). Clonal Hematopoiesis and Its Impact on Human Health. Annu. Rev. Med. *74*, 249–260. https://doi.org/10.1146/annurev-med-042921-112347.
24. Kessler, M.D., Damask, A., O'Keeffe, S., Banerjee, N., Li, D., Watanabe, K., Marketta, A., Van Meter, M., Semrau, S., Horowitz, J., et al. (2022). Common and rare variant associations with clonal haematopoiesis phenotypes. Nature *612*, 301–309. https://doi.org/10.1038/s41586-022-05448-9.
25. Bouzid, H., Belk, J.A., Jan, M., Qi, Y., Sarnowski, C., Wirth, S., Ma, L., Chrostek, M.R., Ahmad, H., Nachun, D., et al. (2023). Clonal hematopoiesis is associated with protection from Alzheimer's disease. Nat. Med. *29*, 1662–1670. https://doi.org/10.1038/s41591-023-02397-2.
26. Choi, J., Park, K.S., Le Guen, Y.L., Park, J.-H., Zhou, Z., Enyenihi, L., Rosen, I., Choi, Y.-H., Walsh, C.A., Kim, J.-W., et al. (2025). Clonal Hematopoiesis Mutations Increase Risk of Alzheimer's Disease with APOE ε3/ε3 Genotype. Preprint at bioRxiv. https://doi.org/10.1101/2025.05.19.654981.
27. Keogh, M.J., Wei, W., Aryaman, J., Walker, L., van den Ameele, J., Coxhead, J., Wilson, I., Bashton, M., Beck, J., West, J., et al. (2018). High prevalence of focal and multi-focal somatic genetic variants in the human brain. Nat. Commun. *9*, 4257. https://doi.org/10.1038/s41467-018-06331-w.
28. Vicario, R., Fragkogianni, S., Weber, L., Lazarov, T., Hu, Y., Hayashi, S.Y., Craddock, B.P., Socci, N.D., Alberdi, A., Baako, A., et al. (2025). A microglia clonal inflammatory disorder in Alzheimer's Disease. eLife *13*, RP96519. https://doi.org/10.7554/eLife.96519.3.
29. Mass, E., Jacome-Galarza, C.E., Blank, T., Lazarov, T., Durham, B.H., Ozkaya, N., Pastore, A., Schwabenland, M., Chung, Y.R., Rosenblum, M.K., et al. (2017). A somatic mutation in erythro-myeloid progenitors

causes neurodegenerative disease. Nature *549*, 389–393. https://doi.org/10.1038/nature23672.

30. Ganz, J., Maury, E.A., Becerra, B., Bizzotto, S., Doan, R.N., Kenny, C.J., Shin, T., Kim, J., Zhou, Z., Ligon, K.L., et al. (2022). Rates and Patterns of Clonal Oncogenic Mutations in the Normal Human Brain. Cancer Discov. *12*, 172–185. https://doi.org/10.1158/2159-8290.CD-21-0245.
31. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. Nature *606*, 343–350. https://doi.org/10.1038/s41586-022-04786-y.
32. Bae, T., Fasching, L., Wang, Y., Shin, J.H., Suvakov, M., Jang, Y., Norton, S., Dias, C., Mariani, J., Jourdon, A., et al. (2022). Analysis of somatic mutations in 131 human brains reveals aging-associated hypermutability. Science *377*, 511–517. https://doi.org/10.1126/science.abm6222.
33. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. *47*, D941–D947. https://doi.org/10.1093/nar/gky1015.
34. Zink, F., Stacey, S.N., Norddahl, G.L., Frigge, M.L., Magnusson, O.T., Jonsdottir, I., Thorgeirsson, T.E., Sigurdsson, A., Gudjonsson, S.A., Gudmundsson, J., et al. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood *130*, 742–752. https://doi.org/10.1182/blood-2017-02-769869.
35. Park, J.S., Lee, J., Jung, E.S., Kim, M.H., Kim, I.B., Son, H., Kim, S., Kim, S., Park, Y.M., Mook-Jung, I., et al. (2019). Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. Nat. Commun. *10*, 3090. https://doi.org/10.1038/s41467-019-11000-7.
36. Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R., and Campbell, P.J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. Cell *171*, 1029–1041.e21. https://doi.org/10.1016/j.cell.2017.09.042.
37. Bick, A.G., Weinstock, J.S., Nandakumar, S.K., Fulco, C.P., Bao, E.L., Zekavat, S.M., Szeto, M.D., Liao, X., Leventhal, M.J., Nasser, J., et al. (2020). Inherited causes of clonal haematopoiesis in 97,691 whole genomes. Nature *586*, 763–768. https://doi.org/10.1038/s41586-020-2819-2.
38. Bencheikh, L., Diop, M.K., Rivière, J., Imanci, A., Pierron, G., Souquere, S., Naimo, A., Morabito, M., Dussiot, M., De Leeuw, F., et al. (2019). Dynamic gene regulation by nuclear colony-stimulating factor 1 receptor in human monocytes and macrophages. Nat. Commun. *10*, 1935. https://doi.org/10.1038/s41467-019-09970-9.
39. Masuda, T., Sankowski, R., Staszewski, O., and Prinz, M. (2020). Microglia Heterogeneity in the Single-Cell Era. Cell Rep. *30*, 1271–1281. https://doi.org/10.1016/j.celrep.2020.01.010.
40. Izzo, F., Myers, R.M., Ganesan, S., Mekerishvili, L., Kottapalli, S., Prieto, T., Eton, E.O., Botella, T., Dunbar, A.J., Bowman, R.L., et al. (2024). Mapping genotypes to chromatin accessibility profiles in single cells. Nature *629*, 1149–1157. https://doi.org/10.1038/s41586-024-07388-y.
41. Dolan, M.J., Therrien, M., Jereb, S., Kamath, T., Gazestani, V., Atkeson, T., Marsh, S.E., Goeva, A., Lojek, N.M., Murphy, S., et al. (2023). Exposure of iPSC-derived human microglia to brain substrates enables the generation and manipulation of diverse transcriptional states in vitro. Nat. Immunol. *24*, 1382–1390. https://doi.org/10.1038/s41590-023-01558-2.
42. Leng, F., and Edison, P. (2021). Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? Nat. Rev. Neurol. *17*, 157–172. https://doi.org/10.1038/s41582-020-00435-y.
43. Bizzotto, S., Talukdar, M., Stronge, E.A., Ramirez, R.B., Yang, Y., Huang, A.Y., Hu, Q., Hou, Y., Hylton, N.K., Finander, B., et al. (2025). Cell-type-informed genotyping of mosaic focal epilepsies reveals cell-autonomous and non-cell-autonomous disease-associated transcriptional programs. Proc. Natl. Acad. Sci. USA *122*, e2509622122. https://doi.org/10.1073/pnas.2509622122.
44. Huang, A.Y., and Lee, E.A. (2021). Identification of Somatic Mutations From Bulk and Single-Cell Sequencing Data. Front. Aging *2*, 800380. https://doi.org/10.3389/fragi.2021.800380.
45. Müller, S., Cho, A., Liu, S.J., Lim, D.A., and Diaz, A. (2018). CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. Bioinformatics *34*, 3217–3219. https://doi.org/10.1093/bioinformatics/bty316.
46. Gao, R., Bai, S., Henderson, Y.C., Lin, Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis, A., et al. (2021). Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat. Biotechnol. *39*, 599–608. https://doi.org/10.1038/s41587-020-00795-2.
47. Gao, T., Soldatov, R., Sarkar, H., Kurkiewicz, A., Biederstedt, E., Loh, P.R., and Kharchenko, P.V. (2022). Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. Nat. Biotechnol. *41*, 417–426. https://doi.org/10.1038/s41587-022-01468-y.
48. Saiki, R., Momozawa, Y., Nannya, Y., Nakagawa, M.M., Ochi, Y., Yoshizato, T., Terao, C., Kuroda, Y., Shiraishi, Y., Chiba, K., et al. (2021). Combined landscape of single-nucleotide variants and copy number alterations in clonal hematopoiesis. Nat. Med. *27*, 1239–1249. https://doi.org/10.1038/s41591-021-01411-9.
49. Loh, P.R., Genovese, G., and McCarroll, S.A. (2020). Monogenic and polygenic inheritance become instruments for clonal selection. Nature *584*, 136–141. https://doi.org/10.1038/s41586-020-2430-6.
50. Bu, G. (2009). Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. Nat. Rev. Neurosci. *10*, 333–344. https://doi.org/10.1038/nrn2620.
51. McQuade, A., Coburn, M., Tu, C.H., Hasselmann, J., Davtyan, H., and Blurton-Jones, M. (2018). Development and validation of a simplified method to generate human microglia from pluripotent stem cells. Mol. Neurodegener. *13*, 67. https://doi.org/10.1186/s13024-018-0297-x.
52. Pan, R.Y., He, L., Zhang, J., Liu, X., Liao, Y., Gao, J., Liao, Y., Yan, Y., Li, Q., Zhou, X., et al. (2022). Positive feedback regulation of microglial glucose metabolism by histone H4 lysine 12 lactylation in Alzheimer's disease. Cell Metab. *34*, 634–648.e6. https://doi.org/10.1016/j.cmet.2022.02.013.
53. Abplanalp, W.T., Cremer, S., John, D., Hoffmann, J., Schuhmacher, B., Merten, M., Rieger, M.A., Vasa-Nicotera, M., Zeiher, A.M., and Dimmeler, S. (2021). Clonal Hematopoiesis-Driver DNMT3A Mutations Alter Immune Cells in Heart Failure. Circ. Res. *128*, 216–228. https://doi.org/10.1161/CIRCRESAHA.120.317104.
54. Block, M.L., Zecca, L., and Hong, J.S. (2007). Microglia-mediated neurotoxicity: uncovering the molecular mechanisms. Nat. Rev. Neurosci. *8*, 57–69. https://doi.org/10.1038/nrn2038.
55. Jakubek, Y.A., Smith, A.P., Leng, X.I., Hall, M.E., Ezzat, D., Pershad, Y., Collins, J.M., Uddin, M.M., Fardo, D.W., Natarajan, P., et al. (2025). Clonal hematopoiesis of indeterminate potential and the risk of cognitive impairment in the Women's Health Initiative Memory Study. Alzheimers Dement. *21*, e70737. https://doi.org/10.1002/alz.70737.
56. Naito, T., Hirata, K., Jang, B., Lakhani, C.M., Buonfiglioli, A., Lee, W.P., Valladares, O., Wang, L.S., Okada, Y., Won, H.H., et al. (2025). Mosaic chromosomal alterations in blood are associated with an increased risk of Alzheimer's disease. Preprint at medRxiv. https://doi.org/10.1101/2025.05.29.25328544.
57. Raulin, A.C., Doss, S.V., Trottier, Z.A., Ikezu, T.C., Bu, G., and Liu, C.C. (2022). ApoE in Alzheimer's disease: pathophysiology and therapeutic strategies. Mol. Neurodegener. *17*, 72. https://doi.org/10.1186/s13024-022-00574-4.
58. Hoeffel, G., and Ginhoux, F. (2018). Fetal monocytes and the origins of tissue-resident macrophages. Cell. Immunol. *330*, 5–15. https://doi.org/10.1016/j.cellimm.2018.01.001.

59. Montagne, A., Nation, D.A., Sagare, A.P., Barisano, G., Sweeney, M.D., Chakhoyan, A., Pachicano, M., Joe, E., Nelson, A.R., D'Orazio, L.M., et al. (2020). APOE4 leads to blood-brain barrier dysfunction predicting cognitive decline. Nature *581*, 71–76. https://doi.org/10.1038/s41586-020-2247-3.

60. Marchetti, L., and Engelhardt, B. (2020). Immune cell trafficking across the blood-brain barrier in the absence and presence of neuroinflammation. Vasc. Biol. *2*, H1–H18. https://doi.org/10.1530/VB-19-0033.

61. Mildner, A., Schmidt, H., Nitsche, M., Merkler, D., Hanisch, U.K., Mack, M., Heikenwalder, M., Brück, W., Priller, J., and Prinz, M. (2007). Microglia in the adult brain arise from Ly-6ChiCCR2+ monocytes only under defined host conditions. Nat. Neurosci. *10*, 1544–1553. https://doi.org/10.1038/nn2015.

62. Pich, O., Bernard, E., Zagorulya, M., Rowan, A., Pospori, C., Slama, R., Huerga Encabo, H., O'Sullivan, J., Papazoglou, D., Anastasiou, P., et al. (2025). Tumor-Infiltrating Clonal Hematopoiesis. N. Engl. J. Med. *392*, 1594–1608. https://doi.org/10.1056/NEJMoa2413361.

63. Du S, Ou F, Drieu A, Xu EZ, Cheng Y, Storck SE, Mamuladze T, Cao J, Abduljawad N, Bhattarai B, et al. Brain-engrafted monocyte-derived macrophages from blood and skull-bone marrow exhibit distinct properties. Neuron. 2026;114:S0896-6273(26)00057-7. doi: 10.1016/j.neuron.2026.01.032.

64. Kim, J.S., Trzebanski, S., Shin, S.H., Schori, L., Frumer Friedman, G.R., Ilani, N.C., Kadam, A., Vicario, R., Aust, O., Bugaeva, P., et al. (2025). Clonal hematopoiesis-associated motoric deficits caused by monocyte-derived microglia accumulating in aging mice. Cell Rep. *44*, 115609. https://doi.org/10.1016/j.celrep.2025.115609.

65. Caunt, C.J., Sale, M.J., Smith, P.D., and Cook, S.J. (2015). MEK1 and MEK2 inhibitors and cancer therapy: the long and winding road. Nat. Rev. Cancer *15*, 577–592. https://doi.org/10.1038/nrc4000.

66. Mayer, I.A., and Arteaga, C.L. (2016). The PI3K/AKT Pathway as a Target for Cancer Treatment. Annu. Rev. Med. *67*, 11–28. https://doi.org/10.1146/annurev-med-062913-051343.

67. Petrone, G., Turker, I., Natarajan, P., and Bolton, K.L. (2024). Clinical and Therapeutic Implications of Clonal Hematopoiesis. Annu. Rev. Genomics Hum. Genet. *25*, 329–351. https://doi.org/10.1146/annurev-genom-120722-100409.

68. Gabitto, M.I., Travaglini, K.J., Rachleff, V.M., Kaplan, E.S., Long, B., Ariza, J., Ding, Y., Mahoney, J.T., Dee, N., Goldy, J., et al. (2024). Integrated multimodal cell atlas of Alzheimer's disease. Nat. Neurosci. *27*, 2366–2383. https://doi.org/10.1038/s41593-024-01774-5.

69. Kotini, A.G., Chang, C.J., Boussaad, I., Delrow, J.J., Dolezal, E.K., Nagulapally, A.B., Perna, F., Fishbein, G.A., Klimek, V.M., Hawkins, R.D., et al. (2015). Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. Nat. Biotechnol. *33*, 646–655. https://doi.org/10.1038/nbt.3178.

70. Wang, T., Pine, A.R., Kotini, A.G., Yuan, H., Zamparo, L., Starczynowski, D.T., Leslie, C., and Papapetrou, E.P. (2021). Sequential CRISPR gene editing in human iPSCs charts the clonal evolution of myeloid leukemia and identifies early disease targets. Cell Stem Cell *28*, 1074–1089.e7. https://doi.org/10.1016/j.stem.2021.01.011.

71. Fleming, S.J., Chaffin, M.D., Arduini, A., Akkad, A.D., Banks, E., Marioni, J.C., Philippakis, A.A., Ellinor, P.T., and Babadi, M. (2023). Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. Nat. Methods *20*, 1323–1335. https://doi.org/10.1038/s41592-023-01943-7.

72. Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. *8*, 281–291.e9. https://doi.org/10.1016/j.cels.2018.11.005.

73. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

74. Li, H., and Birol, I. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

75. Ianevski, A., Giri, A.K., and Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat. Commun. *13*, 1246. https://doi.org/10.1038/s41467-022-28803-w.

76. Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCR: visualizing classifier performance in R. Bioinformatics *21*, 3940–3941. https://doi.org/10.1093/bioinformatics/bti623.

77. Thiele, C., and Hirschfeld, G. (2021). cutpointr: Improved Estimation and Validation of Optimal Cutpoints in R. J. Stat. Softw. *98*, 1–27. https://doi.org/10.18637/jss.v098.i11.

78. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

79. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

80. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

81. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

82. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.

83. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

84. Khanna, A., Larson, D., Srivatsan, S., Mosior, M., Abbott, T., Kiwala, S., Ley, T., Duncavage, E., Walter, M., Walker, J., et al. (2022). Bam-readcount - rapid generation of basepair-resolution sequence metrics. J. Open Source Softw. *7*, 3722. https://doi.org/10.21105/joss.03722.

85. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498. https://doi.org/10.1038/ng.806.

86. Huang, A.Y., Zhang, Z., Ye, A.Y., Dou, Y., Yan, L., Yang, X., Zhang, Y., and Wei, L. (2017). MosaicHunter: accurate detection of postzygotic single-nucleotide mosaicism through next-generation sequencing of unpaired, trio, and paired samples. Nucleic Acids Res. *45*, e76. https://doi.org/10.1093/nar/gkx024.

87. Dunn, T., Berry, G., Emig-Agius, D., Jiang, Y., Lei, S., Iyer, A., Udar, N., Chuang, H.Y., Hegarty, J., Dickover, M., et al. (2019). Pisces: an accurate and versatile variant caller for somatic and germline next-generation sequencing data. Bioinformatics *35*, 1579–1581. https://doi.org/10.1093/bioinformatics/bty849.

88. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

89. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

90. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756. https://doi.org/10.1101/gr.239244.118.

91. Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. J. Stat. Softw. *42*, 1–28. https://doi.org/10.18637/jss.v042.i08.

92. De Jager, P.L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B.N., Felsky, D., Klein, H.U., White, C.C., Peters, M.A., Lodgson, B., et al. (2018). A multiomic atlas of the human frontal cortex for aging and Alzheimer's disease research. Sci. Data *5*, 180142. https://doi.org/10.1038/sdata.2018.142.

93. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311. https://doi.org/10.1093/nar/29.1.308.

94. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65. https://doi.org/10.1038/nature11632.

95. Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64–69. https://doi.org/10.1126/science.1219240.

96. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291. https://doi.org/10.1038/nature19057.

97. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.

98. Zook, J.M., McDaniel, J., Olson, N.D., Wagner, J., Parikh, H., Heaton, H., Irvine, S.A., Trigg, L., Truty, R., McLean, C.Y., et al. (2019). An open resource for accurately benchmarking small variant and reference calls. Nat. Biotechnol. *37*, 561–566. https://doi.org/10.1038/s41587-019-0074-6.

99. Miller, M.B., Huang, A.Y., Kim, J., Zhou, Z., Kirkham, S.L., Maury, E.A., Ziegenfuss, J.S., Reed, H.C., Neil, J.E., Rento, L., et al. (2022). Somatic genomic changes in single Alzheimer's disease neurons. Nature *604*, 714–722. https://doi.org/10.1038/s41586-022-04640-1.

100. Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature *570*, 332–337. https://doi.org/10.1038/s41586-019-1195-2.

101. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049. https://doi.org/10.1038/ncomms14049.

102. Huang, A.Y., Li, P., Rodin, R.E., Kim, S.N., Dou, Y., Kenny, C.J., Akula, S.K., Hodge, R.D., Bakken, T.E., Miller, J.A., et al. (2020). Parallel RNA and DNA analysis after deep sequencing (PRDD-seq) reveals cell type-specific lineage patterns in human brain. Proc. Natl. Acad. Sci. USA *117*, 13886–13895. https://doi.org/10.1073/pnas.2006163117.

103. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249. https://doi.org/10.1038/nmeth0410-248.

104. Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. *4*, 1073–1081. https://doi.org/10.1038/nprot.2009.86.

105. Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat. Rev. Cancer *18*, 696–705. https://doi.org/10.1038/s41568-018-0060-1.

106. Genin, E., Hannequin, D., Wallon, D., Sleegers, K., Hiltunen, M., Combarros, O., Bullido, M.J., Engelborghs, S., De Deyn, P., Berr, C., et al. (2011). APOE and Alzheimer disease: a major gene with semi-dominant inheritance. Mol. Psychiatry *16*, 903–907. https://doi.org/10.1038/mp.2011.52.

107. Herz, J., Filiano, A.J., Wiltbank, A.T., Yogev, N., and Kipnis, J. (2017). Myeloid Cells in the Central Nervous System. Immunity *46*, 943–956. https://doi.org/10.1016/j.immuni.2017.06.007.

108. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006. https://doi.org/10.1101/gr.229102.

109. Chang, C.J., Kotini, A.G., Olszewska, M., Georgomanoli, M., Teruya-Feldstein, J., Sperber, H., Sanchez, R., DeVita, R., Martins, T.J., Abdel-Wahab, O., et al. (2018). Dissecting the Contributions of Cooperating Gene Mutations to Cancer Phenotypes and Drug Responses with Patient-Derived iPSCs. Stem Cell Rep. *10*, 1610–1624. https://doi.org/10.1016/j.stemcr.2018.03.020.

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-NeuN Alexa Fluor 488 (Clone A60)	MilliporeSigma	Cat#MAB377X; RRID: AB_2149209
Rabbit monoclonal anti-CSF1R PE (Clone D3O9X)	Cell Signaling Technology	Cat#65396; RRID: AB_2799682
Mouse monoclonal anti-CD34 PE (Clone 563)	BD Biosciences	Cat#550761; RRID: AB_393871
Mouse monoclonal anti-CD45 APC (Clone HI30)	BD Biosciences	Cat#555485; RRID: AB_398600
Mouse monoclonal anti-CD45 PE-Cy7 (Clone 30-F11)	BD Biosciences	Cat#561868; RRID: AB_10893599
Mouse monoclonal anti-CD68 PE-Cy7 (Clone Y1/82A)	BD Biosciences	Cat#565595; RRID: AB_2739298
Mouse monoclonal anti-CD11b BB515 (Clone ICRF44)	BD Biosciences	Cat#564517; RRID: AB_2744271
Mouse monoclonal anti-CD14 APC (Clone M5E2)	BD Biosciences	Cat#555399; RRID: AB_398596
Rabbit polyclonal anti-CX3CR1 (N-Terminal)	Bio-Rad/AbD Serotec	Cat#AHP1589; RRID: AB_2087421
Rat monoclonal anti-TREM2 (Clone 237920)	R&D Systems	Cat#MAB17291; RRID: AB_2208679
Rabbit polyclonal anti-Iba1 (C-terminal)	FUJIFILM Wako	Cat#019-19741; RRID: AB_839504
Rabbit polyclonal anti-P2RY12	MilliporeSigma	Cat#HPA014518; RRID: AB_2669027
Donkey polyclonal Alexa Fluor 488-AffiniPure	Jackson Immuno Research Labs	Cat#711-545-152; RRID: AB_2313584
Goat polyclonal Alexa Fluor 488-Cross-Adsorbed	Thermo Fisher Scientific	Cat#A-11006; RRID: AB_2534074
Donkey anti-Rabbit IgG, Alexa Fluor 555	Thermo Fisher Scientific	Cat#A-31572; RRID: AB_162543
Donkey anti-Rat IgG, Alexa Fluor 488	Thermo Fisher Scientific	Cat#A-21208; RRID: AB_141709
Biological samples		
Genomic DNA and brain tissue samples from AD and control cases from ROSMAP	ROSMAP	see Table S2
Brain tissue samples from AD and control cases from SEA-AD	SEA-AD	H20.33.017; H21.33.002; H21.33.033; H21.33.017
Genomic DNA from the Genome in a Bottle Consortium	National Institute of Standards and Technology	NA12878; NA24695
Chemicals, peptides, and recombinant proteins		
DAPI, FluoroPure™ grade	Thermo Fisher Scientific	D21490
AMPure XP beads	Beckman Coulter	A63882
SPRIselect beads	Beckman Coulter	B23317
Streptavidin M-270 beads	ThermoFisher	65305
Percoll	Cytiva	Cat# 17-0891-02
SCF	R&D Systems	Cat# 255-SC
Flt3 ligand	R&D Systems	Cat# 308-FK
TPO	R&D Systems	Cat# 288-TP
IL-3	R&D Systems	Cat# 203-IL
BMP4	R&D Systems	Cat# 314-BP
FGF2	R&D Systems	Cat# 233-FB
CD200	Biotechne	Cat# 2724-CD
Fractalkine (CX3CL1)	Biotechne	Cat# 362-CX
IL34	Biotechne	Cat# 5265-IL
MCSF	Biotechne	Cat# 216-MC
TGFβ1	Biotechne	Cat# 7754-BH
VEGF	Thermo Fisher Scientific	Cat# PHC9391
GlutaMax Supplement	Thermo Fisher Scientific	Cat# 35050061
Insulin	Sigma	Cat # I9278

(*Continued on next page*)

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DMEM/F-12, HEPES, no phenol red	Thermo Fisher Scientific	Cat# 11039021
N2 Supplement	Thermo Fisher Scientific	Cat# 17502048
RPMI 1640 Medium, no phenol red	Thermo Fisher Scientific	Cat# 11835030
Matrigel	Corning	Cat# 354277
B27 Supplement	Thermo Fisher Scientific	Cat# 17504044
IMDM (Iscove's Modified Dulbecco's Medium)	Thermo Fisher Scientific	Cat# 12440053
Insulin-Transferrin-Selenium (ITS-G)	Thermo Fisher Scientific	Cat #41400045
KnockOut Replacement Serum (KOSR)	Thermo Fisher Scientific	Cat# 10828028
Monothioglycerol (MTG, 1-TG)	Sigma	Cat# M1753
Non-essential Amino Acids (NEAA)	Thermo Fisher Scientific	Cat# 11140050
CF1 Mouse Embryonic Fibroblasts, irradiated	Thermo Fisher Scientific	Cat# A34180
DAPI	Thermo Fisher Scientific	Cat# 62248
Critical commercial assays		
SureSelect XT HS2 DNA Reagent Kit	Agilent Technologies	G9983A
SureSelect XT Custom 0.5-2.9Mb Library	Agilent Technologies	5190-4819
Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1, 16 rxns	10X Genomics	PN-1000121
ResolveDNA Whole Genome Amplification Kit	BioSkryb Genomics	100136
Phusion Hot Start II DNA Polymerase Kit	Thermo Fisher Scientific	F549L
PrimeSTAR GXL DNA Polymerase	Takara Bio	R050A
EZ1 DNA Tissue Kit	Qiagen	953034
Quant-iT™ dsDNA Assay Kits, high sensitivty	Thermo Fisher Scientific	Q33120
High Sensitivity D5000 Reagents	Agilent Technologies	5067-5593
High Sensitivity D5000 ScreenTape	Agilent Technologies	5067-5592
Chromium Next GEM Single Cell ATAC Kit v2, 4 rxn	10X Genomics	1000406
Deposited data		
Panel sequencing data	This study	https://doi.org/10.7303/syn73904737
GoT-ChA Data	This study	https://doi.org/10.7303/syn73904737
SEA-AD Data	Gabitto et al.[68]	N/A
Experimental models: Cell lines		
iPSC lines	Kotini et al.[69]; Wang et al.[70]	N/A
Oligonucleotides		
Biotinylated primers	IDT	see Table S4
Software and algorithms		
Fiji	Open-source	RRID: SCR_002285
Adobe Illustrator	Adobe	RRID: SCR_010279
FlowJo	TreeStar Inc.	RRID: SCR_008520
Cell Ranger v6.0.0	10X Genomics	RRID: SCR_023221
CellBender v0.3.0	Fleming et al.[71]	RRID: SCR_018461
Scrublet v0.2.3	Wolock et al.[72]	RRID: SCR_018098
Seurat v4.1.1	Hao et al.[73]	RRID: SCR_007322
Minimap2 v2.24	Li and Birol[74]	RRID: SCR_018550
scType v20220909	Ianevski et al.[75]	https://github.com/IanevskiAleksandr/sc-type
ROCR v1.0.11	Sing et al.[76]	RRID: SCR_015843
cutpointr v1.1.2	Thiele and Hirschfeld[77]	https://github.com/Thie1e/cutpointr
CONICSmat v0.0.0.1	Müller et al.[45]	RRID: SCR_016655
BWA-MEM v0.7.15	Li and Durbin[78]	RRID: SCR_010910
BEDTools v2.30.0	Quinlan and Hall[79]	RRID: SCR_006646

(*Continued on next page*)

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Harmony v0.1.1	Korsunsky et al.[80]	RRID: SCR_022048
clusterProfiler v4.4.4	Wu et al.[81]	RRID: SCR_016884
CellRanger-ATAC v2.1.0	10X Genomics	RRID: SCR_017769
ArchR v1.0.3	Granja et al.[82]	RRID: SCR_021172
Samtools v1.15.1	Li et al.[83]	RRID: SCR_002105
bam-readcount v1.0.1	Khanna et al.[84]	https://github.com/genome/bam-readcount
AGeNT v2.0.2	Agilent Technologies	https://www.agilent.com/en/product/next-generation-sequencing/ngs-data-analysis-interpretation/agent-4301558
GATK v3.6	DePristo et al.[85]	RRID: SCR_001876
MosaicHunter v1.0	Huang et al.[86]	https://github.com/zzhang526/MosaicHunter
Pisces v5.3	Dunn et al.[87]	RRID: SCR_022117
Integrative Genomics Viewer v2.3.93	Robinson et al.[88]	RRID: SCR_011793
ANNOVAR v2015Mar22	Wang et al.[89]	RRID: SCR_012821
MAFTools v2.10.1	Mayakonda et al.[90]	RRID: SCR_024519
MatchIt v4.5.5	Ho et al.[91]	RRID: SCR_025618
dNdScv v0.0.1	Martincorena et al.[36]	RRID: SCR_023123
Other		
Ts2R-FL inverted microscope	Nikon Instruments Inc	N/A
DMI4000 B inverted microscope	Leica Microsystems, GmbH	N/A
Stem Cell Engineering Core	Icahn School of Medicine at Mount Sinai	RRID: SCR_027503

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Our study involves samples and sequencing data from two large-scale Alzheimer's disease (AD) studies, ROSMAP and SEA-AD. The ROSMAP study consists of two prospective studies of aging, the Religious Order Study (ROS) and the Memory and Aging Project (MAP), in which the participants were enrolled by the Rush Alzheimer's Disease Center with detailed cognitive and neuroimaging phenotyping as well as structured neuropathologic examination during the autopsy at the time of death.[92] The SEA-AD study performed single-cell multi-omics, quantitative neuropathology, and deep clinical phenotyping on post-mortem brain tissue from 84 aged donors and 5 additional younger neurotypical controls collected at the University of Washington BioRepository and Integrated Neuropathology laboratory and Precision Neuropathology core. Postmortem samples in all studies were collected and de-identified following the protocol of the corresponding Institutional Review Board with informed consent. The diagnosis of AD was based on the consensus conclusion from all postmortem data generated by neurologists with expertise in dementia and neurodegeneration.

For panel sequencing, we obtained genomic DNA (gDNA) extracted from the prefrontal cortex (PFC) samples of 190 AD patients and 121 controls without cognitive impairment from ROSMAP (Table S2), and the detailed demographic and clinical information for these samples were downloaded from the AMP-AD Knowledge Portal. Additional brain samples of multiple brain regions and gDNA from peripheral blood samples were also obtained from ROSMAP to confirm the presence of somatic variants and further study the cell type identity of variant-carrying cells. The raw single-nucleus RNA sequencing (snRNAseq).h5 matrices for SEA-AD and corresponding clinical and technical metadata were also downloaded from the AMP-AD Knowledge Portal, which consists of the middle temporal gyrus of temporal cortex from 31 AD patients and 32 age-matched controls (Table S5). Detailed demographic and sample information for the ROSMAP and SEA-AD datasets, including age and sex, are provided in Tables S2 and S5; however, ethnicity information is not available. Within each dataset, AD and control samples exhibited comparable distributions of sex, age, postmortem interval, and sequencing depth (Tables S2 and S5). In addition, we adjusted for the potential effects of these variables, including age and sex, in the linear regression models described in the following sections. We also sequenced PFC gDNA from 31 individuals with mild cognitive impairment (MCI) from ROSMAP in the same panel sequencing batch, and performed somatic variant calling along with the 190 AD patients and 121 healthy controls (Table S3). However, due to the limited sample size, the MCI group was not included in any downstream analyses, except for amplicon validation experiments used to benchmark the somatic variant calling accuracy of our pipeline.

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

METHOD DETAILS

Panel design and sequencing

For hybrid capture, probes targeting the exons and exon-intron junctions of 149 cancer driver genes (Table S1) were designed using the SureSelect DNA Advanced Design Wizard. The list of targeted genes was designed to include frequently mutated oncogenes and tumor suppressor genes in various types of cancer and clonal hematopoiesis. A total of 23,171 probes with a genomic size of 691 kbp were eventually designed and generated. These probes were then used for gene capture followed by library preparation using the SureSelect XT HS2 DNA Reagent Kit with 30 ng gDNA input. All prepared libraries were sequenced using three Illumina NovaSeq 6000 S4 flow cells with 150 bp paired-end reads.

Somatic variant calling from panel sequencing

The UMI information of each read was first extracted from the fastq files by AGeNT's Trimmer (v2.0.2), and then reads were aligned to the GRCh37 human reference genome by BWA-MEM (v0.7.15) [78] The aligned reads were processed by AGeNT LocatIt (v2.0.2) to generate the consensus read sequence from multiple reads that were derived from the same original DNA fragment and thus carried the same UMI, followed by GATK's indel realignment (v3.6).[85] We only kept the consensus reads that were supported by two or more reads in both strands. As a result, we achieved comparable depth and coverage between the AD and control samples, with more than 1000X average depth and more than 80% coverage of the targeted regions at >500X for consensus reads (Table S2; Figure S1A).

Somatic single-nucleotide variants (sSNVs) and somatic insertions and deletions (sIndels) were called from the consensus reads by MosaicHunter (v1.0)[86] and Pisces (v5.3),[87] respectively. For sSNV, MosaicHunter calculated the likelihood of the presence of a mutant allele, and only the candidates with a 0.5 or higher likelihood, 100 or more total reads, and 4 or more mutant-supporting reads were considered. We further excluded candidates as germline variants if i) they have a 30% or higher variant allele fraction; 2) the counts of mutant-supporting and total reads do not significantly deviate from the binomial distribution for heterozygous variants ($p \geq 0.05$); 3) they are present in the polymorphism databases (dbSNP,[93] the 1000 Genomes Project,[94] the Exome Sequencing Project,[95] and the Exome Aggregation Consortium[96]) or have a 0.01% or higher population allele frequency in the Genome Aggregation Database.[97] sIndels were called by Pisces with its default parameters, and a similar method was used to call variant candidates and remove germline variants.

To balance the sensitivity and specificity of our sSNV and sIndel detection, we developed two different pipelines when considering the recurrent presence across multiple individuals. The "stringent" pipeline only kept the variants that were detected in one sample and were completely absent in any other samples, whereas the "sensitive" pipeline additionally allowed the variants that were exclusively present or specifically enriched (two-sample Z-test of proportion with $p < 0.05$) in the AD or control group.

Benchmarking of variant calling using panel sequencing

A mixing experiment was performed to benchmark the performance of the designed panel and variant calling pipeline. Germline variant calls from two unrelated individuals, NA12878 and NA24695, were downloaded from the website of the Genome in a Bottle Consortium.[98] Genomic sites in the covered regions of panel sequencing that were genotyped as heterozygous in NA24695 but reference-homozygous in NA12878 were considered as the gold-standard list of somatic variants, and gDNA from these two individuals was mixed to reach 10%, 5%, 2%, 1%, 0.5%, and 0.2% mutant allele fractions for these variants. We applied the same experiment and analysis protocols of panel sequencing to the mixed samples with varied allele fractions, and then checked the proportion of gold-standard variants that were identified by our identification pipeline as well as the consistency between expected and observed allele fractions.

Fluorescence-activated nuclei sorting (FANS) for brain samples

Nuclei were prepared following the previously published work.[99] Briefly, fresh frozen human brain tissue samples were first lysed in a dounce homogenizer using a chilled nuclear lysis buffer (10mM Tris-HCl, 0.32M Sucrose, 3mM Mg(Acetate)2, 5mM CaCl2, 0.1mM EDTA, pH 8, 1mM DTT, 0.1% Triton X-100) on ice. Tissue lysates were layered on top of a sucrose cushion buffer (1.8 M sucrose, 3 mM Mg(OAc)2, 10 mM Tris-HCl, 1 mM DTT, pH 8) and ultra-centrifuged for 1 h at 30,000g. Nuclear pellets were resuspended in ice-cold PBS supplemented with 3mM MgCl2, filtered, and then stained with the neuronal marker (NeuN, Millipore MAB377) or brain macrophage marker (CSF1R, Cell Signaling 65396) together with DAPI. For each AD PFC sample, neuronal (NeuN+), glial (NeuN-), brain macrophage (CSF1R+), and total (DAPI+) nuclei populations were sorted into 96-well plates by flow cytometry. For AD brain samples from other brain regions and control brain samples, only CSF1R+ and NeuN+ nuclei populations were sorted. To note, the purity of our microglia-enriched brain macrophage sorting was slightly reduced in some AD samples, exhibiting a marginally higher proportion of non-microglia/CAM nuclei as determined by snRNAseq (Figure S3A). This likely stems from autofluorescence caused by apoptotic cells and increased cellular debris. Our subsequent single-cell multi-omic analyses showed that somatic driver mutations are generally absent in non-microglia/CAM brain cell types (Figures S5E and S5F), indicating that they are unlikely to account for the observed VAF increase in CSF1R+ nuclei in AD brains.

Cell type analysis from 10X snRNAseq

For the PFC sample of one AD patient (with a *TET2* p.Pro1194Ser sSNV) and one healthy control, ten thousand CSF1R+ nuclei were sorted separately into a well of the 96-well plate and used for droplet generation and sequencing library preparation using the 10x Genomics Next GEM Single Cell 3′ GEM Kit v3.1 and Chromium Controller, following the manufacturer's manual. The snRNA-seq libraries were sequenced by Illumina HiSeq X, and down-sampled to have a comparable sequencing throughput. We also downloaded a large-scale snRNAseq dataset,[100] consisting of 80,660 nuclei isolated from 24 AD and 24 control PFC samples collected by ROSMAP, to serve as the reference. The sequencing data of our AD and control samples were first processed by Cell Ranger (v6.0.0)[101] and then integrated and analyzed along with the reference dataset by Seurat (v4.9.9),[73] for variance normalization, anchor-based reciprocal principal-component analysis (PCA) integration, PCA clustering, and UMAP visualization. Cell clusters were manually annotated into different cell types based on the expression profile of marker genes (Figure S4B) for the major brain[102] and blood[8] cell types (*HBA1*: red blood cell; *CD3E*: T-cell; *CCR7*: B-cell; *FCN1*: monocyte). Our snRNAseq result confirmed that 75-77% of CSF1R+ sorted nuclei from the AD and control brains are transcriptionally microglia, with an additional 4-9% being CNS-associated macrophages (CAM; Figure 3A). We also observed minimal blood contamination in the sorted CSF1R+ population, with only 1% monocytes and the absence of other major blood cell types including red blood cells, T-cells, and B-cells (Figures 3A and S4B). Using this reference dataset, we also confirmed the minimal contamination of blood cells (< 0.3%) in ROSMAP brain samples.

Amplicon sequencing

Amplicon sequencing was performed for validation and mutant allele fraction estimation in both bulk gDNA samples and sorted nuclei. Bulk gDNA was extracted from frozen brain samples using the EZ1 DNA Tissue Kit (Qiagen 953034). Five hundred nuclei of each cell type from each brain sample were sorted into 96-well plates with four replicates. Whole-genome amplification was then performed for sorted nuclei using the ResolveDNA Whole Genome Amplification Kit (BioSkryb Genomics) to meet the minimal DNA amount for panel sequencing. For each identified sSNV, three sets of primers were designed for PCR amplification of the targeted genomic region. PCR amplification was performed using the Phusion Hot Start II DNA Polymerase kit (Thermo Fisher F549L) with the following cycles: 98 °C for 30 sec; 5 cycles of 98 °C for 10 sec, 68 °C for 30 sec (decrease 1 °C/cycle), and 72 °C for 30 sec; 25 cycles of 98 °C for 10 sec, 63 °C for 30 sec, 72 °C for 30 sec; 72 °C for 10 min. The annealing temperatures of primers varied for each design which was determined by a testing PCR. PCR products were then purified using AMPure XP beads (Beckman Coulter A63882) and pooled for Amplicon-EZ sequencing (GENEWIZ).

The sequencing reads were first aligned to the GRCh37 human reference genome by BWA-MEM (v0.7.15) and then processed by GATK (v3.6) for indel realignment. For each somatic variant candidate, the number of reads supporting each allele was calculated by MosaicHunter (v1.0) and manually verified by Integrative Genomics Viewer (v2.3.93).[88] A candidate was considered validated as somatic variant (Figures S1E–S1G) if 1) the read fraction of the mutant allele was more than three times as high as the fractions of the other two error alleles in all three amplicons (somatic-I); or 2) the read fraction of the mutant allele in the corresponding brain sample was significantly higher than the fraction in an unrelated negative control brain sample for all three amplicons (somatic-II).

Functional annotation of sSNV and sIndel

ANNOVAR (v2015Mar22)[89] was applied to annotate somatic variants into different genic categories: 5' UTR, exonic (coding sequence), 3' UTR, splicing (within intronic 2 bp of a splicing junction), and intronic. Exonic somatic variants were further classified into multiple categories based on their predicted impacts on amino acids. A somatic variant was labeled as deleterious if 1) it was annotated as splicing or predicted to cause stop-codon gain/loss; 2) it was a frameshift insertion or deletion; or 3) it was a missense variant whose amino acid change was predicted to be deleterious by either PolyPhen2[103] or SIFT.[104] For 149 cancer driver genes, we grouped them into (proto-)oncogenes and tumor suppressor genes (TSGs) according to the annotation of the COSMIC Cancer Gene Census.[105] Genes annotated as both oncogenes and TSGs were not considered in calculating the variant burdens plotted in Figure 2D. MAFTools (v2.10.1)[90] was used to illustrate the gene-level distribution of somatic variants. Genes and driver variants involved in clonal hematopoiesis of indeterminate potential (CHIP) were extracted from a curated list summarized in Table S1 of Bouzid et al.,[25] derived from the blood whole-genome sequencing data from 97,691 people.[37]

Burden analysis of sSNV and sIndel

Somatic variant density in each clinical group was calculated by counting the total number of somatic variants and dividing it by the total size of powered genomic regions with $\geq$500X UMI-aggregated reads, and the odds ratio and the two-sample Z-test of proportion were used to test whether the AD group had a higher mutational burden than the control group. In the gene-level analysis, we compared the somatic variant burden between AD and control groups using a similar two-sample Z-test of proportion, in which the total genomic size for each gene was calculated as the product of the exonic length and the number of individuals in AD or control group.

For the linear regression analysis, the count of somatic variants in each sample was modeled as a continuous outcome, whereas clinical status and other covariates of interest (e.g., age, sex, sequencing depth, and post-mortem interval) were modeled as independent variables. Our linear regression results from panel sequencing confirmed the increased burden of somatic variants in AD brains after controlling for all of these potential confounding factors (Figure 2C). We only considered donors with ages less than 90, because all donors aged 90 or higher were labeled as "90+" in the demographic tables of the ROSMAP study.

To further control for the potential impact of age on our burden analysis, we selected a subset of AD patients (n = 121) with an age distribution matched to that of the healthy controls. This was done using the R package MatchIt (v4.5.5),[91] with the function

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

matchit(Cogdx ~ Age + Sex + PMI, distance = "glm", method = "optimal"), where Cogdx indicates clinical status (AD or control), and PMI denotes post-mortem interval. The resulting subset of AD patients closely mirrored the age distribution of the control group (Figure S3E). We replicated our burden analyses using this matched subcohort and observed results that were highly consistent with those obtained from the full cohort (Figures S3F–S3H).

In addition to the linear regression analysis, we performed logistic regression, modeling clinical status as a binary outcome and including the count of somatic variants along with potential confounding factors such as age, sex, sequencing depth, post-mortem interval, and *APOE* genotype as independent variables. We observed a significant positive association ($p < 0.05$) between somatic variant burden and AD risk, regardless of whether *APOE* genotype was modeled additively for the APOE4 allele[106] (Figure S3K) or as categorical factors representing all genotypes (22, 23, 33, 34, 44).

Positive selection analysis

Signals of positive selection were assessed for sSNVs identified from AD and control samples separately by dNdScv (v0.0.1).[36] The dN/dS ratios and p-value for missense, nonsense, and splicing variants were calculated at the levels of individual genes and groups of genes, by comparing against the background synonymous variant rate with the consideration of the sequence composition of genes. For each gene in AD or control group, we 1) calculated the number of missense and truncating (nonsense and splicing) variants under positive selection by multiplying the number of all variants in that gene by the proportion of positively selected variants inferred from the gene-specific dN/dS ratio; 2) determined the proportion of positively selected cells by multiplying the number of positively selected variants by the average mutant allele fraction in that gene × 2 (given that almost all the sSNVs should be heterozygous in carrier cells). Assuming a consistent number of profiled cells in panel sequencing for each brain, we further estimated the number of positively selected cells in each AD and control brain by aggregating the number of positively selected cells across the group of genes and normalizing this number based on the count of brain samples in AD and control groups.

Simultaneous genotyping and transcriptomic analysis of single cells

The PFC tissue of an AD case with a somatic p.Pro1194Ser *TET2* variant was used for sorting ~10,000 CSF1R+ nuclei. The sorted nuclei were then followed by snRNAseq library preparation using 10x Genomics Next GEM Single Cell 3′ GEM Kit v3.1 and Chromium Controller, following the manufacturer's manual. Ten ng of the snRNAseq cDNA was used to perform semi-nested long-range PCR to capture the target variant using the PrimeSTAR GXL DNA Polymerase, following the manufacturer's manual. The long-range PCR product was then used for long-read sequencing library preparation using the ONT Kit12 chemistry. The final library was sequenced on one MinION Flow Cell.

To perform transcriptomic analysis and cell type identification, CellBender (v0.3.0)[71] with all default parameters was used to remove ambient RNA background from the generated 10X snRNAseq data, and Scrublet (v0.2.3)[72] was used to remove putative doublets. All downstream analysis was performed with Seurat (v4.1.1). High-quality nuclei with 1) $\leq$ 5% mitochondrial gene expression and 2) > 250 expressed genes and < 7500 expressed genes were extracted, where expressed genes were defined as those expressed in $\geq$ 3 cells. SCTransform was used to perform normalization and regression of the technical covariates percent.mt, nFeature_RNA, and nCount_RNA. Unsupervised clusters were identified with the Louvain algorithm (resolution = 0.6), and brain cell types were annotated based on canonical cell type markers.

To perform genotyping of the *TET2* variant, MinION reads were aligned with Minimap2 (v2.24)[74] to the GRCh38 human reference genome. Reads were parsed using regular expressions to identify known features (i.e., forward primer, 10X cell barcode, 10X UMI, poly-A tail), and the genotype status of *TET2* (i.e., mutant or wild-type) was assessed on a per-read basis. Reads that did not carry a 10X cell barcode or UMI retained in the snRNAseq analysis were filtered. We first calculated the consensus genotype (reference or mutant) for each UMI by using the Bayesian genotyper of MosaicHunter with the consideration of base quality and number of supporting reads, and then assigned genotype status (reference-homozygous or heterozygous) to each cell based on UMI information and the mutant cell fraction of this *TET2* variant using the framework of scMosaicHunter.[102]

Simultaneous genotyping and chromatin accessibility analysis of single cells

Nuclei were prepared as described previously.[99] Nuclei were stained with CSF1R+ PE-conjugated antibody before sorting with flow cytometry into 1.5 mL tubes containing 300 μL lysis dilution buffer. Detergent was added to lyse cells followed by centrifugation and buffer exchange into diluted 10X Single Cell Nuclei Buffer. Nuclei were subsequently processed according to the Chromium Next GEM Single Cell ATAC Solution user guide (version CG000209 Rev F, 10x Genomics) with the modifications described as published to generate ATAC and GoT-ChA libraries.[40] Namely, locus-specific primers at 22.5 uM concentration were added to the cell barcoding PCR reaction mixture prior to droplet generation. A 5 μL aliquot of eluted material post GEM incubation cleanup (Step 3.2 of 10X protocol) was further amplified with locus-specific primers to add sequencing adapters and generate a GoT-ChA genotyping library. Single-nucleus ATAC sequencing (snATACseq) libraries were sequenced to a depth of 25,000 read pairs per nucleus, and GoT-ChA libraries were sequenced to 5,000 read pairs per nucleus.

snATACseq data alignment and chromatin accessibility quantification were performed using cellranger-ATAC (v2.1.0). All subsequent downstream processing was performed using ArchR (v1.0.3).[82] Low-quality nuclei with less than 1000 fragments or a transcription start site enrichment of less than four were filtered, as were nuclei identified as putative doublets with ArchR. Dimensionality reduction was performed with iterative latent semantic indexing, followed by batch correction with harmony across samples and unsupervised clustering. Finally, a UMAP embedding was created to facilitate 2-D visualization. High-quality MLBMs or neurons were annotated by examining both genome browser tracks and gene activity for microglia marker genes *CX3CR1*, *CSF1R*, *TMEM119*, *P2RY12*, *MS4A7*, and *MRC1* as well as neuron marker genes *RBFOX3*, *MAP2*, and *SYN1*. MLBMs were assigned to putative microglial states based on gene activity of five known microglial state modules.[40] In brief, average gene activity was calculated per unsupervised cluster for each of these five microglial state modules. Min-max scaling was then applied individually to each microglial state module's gene activity vector to control for differences in overall gene activity between different microglial state modules. Finally, unsupervised clusters were assigned to microglial states based on the cosine similarity of each unsupervised cluster's gene activity vector to individual unit vector representations of each of the five tested microglial state modules.

To perform genotyping, GoT-ChA libraries were filtered to reads from the nuclei retained in the snATACseq analysis that overlapped the variant site and matched the expected primer sequence within an edit distance of <=3. Alt and ref reads were counted per nucleus using Samtools mpileup (v1.15.1)[83] for sSNVs or bam-readcount (v1.0.1)[84] for sIndels. Final genotype assignments were determined using the Bayesian genotyper of scMosaicHunter described in the previous section, but without the step of UMI-based consensus calling to account for the lack of UMIs in 10X snATACseq data. One-tailed Fisher's exact tests were used to test for cell type and microglial state enrichment analyses and were performed both with all samples pooled to maximize power as well as per individual sample. For per-individual sample analysis, a Benjamini-Hochberg correction was used to control for multiple-hypothesis testing.

Automatic cell-type identification with scType

Brain macrophages include both microglia and CAMs, such as meningeal, choroid plexus, and perivascular macrophages (PVMs).[107] Microglia-perivascular macrophages, hereby referred to as microglia-CAMs, represented 3.37% of all pre-annotated cells within SEA-AD, which is slightly lower than past estimates of microglia making up about 5% of brain cells in cortex.[6] scType (v20220909)[75] was used to automatically identify any additional high-quality microglia-CAMs beyond those originally annotated in SEA-AD ("pre-annotated" cells) to increase statistical power for calling mosaic chromosomal alterations (mCAs). Excitatory neurons (ExNs) were also automatically typed as a cell-type out-group to further facilitate accurate identification of microglia-CAMs, as scType'd microglia-CAMs should have high microglia-CAM scType scores but low ExN scType scores.

Prior to running scType, each SEA-AD sample was processed, normalized, and clustered with the Louvain algorithm using Seurat (v4.1.1). Each sample underwent quality control with the following metrics: retain only 1) genes expressed in $\geq$ 3 cells, 2) cells with $\geq$ 10 expressed genes, 3) cells with $\leq$ 5% mitochondrial gene expression, 4) cells with > 250 expressed genes and < 7500 expressed genes. Positive markers for microglia-CAMs (*P2RY12, ITGAM, CD40, PTPRC, CD68, AIF1, CX3CR1, TMEM119, ADGRE1, C1QA, NOS2, TNF, ISYNA1, CCL4, ADORA3, ADRB2, BHLHE41, BIN1, KLF2, NAV3, RHOB, SALL1, SIGLEC8, SLC1A3, SPRY1, TAL1*) and ExNs (*SLC17A7, SLC17A6, GRIN1, GRIN2B, GLS, GLUL, GRIN2A*) were downloaded from the scType marker database and used to calculate microglia-CAM and ExN scType scores for each individual cell.

In brief, scType calculates cell-type-specific scores for each cell using a weighted and normalized aggregation of marker gene expression, where marker genes are weighted more highly if they are more specific for a given cell type (expressed in one cell type of interest, rather than several). For each sample, both ExN and microglia-CAM scType scores were calculated for cells that were pre-annotated as either ExNs or microglia-CAMs. Taking these pre-annotations as ground truth, ROCR (v1.0.11)[76] and cutpointr (v1.1.2)[77] were used to calculate the optimal cutpoint for ExN and microglia-CAM scType scores that maximized the sum of sensitivity and specificity of classification over 1000 bootstraps. Using these learned ExN and microglia-CAM cutpoints, cells that were not pre-annotated were assigned as ExNs, microglia-CAMs, or neither. A small number of cells had both microglia-CAM and ExN scType scores greater than the corresponding optimal cutpoints; these cells were discarded due to ambiguity in assignment.

In addition to filtering of individual cells, 6 samples were filtered out due to not meeting at least one of the following sample-specific metrics: 1) microglia-CAM AUC > 0.9, 2) ExN AUC > 0.9, 3) fraction of pre-annotated ExN typed by scType as microglia-CAM < 0.1, and 4) total number of pre-annotated and scType'd microglia-CAMs > 50. This analysis filtered one individual H20.33.008, as this donor had only one associated sample that was filtered due to not meeting the above sample-specific metrics.

As a final step to ensure that scType'd microglia-CAMs were highly similar to their corresponding pre-annotated counterparts, pre-annotated and scType'd microglia-CAMs derived from the same donor were merged into a single Seurat object and processed, normalized, and clustered using the Louvain algorithm. Clusters in which over 50% of cells were pre-annotated microglia-CAMs were identified, and only scType'd microglia-CAMs in these clusters were retained as high-confidence scType'd microglia-CAMs. Only pre-annotated microglia-CAMs and these high-confidence scType'd microglia-CAMs were used for mCA-calling and all subsequent downstream analyses.

mCA calling from brain snRNAseq

Genomic regions of non-uniparental disomy CH-related mCA listed in Extended Data Figure 4d and 4e of Saiki et al.[48] were extracted, and genomic coordinates of these regions were downloaded from the hg38 reference genome accessed through the UCSC Genome Browser.[108]

mCA calling was done for microglia-CAMs, astrocytes, oligodendrocytes, oligodendrocyte precursor cells (OPCs), and ExNs. For each cell type, raw count matrices (gene × cell) were extracted for the 31 AD cases and 31 age-matched healthy controls that passed filtering as described above. Each of these matrices was processed and normalized using Seurat (v4.1.1) and then independently used as input for mCA-calling with CONICSmat (v0.0.0.1).[45]

The aforementioned mCA regions identified in Saiki et al. were tested with CONICSmat (Table S6), and raw mCA calls were further filtered to increase specificity of calls. In brief, a putative mCA was retained if it met the following criteria: 1) Bonferonni adjusted p-value < 0.05; 2) $<25\%$ ambiguous cells (cells with a posterior probability > 0.25 and < 0.75); 3) median expression of putative mCA-carrying cells is $>$ or $<$ 1.96 standard deviations of putative normal cells of the same type for amplifications or deletions, respectively; 4) no negative control regions (i.e. whole chromosome regions that have not been associated with mCA in past literature) showed a larger difference in expression between putative normal and mCA-carrying cells than the called mCA; 5) the expression of putative normal cells was within 1.96 standard deviations of baseline expression of cells of the same type across all other individuals; and 6) the same mCA was not called in a different cell-type from the same individual. For microglia-CAMs, putative mCAs were additionally filtered if the number of scType'd non-ambiguous cells (posterior probability < 0.25 or > 0.75) was ≤ 1.5 times the number of pre-annotated non-ambiguous cells for both altered and wild-type cells. This filtering criterion was added to ensure that mCA calls identified from scType'd and pre-annotated microglia-CAMs were not driven by added scType'd cells.

mCA validation using whole-genome sequencing

The middle temporal gyrus tissues from AD and control cases from the SEA-AD cohort were used for sorting ~500 neuronal (NeuN+) and brain macrophage (microglia and CAM; CSF1R+) nuclei, followed by scWGA using the ResolveDNA Whole Genome Amplification Kit (BioSkryb Genomics) as mentioned above. The amplified single-cell gDNA was then used for library preparation using the ResolveDNA Library Preparation Kit (BioSkryb Genomics), followed by 30X whole-genome sequencing on Illumina NovaSeq X.

Whole-genome sequencing reads from sorted NeuN+ and CSF1R+ nuclei were aligned to the GRCh37 human reference genome separately by BWA-MEM (v0.7.15). We calculated the coverage of sorted NeuN+ and CSF1R+ nuclei for each donor over 10^6 genomic windows using BEDTools (v2.30.0).[79] We then calculated the ratio of normalized, average coverage in the regions of called mCAs between sorted NeuN+ and CSF1R+ nuclei (macrophage-to-neuron ratio) from the same donor. This calculation was performed across all donors, with donors in which a given mCA was not called serving as a negative control. An mCA was considered validated if its macrophage-to-neuron ratio significantly deviated from the window of 0.99-1.01, and such deviation could only be seen in the donors in which the mCA was originally called.

Burden analysis of mCA

Per cell type, the number of cells with mCAs from AD donors, the number of cells without mCAs from AD donors, the number of cells with mCAs from control donors, and the number of cells without mCAs from control donors were counted, and an odds ratio (OR) of mCA-carrying cells in AD donors vs control donors was calculated. For two cell types, CAMs and oligodendrocytes, all mCA-carrying cells were in AD donors and the OR was thus infinite. To facilitate comparison of the actual OR against an empirical null as described below, a pseudocount of 1 was added to the number of mCA-carrying cells in AD and control groups separately for these two cell types. To calculate the significance level of this calculated odds ratio, an empirical null was generated using permutation. In brief, for each cell type, diagnosis labels were permuted over the set of all cells from each donor, including both mCA-carrying and wild-type cells. If a donor had multiple called mCAs, diagnosis labels were permuted over each mCA individually. Specifically, for each called mCA in a given individual, cells were divided into wild-type or mCA-carrying for that specific mCA. Each of these partitions of wild-type versus mCA-carrying cells was then randomly assigned a diagnosis status. OR was calculated for each set of permutated data. Permutations were repeated 1000 times and the p-value of the actual OR was calculated as 1 – the percentile rank of actual OR against the empirical null distribution of permutation ORs. Ten trials of 1000 permutations were completed to ensure the robustness of p-values.

Creation of an integrated snRNAseq microglia-CAM atlas

All scType'd and pre-annotated microglia-CAMs from AD and healthy control samples, except the one associated with *H20.33.008* as described above, were individually processed with Seurat (v4.1.1). In brief, each sample underwent quality control, and only cells with 1) $\leq 5\%$ mitochondrial gene expression and 2) > 250 expressed genes and < 7500 expressed genes were retained, where expressed genes were defined as those expressed in ≥ 3 cells. Variance-stabilizing normalization and regression of the technical covariates percent.mt, nFeature_RNA, and nCount_RNA were performed with Seurat function SCTransform, and clustering was done using the Louvain algorithm.

Individual samples were then merged into a single Seurat object, and dimensionality reduction was performed using PCA. This merged object was then integrated over constituent individual samples using Seurat's wrapper function for Harmony (v0.1.1).[80] UMAP visualization of the integrated object showed no visible clustering by sample ID or individual ID, consistent with successful integration (Figure S7).

Please cite this article in press as: Huang et al., Somatic cancer variants enriched in Alzheimer's disease microglia-like cells drive inflammatory and proliferative states, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.040

Differential expression analysis and functional annotation of integrated microglia-CAM snRNAseq atlas

Differential expression analysis was performed between microglia-CAMs with and without called mCAs from mCA-carrying AD individuals using the FindMarkers function of Seurat (v4.1.1) with a min.pct cutoff of 0.10 and no fold-change cutoff. Genes with a Bonferroni-adjusted p-value < 0.05 were called as differentially-expressed genes (DEGs).

clusterProfiler (v4.4.4)[81] was used to perform all enrichment analyses. GO enrichment analysis was performed using standard parameters and a universe of all genes expressed in >10% of microglia-CAMs in the integrated atlas. Terms were deemed significant if they had a Bonferroni-adjusted p-value < 0.05.

DEGs were also tested for enrichment of previously defined microglial state gene modules.[41] A minority of genes (107/905; 11.9%) within these microglial state gene modules were shared between multiple modules. To ensure specificity of module enrichment, genes were weighted by the inverse of the number of modules in which they were present. Non-integer values were rounded and module enrichment was tested using a hypergeometric test.

Gene editing of human iPSCs

The previously described normal induced pluripotent stem cell (iPSC) line N-2.12-D-1-1 served as the parental line of all CRISPR/Cas9-edited lines used in this study.[69] The gene editing strategy used to generate iPSCs with *ASXL1* C-terminus truncation was previously described.[70] CRISPR/Cas9-mediated homology-directed repair (HDR) was used to introduce the *DNMT3A* R882H variant using co-delivery of a mutant and a wild-type (WT) donor template, as previously described.[70] Nucleofection of a plasmid expressing the gRNA and Cas9 with mCitrine and clone selection by Restriction Fragment Length Polymorphism (RFLP) were performed as previously described.[70] Briefly, the N-2.12-D-1-1 iPSC line was cultured in human pluripotent stem cell (hESC) media containing 10 mM Y-27632 for at least one hour before nucleofection. The cells were dissociated into single cells with accutase and 1 million cells were used for nucleofection with 5 μg of gRNA/Cas9 plasmid and 5 μg of each donor plasmid (WT and G12D) using Nucleofector II (Lonza). mCitrine+ cells were sorted by fluorescence-activated cell sorting (FACS) 48 hr after transfection and plated at clonal density. Single colonies were screened by PCR and RFLP with DdeI restriction enzyme. We used CRISPR/Cas9 to generate iPSC lines with a mono-allelic genomic deletion of ∼45 kb spanning exons 3-11 of the *TET2* gene. Two different gRNAs targeting the *TET2* locus—gRNA1 targeting intron 2 and gRNA2 targeting intron 11—were designed and cloned downstream of the U6 promoter in the gRNA/Cas9 plasmid. Nucleofection was performed as above with 5 μg of each gRNA/Cas9 plasmid. Single-cell clones were screened with PCR with primers F1: CGGACATTTAGGTTGATTCCA and R2: GCAGACTGCCATTCAGTTCTC amplifying the junction generated after deletion. Selected clones were further screened to identify clones with mono-allelic deletion with PCR with primers F2: CGGACATTTAGGTTGATTCCA; R2: CTTGGCTTACCCCGAAGTTA (amplifying the area around the cutting site of gRNA1) and F3: CATGAGATGGATGGCCACTT; R3: GCAGACTGCCATTCAGTTCTC (amplifying the area around the cutting site of gRNA2). Multiple independent clones with each variant were isolated and their genotype were verified by Sanger sequencing. All clones were confirmed to have a normal karyotype.

Human iPSC culture and iMGL differentiation

Culture of human iPSCs on mitotically inactivated mouse embryonic fibroblasts (MEFs) was performed as previously described.[109] Hematopoietic differentiation was performed using a spin-EB protocol previously described.[109] For differentiation into iMGLs, day 11 HSPCs were transferred to DMEM/F12 1:1 medium with 4% B27, 0.5% N2, 1% nonessential amino acids (NEAA), 1 mM L-glutamine, Insulin-Transferrin-Selenium (ITS-G), 400uM monothioglycerol (1-Thioglycerol), and 5 μg/ml insulin, supplemented with 100 ng/ml interleukin (IL)-34, 50 ng/ml TGFβ, and 25 ng/ml macrophage colony-stimulating factor (M-CSF) for 40 days with media changes every two days and splitting/replating every week. After 40 days, 100 ng/ml CX3CL1 and 100 ng/ml CD200 were added to the media to induce further maturation of iMGLs for 3 days. Human iPSC experiments were supported by the Stem Cell Engineering Core (RRID: SCR_027503) at the Icahn School of Medicine at Mount Sinai.

Flow cytometry of iPSC-derived hematopoietic stem and progenitor cells and iMGLs

The following antibodies were used: CD34-PE (clone 563, BD Pharmingen), CD45-APC (clone HI30, BD Pharmingen), mCD45-PE-Cy7 (clone 30-F1, BD Pharmingen), CD68-PE-Cy7 (clone Y1/82A, BD Pharmingen), CD11b-BB515 (clone ICRF44, BD Horizon), CD14-APC (clone M5E2, BD Pharmingen). Cell viability was assessed with DAPI (Life Technologies). Cells were assayed on a BD Fortessa or BD Symphony A5 SE and data were analyzed with FlowJo software (Tree Star).

Immunofluorescence staining and imaging

iMGLs were cultured on round coverslips inside 24-well plates. On the day of maturation, cells were washed 3 times with 1X HEPES solution followed by fixation for 10 minutes at room temperature with 4% (wt/vol) paraformaldehyde (PFA). Following fixation, cultures were washed 3 times with DPBS. Cells were then permeabilized and blocked in 0.3% triton X-100 and 5% normal goat serum in DPBS. Incubation with primary antibodies anti-CX3CR1 (NC0789871, Fisher Scientific), anti-TREM2 (MAB17291, R&D), anti-IBA1 (019-19741, Wako Pure Chemical Corporation), anti-P2PY12 (HPA014518, MilliporeSigma) was performed overnight at 4ºC, followed by incubation for 2 hours with secondary antibodies Alexa Fluor 488 Donkey Anti-Rabbit IgG (711-545-152, Jackson Immuno) and Goat anti-Rat IgG Secondary Antibody, Alexa Fluor 488 (A-11006, ThermoFisher) at room temperature. Imaging was done with a Zeiss LSM 780 microscope and quantification of immunofluorescence was performed with ImageJ.

scRNAseq of iMGLs

Single-cell RNA-sequencing (scRNAseq) was performed on freshly dissociated iMGLs with the Chromium 10x Genomics 3' gene expression (3' GEX) V3 protocol, using an input of ~10,000 cells. Briefly, Gel-Bead in Emulsions (GEMs) were generated on the sample chip in the Chromium controller. Barcoded cDNA was extracted from the GEMs by Post-GEM RT-cleanup and amplified for 12 cycles. Amplified cDNA was fragmented and subjected to end-repair, poly A-tailing, adapter ligation, and 10X-specific sample indexing following the manufacturer's protocol. Libraries were quantified using Bioanalyzer (Agilent) and QuBit (Thermofisher) analysis and were sequenced in paired-end mode on NovaSeq targeting a depth of 50,000-100,000 reads per cell. Sequencing data were aligned and quantified using Cell Ranger (v7.1) against the GRCh38 human reference genome.

Creation and analysis of integrated scRNAseq atlas of iMGLs

Analysis of scRNAseq data was performed using Seurat (v4.4.1). High-quality cells were retained based on meeting the following quality-control metrics: 1) percent mitochondrial gene expression < 20% and 2) 2000 < nFeature_RNA < 7000. Cells from all lines (3 mutant and 1 WT) were merged into a single Seurat object and underwent log-normalization and regression of the technical covariates percent.mito, nCount_RNA, and nFeature_RNA. Integration across all four iMGL genotypes was performed using Harmony (v1.2.0). Unsupervised clusters were identified using the Louvain algorithm and previously defined microglial state markers[41] were annotated based on canonical state marker genes.

For each mutant line, the enrichment of each microglial state was assessed by performing a hypergeometric test of the overlap between iMGLs from a given mutant line and iMGLs of a given microglial state, as compared to the overlap between WT iMGLs and iMGLs of a given microglial state. Differential expression analysis was performed using Seurat's FindMarkers function (min.pct = 0.10, no log-fold change cutoff) between each mutant iMGL line individually and all combined, versus the WT iMGL line. Gene set enrichment analysis was performed using clusterProfiler (v4.4.4) and the 50 Hallmark Molecular Signature Database pathways. Significance for all scRNAseq analysis was defined as Benjamini-Hochberg adjusted p-value < 0.05.

QUANTIFICATION AND STATISTICAL ANALYSIS

All of the quantitative and statistical methods, strategies, and analyses are described in the relevant sections of the method details or in the table and figure legends.

Supplemental figures

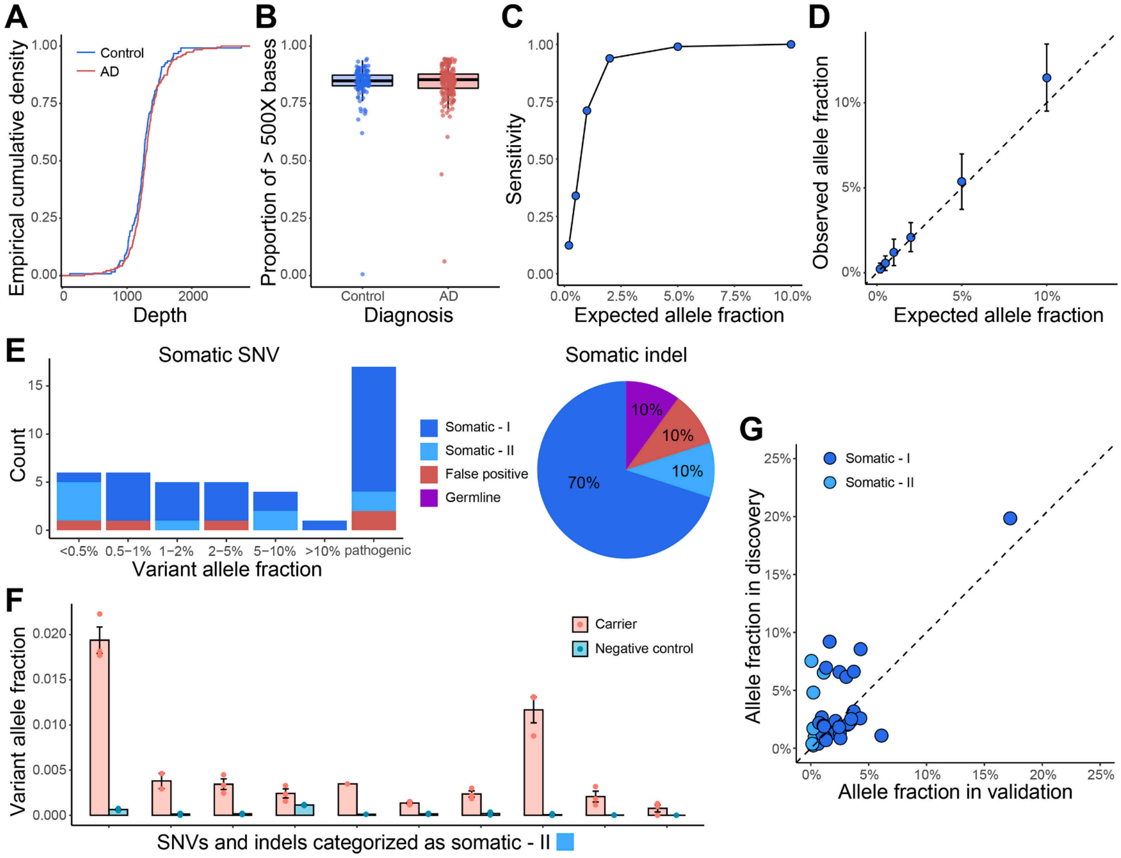


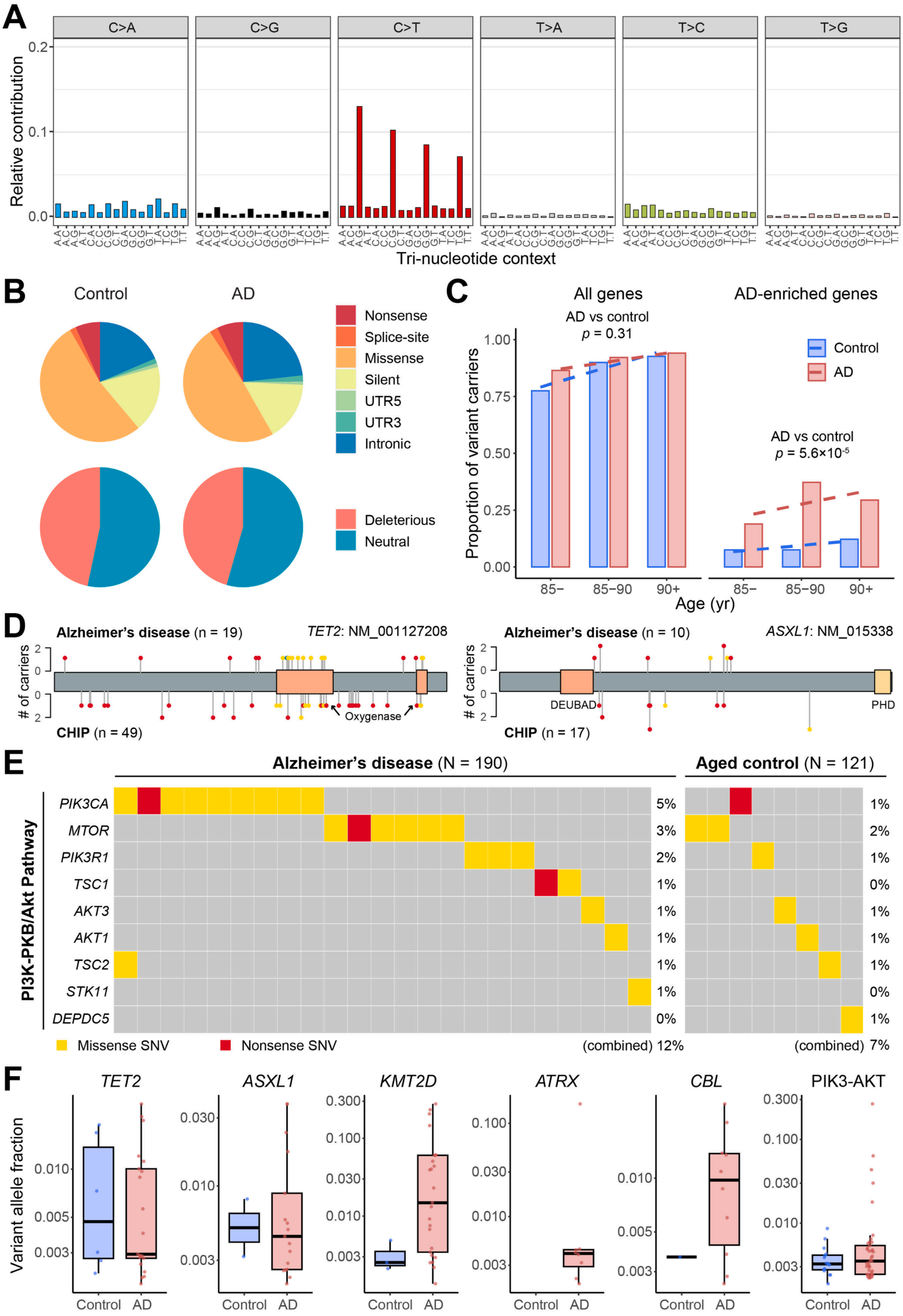
Figure S1. Benchmarking and validation results of sSNVs and sIndels identified from panel sequencing, related to Figure 2

(A and B) Comparable sequencing depth (A) and coverage (B) between AD and control PFC samples, calculated based on consensus reads after UMI-based read collapsing.

(C and D), Detection sensitivity (C) and accuracy of allele fraction estimation (D) for our panel sequencing and somatic variant identification pipeline, benchmarked by an *in vitro* mixture of the DNA samples of two unrelated individuals with varied genome ratios. Error bar, SD.

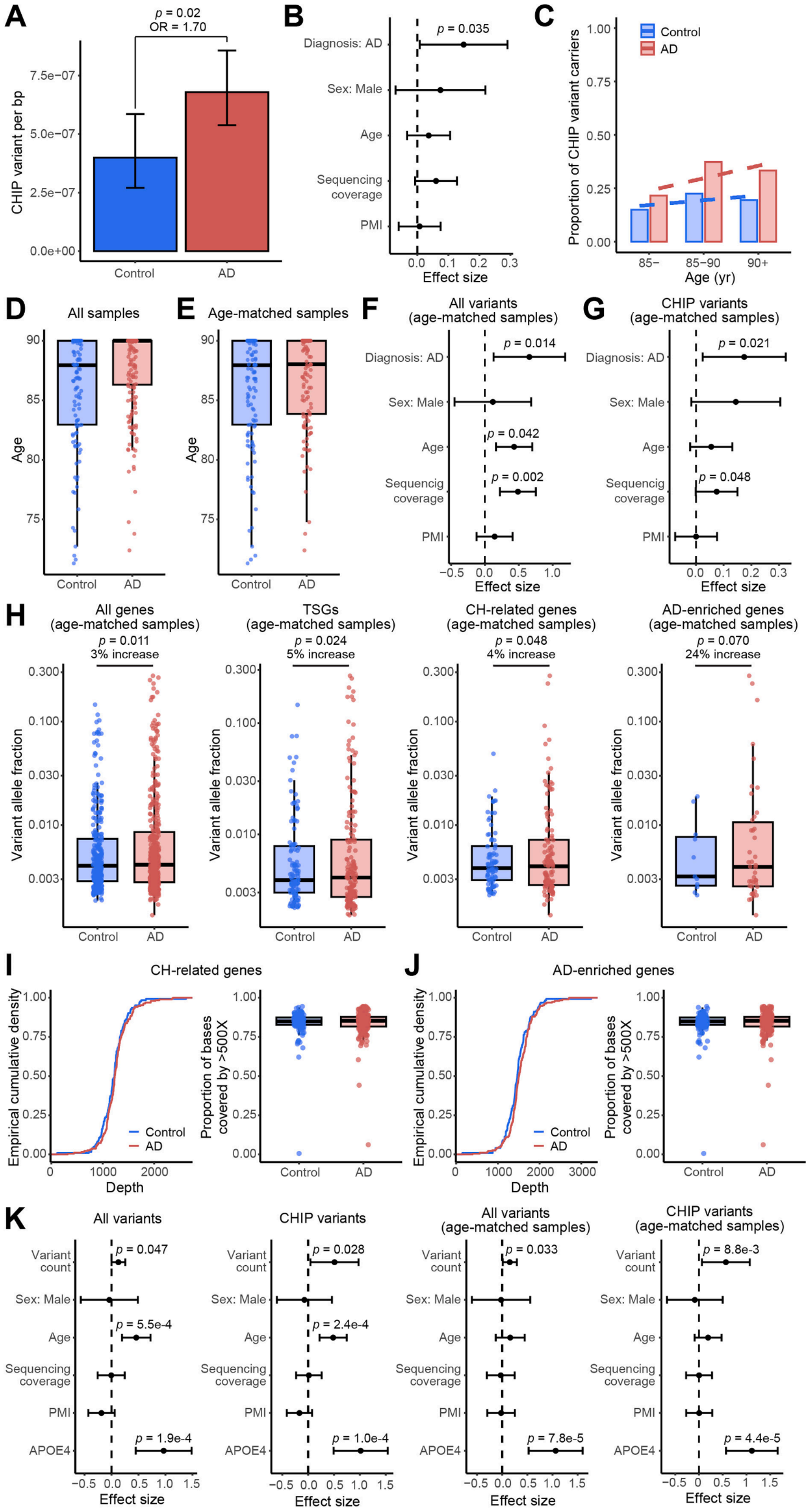
(E and F) Amplicon sequencing validation confirmed high accuracy for identified sSNVs and sIndels in AD and control samples (E). Somatic-I variants are those with VAFs at least 3× larger than the fractions of the other two error alleles of the same genomic position, whereas somatic-II are those that were further validated by comparing their VAFs in a negative control sample (F). Error bar, SE.

(G) Variant allele fraction of validated somatic variants between panel sequencing (discovery) and amplicon sequencing (validation). Amplicon sequencing was performed using newly extracted DNA from the corresponding brain sample; therefore, the allele fractions may vary between the discovery and validation stages.



(legend on next page)

Figure S2. Identification and functional annotation of sSNVs in panel sequencing data, related to Figure 2
(A) Variant type and tri-nucleotide context of sSNVs.
(B) Genic annotation and functional impact prediction of sSNVs identified from AD and control PFC samples.
(C) The proportion of somatic variant carriers increases with age. AD patients had a significantly larger proportion of carriers with somatic variants in AD-enriched genes than matched controls ($p = 5.6 \times 10^{-5}$, linear regression).
(D) Similar distributions between somatic variants identified in AD brains and previously reported CH-related variants in blood.
(E) Genes in the PI3K-PKB/Akt pathway contained significantly more somatic variants in AD brains (12% of AD samples versus 7% of control samples; $p < 0.05$, one-tailed proportion test).
(F) AD brains generally showed a trend toward carrying more somatic variants with higher VAF (e.g., VAF > 0.01) compared with controls, for each AD-enriched gene, as well as for genes in the PI3K-PKB/Akt pathway, despite inter- and intra-gene variability.



(legend on next page)

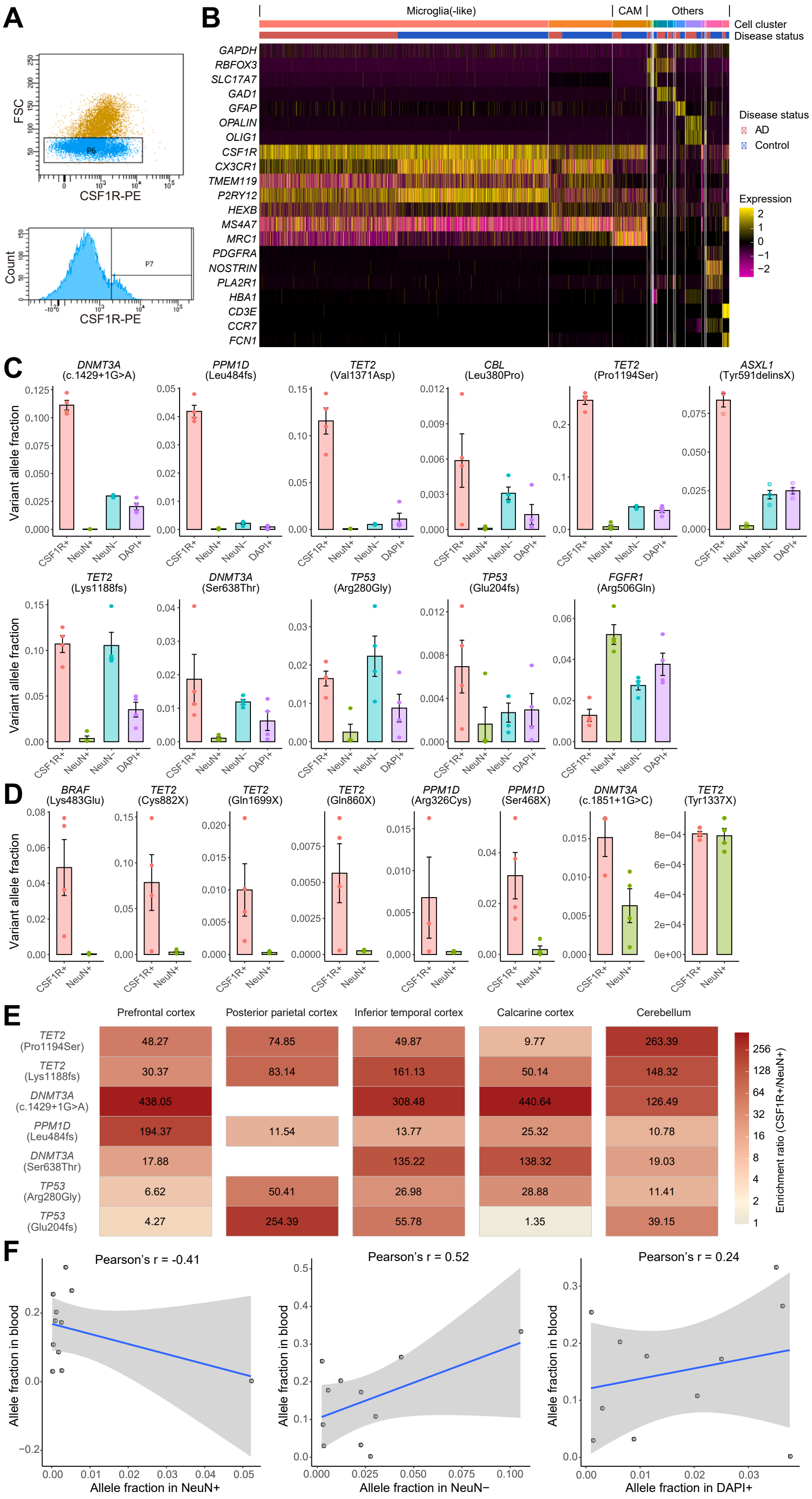
Figure S3. Confirmation of the higher burden of somatic CHIP variants in AD brains, related to Figure 2

(A–C) Elevated burden of somatic CHIP variants in AD brains, shown as number of variants per bp (A), linear regression model (B), and proportion of variant carriers (C). CHIP variants were defined based on a curated list of CHIP variants previously identified from blood WGS datasets.[25,37]

(D–H) Robust results after age matching in the panel sequencing of ROSMAP samples. Age distribution of ROSMAP PFC samples before (D) and after (E) age matching. A subset of AD samples (121 out of 170) shows a similar age distribution to control samples ($p = 0.35$, two-tailed Wilcoxon test). Linear regression models with age-matched samples confirm a significant increase in AD after controlling for other potential confounding factors, whether considering all variants (F) or CHIP variants only (G). Higher VAFs were observed in AD compared with age-matched samples for somatic variants across all genes, tumor suppressor genes (TSGs), CH-related genes, and AD-enriched genes (H).

(I and J) Comparable sequencing depth and coverage between AD and control samples across CH-related (I) or AD-enriched genes (J). Depth and coverage were calculated based on consensus reads after UMI-based read collapsing.

(K) Logistic regression models for AD diagnosis as a binary outcome. The *APOE* genotype was modeled additively for the APOE4 allele. We confirmed a consistently significant positive association between somatic variant burden and AD risk, whether considering all variants or CHIP variants only, both before and after age matching. (B) and (K), PMI, post-mortem interval.



	Prefrontal cortex	Posterior parietal cortex	Inferior temporal cortex	Calcarine cortex	Cerebellum
TET2 (Pro1194Ser)	48.27	74.85	49.87	9.77	263.39
TET2 (Lys1188fs)	30.37	83.14	161.13	50.14	148.32
DNMT3A (c.1429+1G>A)	438.05		308.48	440.64	126.49
PPM1D (Leu484fs)	194.37	11.54	13.77	25.32	10.78
DNMT3A (Ser638Thr)	17.88		135.22	138.32	19.03
TP53 (Arg280Gly)	6.62	50.41	26.98	28.88	11.41
TP53 (Glu204fs)	4.27	254.39	55.78	1.35	39.15

(legend on next page)

Figure S4. Cell-type composition and variant allele fraction of FANS-sorted nuclei populations, related to Figure 3

(A) Selectively isolated microglia-enriched brain macrophages from frozen brain tissues using FANS with an antibody targeting epitopes of CSF1R.

(B) Marker gene expression profile for 10× snRNAseq of CSF1R+ sorted nuclei. Each column represents a single nucleus, clustered by PCA based on its expression similarity. About 75%–77% of the sorted nuclei transcriptionally resemble microglia with high expression of *CX3CR1*, *TMEM119*, and *P2RY12*, whereas another 4%–9% are CNS-associated macrophages (CAMs). Markers for blood cell types (*HBA1*: red blood cell; *CD3E*: T cell; *CCR7*: B cell; *FCN1*: monocyte) confirm the minimal presence of blood cells in sorted nuclei. CNS, central nervous system. AD brain macrophages showed generally reduced expression of *CX3CR1* and *P2RY12*, consistent with previous findings in AD.[7]

(C and D) VAFs across different sorted nuclei populations for all 18 profiled somatic variants in AD (C) and control (D) PFC samples. Four AD variants are shown in Figure 3D as examples. Except for the non-driver *FGFR1* variant, we observed higher allele fractions in brain macrophages (CSF1R+) than in neurons (NeuN+) in 16 of the 17 potentially driver variants. Each population of nuclei was typically sorted four times from each brain sample to serve as replicates. Error bar, SE.

(E) The enrichment ratio of profiled driver variants between CSF1R+ and NeuN+ nuclei across multiple brain regions of AD individuals.

(F) The correlation of VAFs between blood and three nuclei populations (NeuN+, NeuN-, and DAPI+) sorted from matched AD PFC samples.

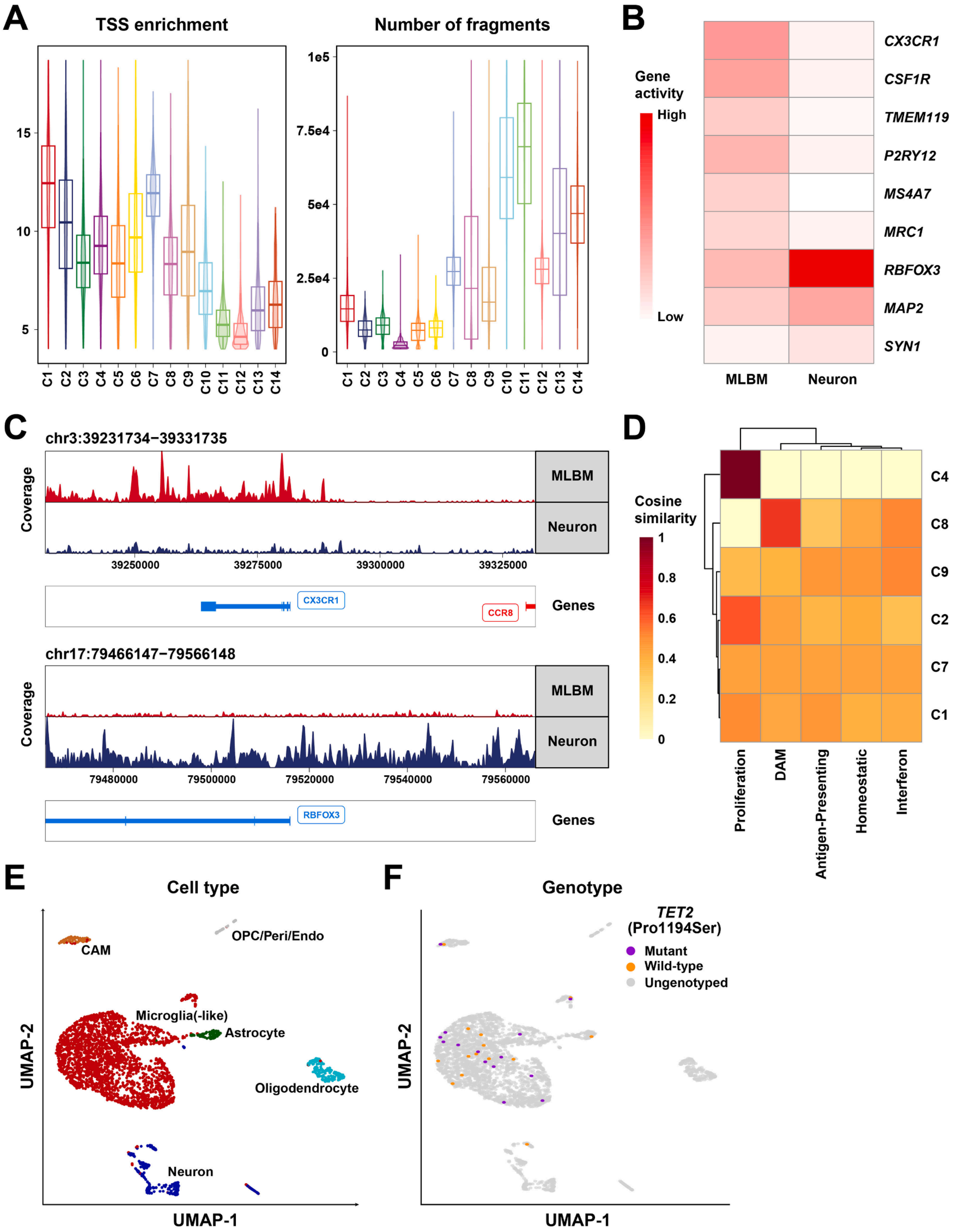


Figure S5. Multi-omic analysis of single-nucleus sequencing data from brain samples with somatic driver variants, related to Figure 4

(A) Quality-control metrics of transcription start site (TSS) enrichment and number of fragments per nucleus per unsupervised cluster.

(B and C) Gene activity (B) and chromatin accessibility (C) of canonical microglia and neuron marker genes in designated MLBM and neuron clusters, respectively.

(D) Cosine similarity of gene activity between known microglial state modules[41] and unsupervised clusters. Clusters C1, C2, and C4 most closely resemble the proliferation state; C8 most closely resembles the DAM state; and C9 most closely resembles the interferon state.

(E and F) Simultaneous transcriptomic (E) and genotyping (F) analysis of snRNAseq data from CSF1R+ sorted brain sample of an AD patient with a deleterious missense variant in *TET2* (p.Pro1194Ser). Mutant alleles are observed in both microglia(-like cells) and CAMs, but not in other brain cell types. Peri, pericyte. Endo, endothelial cell.

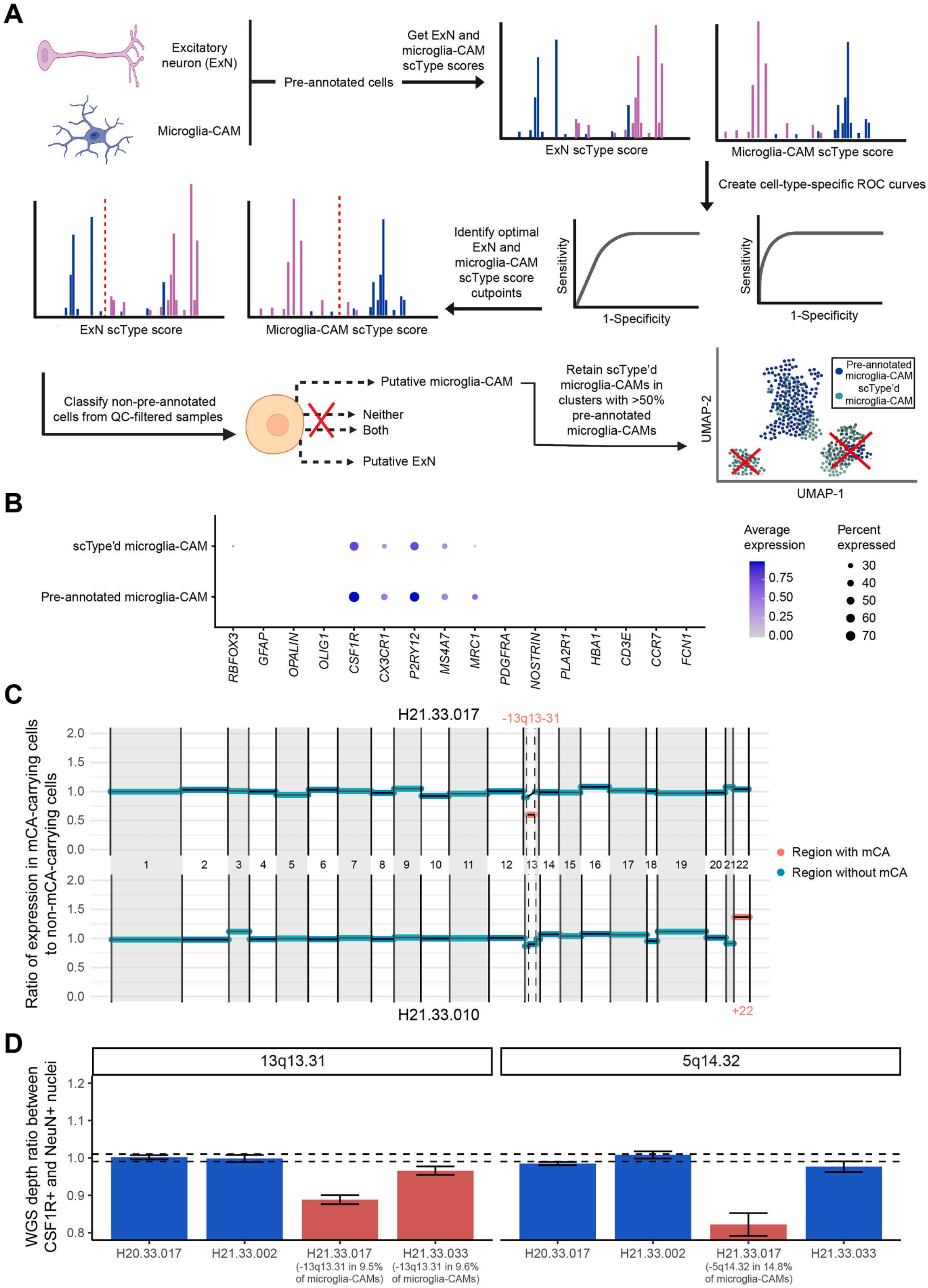


Figure S6. Identification and validation of mCA in the SEA-AD cohort, related to Figure 5

(A) Schematic representation of the supervised learning framework and quality-control metrics used to detect additional high-quality microglia-CAMs from SEA-AD.

(B) scType'd, and pre-annotated microglia-CAMs show similar marker gene expression profiles, with specific expression of microglia and CAM marker genes.

(legend continued on next page)

(C) Examples of mCA called in two AD individuals, H21.33.017 (chr13p13-31 deletion) and H21.33.010 (chr22 amplification). Normalized median ratio of expression in mCA-carrying cells versus non-carrying cells displayed per chromosomal region, with chromosome size proportional to the number of expressed genes in microglia-CAMs from that chromosome.

(D) Validation of mCA by whole-genome sequencing (WGS) of sorted microglia-CAMs (CSF1R+) and neurons (NeuN+) from four individuals. Individuals with called mCA deletion events in CSF1R+ nuclei (red) exhibit reduced WGS depth within predicted chromosomal regions, in contrast to negative control individuals (blue). Dotted lines denote 0.99 and 1.01. Error bar, SE.

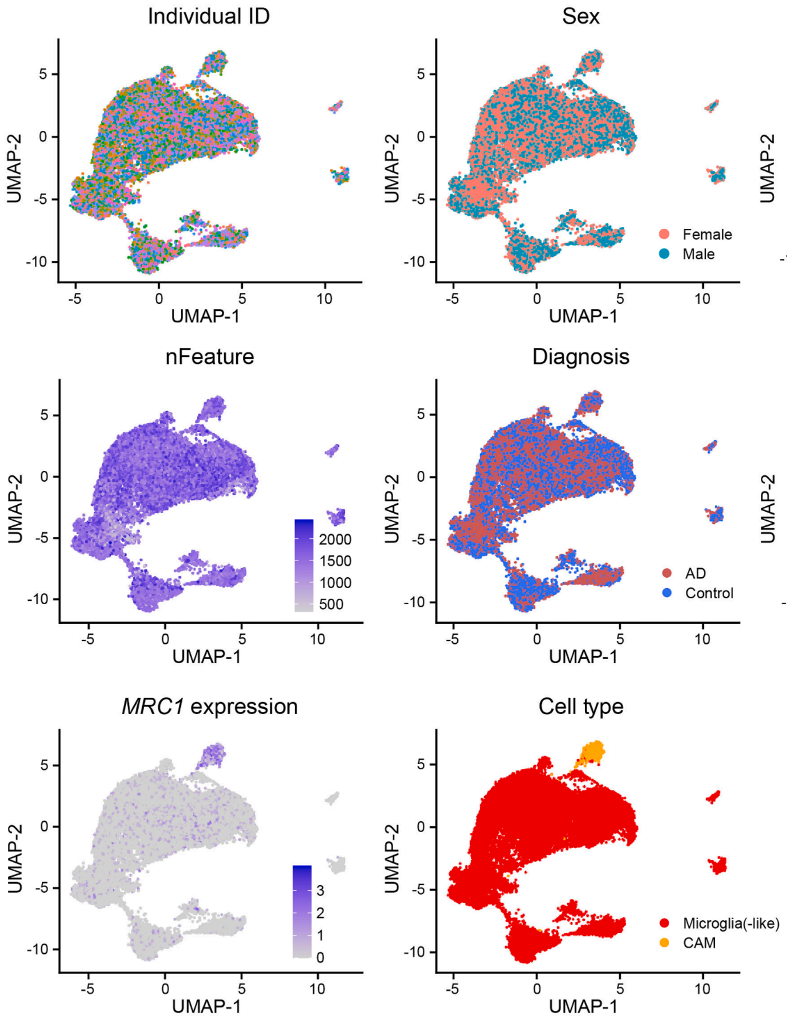


Figure S7. Integrated snRNAseq atlas of microglia-CAMs in AD and healthy controls, related to Figure 5

UMAP visualization of covariates of interest does not reveal significant clustering by individual ID, nFeature, or nCount, consistent with successful integration across samples. Microglia(-like cells) and CAMs (with high *MRC1* expression) separate into distinct clusters.

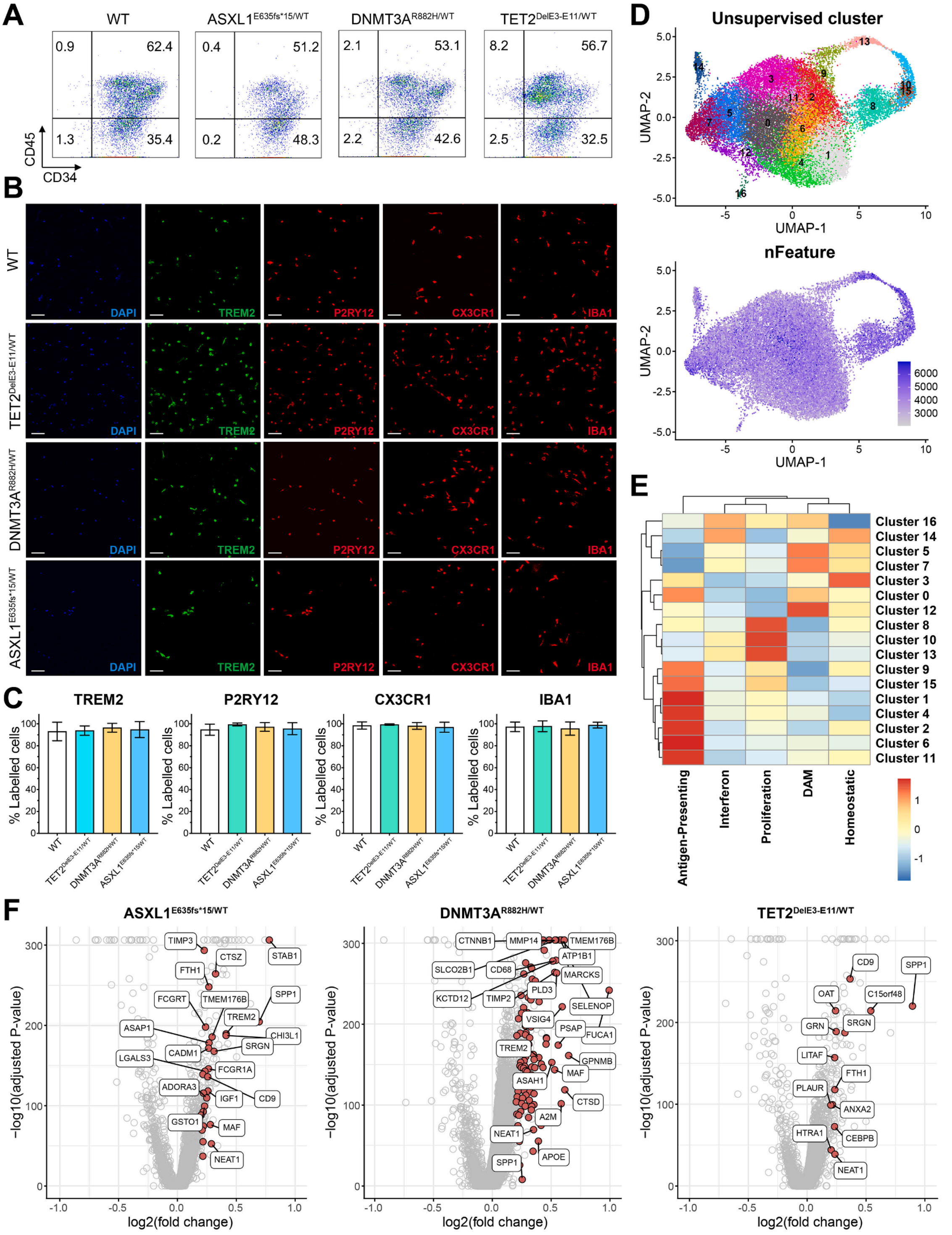


Figure S8. Quality control and annotation of iMGLs, related to Figure 6

(A) Flow cytometry assessment of HSPCs derived from induced pluripotent stem cells (iPSCs). iPSC lines of the indicated genotypes on day 11 of differentiation.

(B) Representative images of iMGLs derived from iPSC lines of the indicated genotypes immunolabeled for the indicated microglia markers on day 53 of differentiation. Scale bars, 50 μm.

(C) Quantification of the immunofluorescence data from (B).

(D) UMAP embedding of unsupervised clustering and nFeature for integrated scRNAseq atlas of iMGLs from the three mutant lines and the isogenic wild-type (WT) lines.

(legend continued on next page)

(E) Microglial state marker gene scaled expression used for state annotation in Figure 6C.

(F) Volcano plots of DEGs between iMGL from each mutant line and WT iMGLs. Genes that are upregulated in mutant iMGLs include many canonical DAM genes (highlighted in red), many of which are shared across mutant lines.

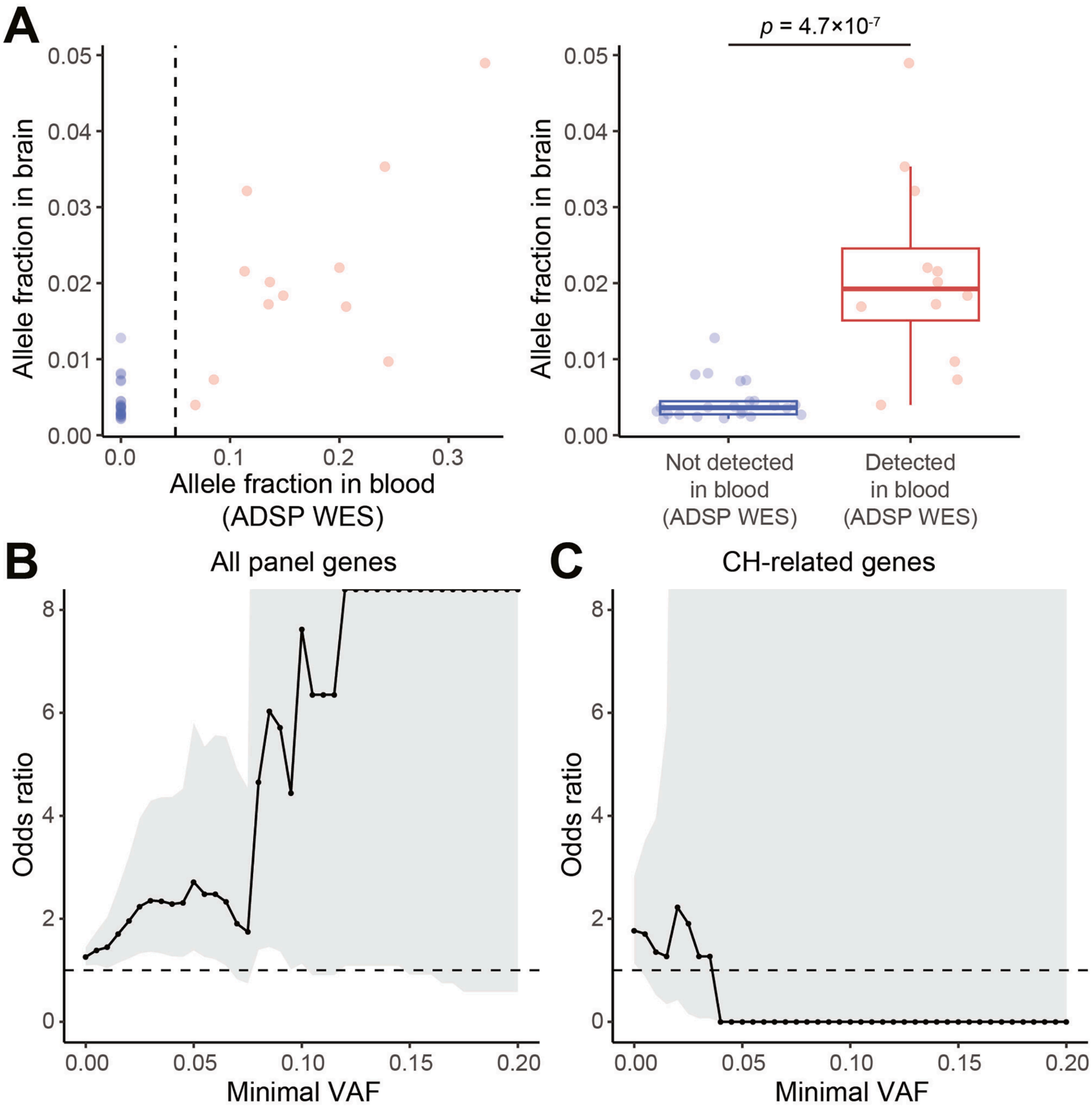


Figure S9. Low-fraction CHIP variants contribute predominantly to AD risk but are missed in the WES of blood samples

(A) Only somatic variants with high variant allele fractions (VAFs) in brain samples can be called by the pipeline of Bouzid et al.[25] from the whole-exome sequencing (WES) of matched blood samples generated by ADSP. The ADSP WES data can only detect somatic variants with >5% VAF, due to its limited sequencing depth.

(B and C) The OR of AD enrichment for sSNVs with different VAF cutoffs. When we consider all 149 genes targeted by the panel sequencing, we observe a consistent trend of AD enrichment even for sSNVs with 5% or more VAF (B). In comparison, when we only consider deleterious sSNVs in CH-related genes, the OR becomes smaller than 1 when VAF is larger than 4%, though with a very large confidence interval (C). The dashed line represents the OR of 1, and odds ratios larger and smaller than 1 denote the enrichment and depletion of sSNV in AD, respectively.