A microglia clonal inflammatory disorder in Alzheimer's disease

Rocio Vicario[1]*[†], Stamatina Fragkogianni[1†], Leslie Weber[1], Tomi Lazarov[1], Yang Hu[2], Samantha Y Hayashi[3], Barbara Craddock[3], Nicholas D Socci[4], Araitz Alberdi[1], Ann Baako[1], Oyku Ay[1], Masato Ogishi[5], Estibaliz Lopez-Rodrigo[1], Rajya Kappagantula[6], Agnes Viale[4], Christine A Iacobuzio-Donahue[6,7], Ting Zhou[8], Richard M Ransohoff[9], Richard Chesworth[9], Netherlands Brain Bank[10], Omar Abdel-Wahab[6], Bertrand Boisson[5], Olivier Elemento[2], Jean-Laurent Casanova[5], W Todd Miller[3], Frédéric Geissmann[1]*

[1]Immunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York, United States; [2]Department of Physiology and Biophysics, Institute for Computational Biomedicine, Weill Cornell New York, New York, United States; [3]Department of Physiology and Biophysics, Stony Brook University School of Medicine, Stony Brook, New York, United States; [4]Marie-Josée & Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, United States; [5]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, New York, United States; [6]Human Oncology & Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York, United States; [7]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York, United States; [8]SKI Stem Cell Research Core, Memorial Sloan Kettering Cancer Center, New York, New York, United States; [9]Third Rock Ventures, Boston, United States; [10]Netherlands Brain Bank, Amsterdam, Netherlands

*For correspondence: rociovicario@gmail.com (RV); rociovicario@gmail.com (RV); geissmaf@mskcc.org (FG)

[†]These authors contributed equally to this work

eLife assessment

This **fundamental** study enhances our understanding of how somatic variants in microglia might influence the onset and progression of neurodegenerative diseases such as Alzheimer's. The evidence supporting the conclusions is **compelling**, with the authors employing a multi-faceted approach to identify an enrichment of potentially pathogenic somatic mutations in Alzheimer's disease microglia. This research will be of significant interest to those investigating somatic mutations, Alzheimer's disease, microglial biology and cell signalling pathways.

Abstract Somatic genetic heterogeneity resulting from post-zygotic DNA mutations is widespread in human tissues and can cause diseases, however, few studies have investigated its role in neurodegenerative processes such as Alzheimer's disease (AD). Here, we report the selective enrichment of microglia clones carrying pathogenic variants, that are not present in neuronal, glia/stromal cells, or blood, from patients with AD in comparison to age-matched controls. Notably, microglia-specific AD-associated variants preferentially target the MAPK pathway, including recurrent CBL ring-domain mutations. These variants activate ERK and drive a microglia transcriptional program characterized by a strong neuro-inflammatory response, both in vitro and in patients. Although the natural history of AD-associated microglial clones is difficult to establish in humans, microglial expression of a MAPK pathway activating variant was previously shown to cause neurodegeneration

in mice, suggesting that AD-associated neuroinflammatory microglial clones may contribute to the neurodegenerative process in patients.

Introduction

Neurodegenerative diseases are a frequent cause of progressive dementia. AD is diagnosed in ~90% of cases, with an estimated prevalence of ~10% in the population over 65 y of age (*Hebert et al., 2013*; *Alzheimer's Association, 2019*). The role of germline genetic variation in neurodegenerative diseases and AD has been studied intensely. Although autosomal dominant forms of AD due to rare germline variants with high penetrance account only for an estimated ~1% of cases (*Lanoiselée et al., 2017*; *Goate et al., 1991*; *Chartier-Harlin et al., 1991*; *Levy-Lahad et al., 1995a*; *Levy-Lahad et al., 1995b*; *Rogaev et al., 1995*), a number of common variants were also shown to contribute to disease risk. Carriers of one germline copy of the epsilon4 (E4) allele of the apolipoprotein E gene (APOE4), present in ~15 to 20% of the population, have a threefold higher risk of AD, while two copies (~2 to 3% of the population) increase the risk by ~10 fold (*Saunders et al., 1993*; *Murrell et al., 2006*; *Sando et al., 2008*; *Lumsden et al., 2020*). Genome-wide association studies (GWAS) have identified an additional ~50 common germline variants that more moderately increase the risk of AD, including TREM2, CD33, and MS4A6A variants (*Jonsson et al., 2013*; *Guerreiro et al., 2013*; *McQuade and Blurton-Jones, 2019*). Interestingly, the APOE4 allele is responsible for an increased inflammatory and neurotoxic response of microglia and astrocytes in the brain of carriers (*Arnaud et al., 2022*; *Serrano-Pozo et al., 2021*; *Rodriguez et al., 2014*), and it was noted that the majority of the other germline AD-risk variants are located within or near genes expressed in microglia *McQuade and Blurton-Jones, 2019* and in particular at microglia-specific enhancers (*Nott et al., 2019*). These data, together with transcriptional studies (*Krasemann et al., 2017*; *Mathys et al., 2019*; *Keren-Shaul et al., 2017*) support the hypothesis that genetic variation in microglia may contribute to the pathogenesis of neurodegeneration and AD.

Somatic genetic heterogeneity (mosaicism), resulting from post-zygotic DNA mutations, is widespread in human tissues, and a cause of tumoral, developmental, and immune diseases (*Miller et al., 2021*; *Martincorena et al., 2015*; *Martincorena and Campbell, 2015*; *Behjati et al., 2014*). Additionally, the role of somatic variants in neuropsychiatric disorders is also suspected (*McConnell et al., 2017*). Mosaicism has been documented in the brain tissue of AD patients in several deep-sequencing studies (*Keogh et al., 2018*; *Wei et al., 2019*; *Park et al., 2019*), showing that the enrichment of putative pathogenic somatic mutations in the PI3K-AKT, MAPK, and AMPK pathway do occur in the brain of patients in comparison to controls (*Park et al., 2019*). However, these studies performed in whole brain tissue lacked cellular resolution and mechanistic insights, and the role of somatic mutants in neurodegenerative diseases remains poorly understood (*Miller et al., 2021*). Somatic variants that activate the PI3K-AKT-mTOR or MAPK pathways in neural progenitors are a cause of cortical dysplasia and epilepsy *D'Gama et al., 2017*; *Khoshkhoo et al., 2023*; *Koh et al., 2018*; *Lim et al., 2015* and developmental brain malformations (*Poduri et al., 2012*), while somatic variants that activate the MAPK pathway in brain endothelial cells are associated with arteriovenous malformations (*Nikolaev et al., 2018*). Interestingly, we reported that expression of a somatic variant activating the MAPK pathway in microglia causes neurodegeneration in mice (*Mass et al., 2017*), but the presence and contribution of microglial somatic clones in neurodegenerative diseases and AD remains unknown.

Here, we investigated the presence and nature of somatic variants in brain cells from control and AD patients. In an attempt to examine all brain cells at the same resolution, nuclei from neurons, glia cells, and microglia, which only represent ~5% of brain cells, were pre-sorted. Human microglia are reported to develop in embryos and renew by local proliferation within the brain (*Askew et al., 2017*; *Réu et al., 2017*; *Bian et al., 2020*). However, bone marrow-derived myeloid cells can enter the brain, in particular during pathological processes, and may not be distinguishable from resident microglia by transcriptomics alone *Kim, 2023*. In order to distinguish somatic variants carried by resident microglia from the ones carried by myeloid cells of peripheral origin, we analyzed matched peripheral blood from control and patients, to 'barcode' somatic mutants shared between microglia and blood. Finally, in order to achieve high sensitivity in the detection of variants that confer a proliferative or activation advantage (pathogenic mutations) and support the emergence or pathogenicity of mosaic clones (*Frank, 2010*), and/or that have been previously associated with neurological diseases, we initially

eLife digest Around 10% of people aged over 65 are estimated to have Alzheimer's disease. This progressive neurodegenerative condition leads to death of brain cells, memory loss, confusion and other life-altering symptoms.

Somatic mutations are changes in the genetic information of a cell other than sperm or eggs, which can result in alterations in gene function. As the mutant cells multiply, they form clones that also carry these changes – potentially resulting in groups of cells that behave differently from those in which those mutations are absent. Despite their importance, the role of somatic mutations in Alzheimer's disease remains poorly understood.

To investigate this question, Vicario, Fragkogianni, Weber, Lazarov et al. examined the genetic material of brain and blood cells obtained from individuals who had died either of Alzheimer's disease, or of other causes. The team focused their analysis on around 700 genes previously associated with neurodegenerative conditions. The results showed that, compared to individuals whose death was not due to neurological illnesses, harmful variants of those genes were present in higher numbers in the microglia cells of around 25% of Alzheimer's patients in their series. No such increase was detected in other blood or brain cell populations, regardless of the individuals' cause of death.

Microglia are cells tasked with helping to repair damage and fight off infections in the brain. Many of the harmful gene variants found in this population switched on a cell pathway known as the MAP Kinase pathway, which activated the cells and caused them to multiply. This, in turn, led to inflammation and may contribute to the death of neurons.

Together these findings indicate that developing a new class of therapeutics that inhibits the MAP Kinase pathway in microglia may help prevent irreversible brain damage in some patients with Alzheimer's disease.

performed a targeted deep-sequencing of a panel of 716 genes covering somatic variants reported in clonal proliferative disorders and genes associated with neurodegenerative diseases diseases.

We found that microglia from AD patients were enriched for pathogenic variants in comparison to age-matched controls. Furthermore, we found that these microglia-specific AD-associated variants preferentially target the MAPK pathway, including recurrent CBL ring-domain mutations. In addition, we showed that these variants drive a microglia transcriptional program characterized by a strong neuro-inflammatory response previously associated with neurotoxicity, including the production of IL1 and TNF, both in in vitro microglia models and in patients. The natural history of the AD-associated microglia clonal inflammatory disorder we describe here is difficult to establish. Specifically, we do not know whether it contributes to the onset of the neuro-inflammatory process at an early stage of the disease, or if microglia carrying pathogenic mutations preferentially expand later during the course of the disease in response to tissue inflammation. Under both hypotheses, however, the presence of neuro-inflammatory microglial clones may contribute to the neurodegenerative process in a subset of AD patients. This report reveals a previously unrecognized presence of AD-associated microglia harboring pathogenic somatic variants in humans and provides mechanistic insight for neurodegenerative diseases by delineating cell-type specific variant recurrence.

Results

Clonal diversity among brain cells and blood from controls and AD patients

We examined post-mortem frozen brain samples and matching blood from 45 patients with intermediate-onset sporadic AD and 44 control individuals who died of other causes, including 27 donors age and sex-matched donors with the AD cohort (*Figure 1A*; *Figure 1—figure supplement 1A* and *Supplementary file 1*). APOE risk allele frequency for patients and controls was comparable to published series (*Murrell et al., 2006*; *Sando et al., 2008*; *Lumsden et al., 2020*; *Figure 1—figure supplement 1A*), and analysis of germline mutations did not identify deleterious variants in the 140 genes associated with neurological diseases. Myeloid/microglia, neurons, and glia/stromal cells were purified by flow cytometry using antibodies against PU.1 and NeuN (*Evrony et al., 2012*; *Figure 1B*,
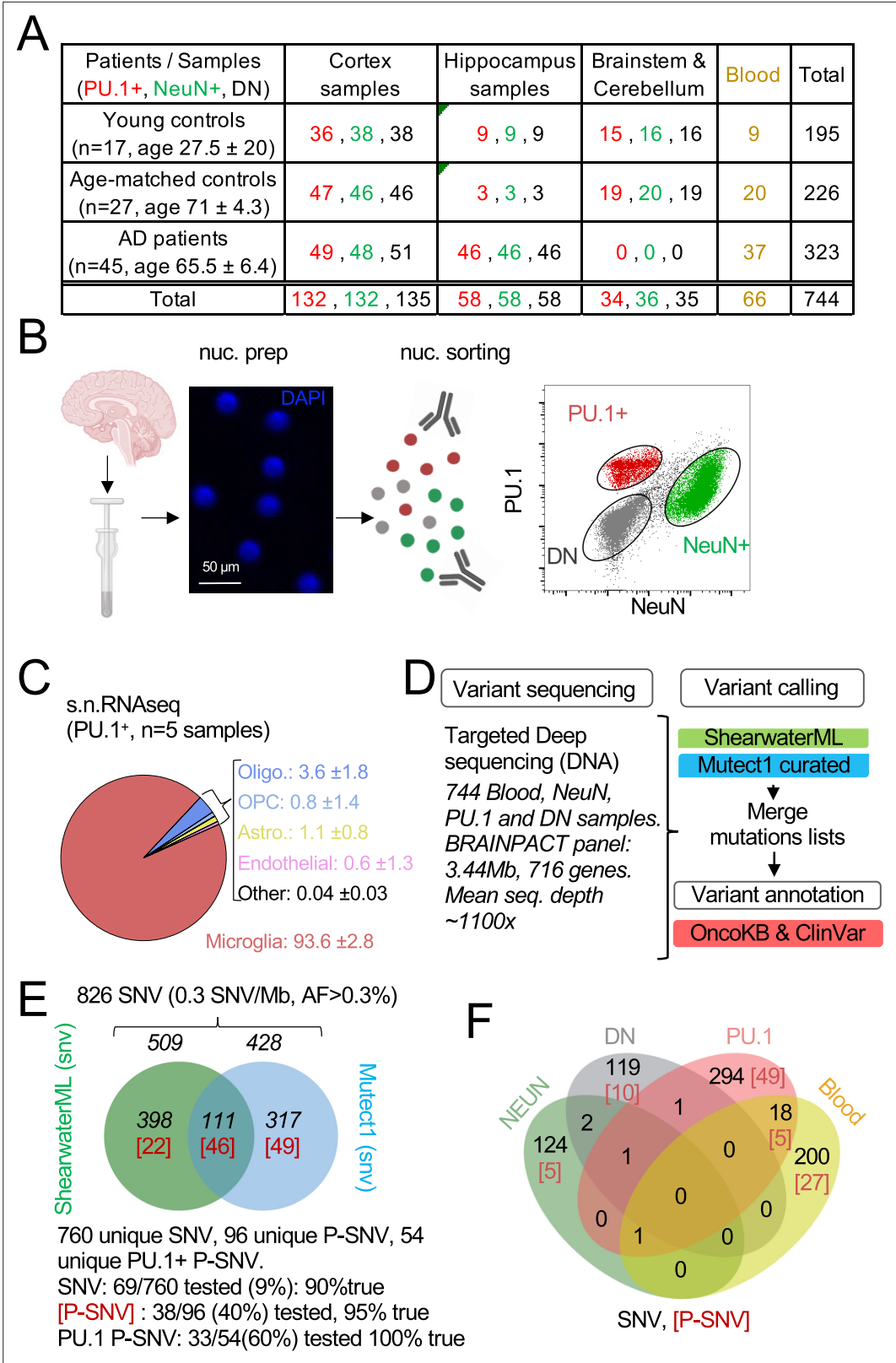
A

Patients / Samples (PU.1+, NeuN+, DN)	Cortex samples	Hippocampus samples	Brainstem & Cerebellum	Blood	Total
Young controls (n=17, age 27.5 ± 20)	36 , 38 , 38	9 , 9 , 9	15 , 16 , 16	9	195
Age-matched controls (n=27, age 71 ± 4.3)	47 , 46 , 46	3 , 3 , 3	19 , 20 , 19	20	226
AD patients (n=45, age 65.5 ± 6.4)	49 , 48 , 51	46 , 46 , 46	0 , 0 , 0	37	323
Total	132 , 132 , 135	58 , 58 , 58	34 , 36 , 35	66	744

B

nuc. prep — DAPI — 50 µm

nuc. sorting

PU.1+ / NeuN+ / DN

PU.1 vs NeuN

C

s.n.RNAseq (PU.1+, n=5 samples)

Oligo.: 3.6 ±1.8
OPC: 0.8 ±1.4
Astro.: 1.1 ±0.8
Endothelial: 0.6 ±1.3
Other: 0.04 ±0.03
Microglia: 93.6 ±2.8

D

Variant sequencing | Variant calling

Targeted Deep sequencing (DNA)
744 Blood, NeuN, PU.1 and DN samples. BRAINPACT panel: 3.44Mb, 716 genes. Mean seq. depth ~1100x

ShearwaterML
Mutect1 curated
↓
Merge mutations lists
↓
Variant annotation
OncoKB & ClinVar

E

826 SNV (0.3 SNV/Mb, AF>0.3%)

509   428

ShearwaterML (snv) / Mutect1 (snv)

398 [22]   111 [46]   317 [49]

760 unique SNV, 96 unique P-SNV, 54 unique PU.1+ P-SNV.
SNV: 69/760 tested (9%): 90%true
[P-SNV] : 38/96 (40%) tested, 95% true
PU.1 P-SNV: 33/54(60%) tested 100% true

F

DN   PU.1
119 [10]   294 [49]

NEUN   Blood
18 [5]

124 [5]   2   1   200 [27]

1   0

1   0

0   0

0   0

1

0

SNV, [P-SNV]

Figure 1. Detection of mutations in brain cell types and blood. (**A**) Table with patient and sample information. (**B**) Schematic represents the isolation and labeling of nuclei from post-mortem frozen brain samples from controls and Alzheimer's disease patients with DAPI and antibodies against PU.1+ (myeloid/microglia) and NeuN+ (neurons). Representative flow cytometry dot-plot of nuclei separation. Double negative nuclei are labeled 'DN.'

Figure 1 continued on next page

Figure 1 continued

(**C**) Percentage of cell types obtained in sorted PU.1$^+$ nuclei determined by single-nuclei RNAseq in five brain samples from four individuals. (**D**) Schematic represents the sequencing strategy. Two algorithms (ShearwaterML and Mutect1) were used for variant calling. After annotation, pathogenicity was determined using OncoKb and ClinVar. (**E**) Venn diagram represents the number of variants and overlap between the ShearwaterML and Mutect1. Numbers in red indicate pathogenic variants (P-SNV). Validation of variants was performed by droplet digital (dd) PCR on pre-amplified DNA when available. (**F**) Venn diagrams represent the repartition per cell type of the 826 single-nucleotide variations (SNVs) identified in NeuN$^+$: Neurons, PU.1$^+$: microglia, DN: glia, and matching blood. [Numbers] in red indicate pathogenic variants P-SNV.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. Source data for panel 1C.

Figure supplement 1. Quality control for DNA analysis and snRNA-seq.

Figure supplement 1—source data 1. Source data corresponding to panels A, B, C, and I.

Figure 1—figure supplement 1B and C). Single nuclei (sn)RNA-seq was performed on PU.1$^+$ nuclei from one control and three AD patients to evaluate microglia enrichment following PU.1$^+$ purification, and a cell-type annotation analysis indicated that ~94% of PU.1$^+$ nuclei correspond to microglia (*Figure 1C* and *Figure 1—figure supplement 1D–1H*). Cortex samples were obtained from all donors but hippocampus samples were mostly obtained from AD patients (*Figure 1A*; *Supplementary file 1*). A total of 744 DNA samples from blood, PU.1$^+$ nuclei, NeuN$^+$ nuclei, and Double Negative nuclei (glia/stromal cells) from patients and controls (*Figure 1A*) were submitted to targeted hybridization/ capture and deep-DNA targeted sequencing (TDS, *Figure 1D*, see Materials and methods), at mean coverage of ~1100 x (*Figure 1—figure supplement 1I*), for a panel of 716 genes (3.43 Mb, referred to below as BRAIN-PACT) which included genes reported to carry somatic variants in clonal proliferative disorders (n=576 genes) *Cheng et al., 2015*; *Durham et al., 2019* or that have been reported to be associated with neurodegenerativediseases (n=140) (*Bras et al., 2012*; *Renton et al., 2014*; *Karch et al., 2014*; *Karch and Goate, 2015*; *Turner et al., 2013*; *Ferrari et al., 2015*; *Kouri et al., 2015*; *Scholz and Bras, 2015*; *Nalls et al., 2014*; *Supplementary file 2*, see Materials and methods).

After QC and filtering of germline variants, variant calling using ShearwaterML, and a curated Mutect1 analysis identified 826 somatic synonymous and non-synonymous single-nucleotide-variations (SNVs), at an allelic frequency >0.3% (mean 1.3%) in the 744 samples, corresponding to an overall variant burden of 0.3 mut/Mb (*Figure 1E*). Sixty-six/826 SNV were present in more than one sample (*Supplementary file 3*). Droplet digital-PCR performed on pre-amplification DNA for ~10% of the 760 unique SNV was positive in 90% of cases (*Figure 1E*; *Supplementary file 3*). After annotation using the OncoKB *Chakravarty et al., 2017* and ClinVar (*Landrum et al., 2014*) databases for disease-associated or causative variants (*Figure 2D and F*; *Supplementary file 3*), 96 unique SNV were classified as Pathogenic (P)-SNV. 40% of these P-SNV were tested by droplet digital-PCR and confirmed in 95% of cases (*Figure 1E* and *Supplementary file 3*). Positive and negative results in matching brain samples from individual donors were confirmed in 100% of samples at a mean depth of ~5000 x (range 648–23.000 x) (*Supplementary file 3*). A venn-diagram analysis of SNVs detected in PU.1$^+$, NeuN$^+$, DN, and blood samples indicated that most (>90%) SNV and P-SNV were cell-type or tissue-specific, with ~5% of SNV and ~8% of P-SNV shared between the blood and brain of individual donors (*Figure 1F*; *Supplementary file 3*). These data indicate that targeted deep-sequencing of purified nuclei allows to detect of clonal mosaic variants with high sensitivity and specificity. In addition, 'barcoding' of clonal variants across tissues suggests that infiltrating myeloid cells of peripheral origin account for ~5% of microglia somatic diversity, and therefore that blood clones have a detectable but minor contribution to microglia, consistent with its local maintenance and proliferation (*Askew et al., 2017*; *Réu et al., 2017*).

Somatic clonal diversity of the different cell types, as evaluated by the SNV/megabase burden was higher in blood (1 mut/Mb) and PU.1$^+$ nuclei (0.5 mut/Mb) than for DN and neurons (0.18 mut/Mb) (*Figure 2—figure supplement 1A*). The SNV/mb burden of blood and PU.1$^+$ nuclei increased as a function of age (*Figure 2A* and *Supplementary file 3*) as previously reported for proliferating cells (*Martincorena et al., 2015*; *Martincorena et al., 2018*; *Jaiswal et al., 2014*; *Genovese et al., 2014*). Interestingly, the SNV/mb burden of blood cells from age-matched controls was higher than for AD patients (*Figure 2B*). In contrast, there was no difference in SNV/mb burden between PU.1$^+$, NEUN$^+$,
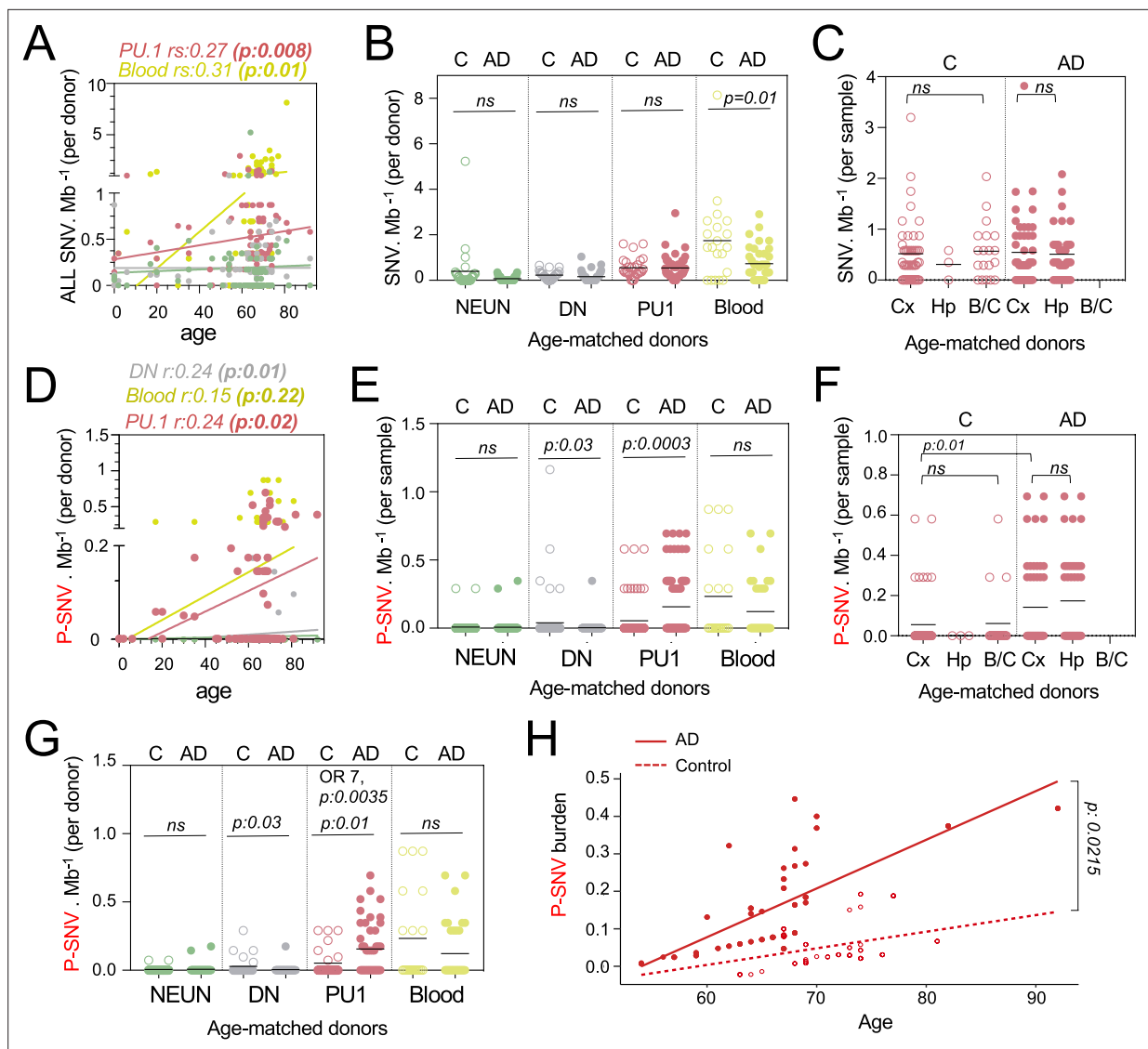
Figure 2. Pathogenic variants are enriched in microglia from Alzheimer's disease (AD) patients. (**A**) Correlation plot represents the mean number of variants per cell type and donor (n=89) (Y-axis), as a function of age (X-axis). Each dot represents mean value for a donor. Statistics: fitted lines, the correlation coefficients (rs), and associated p-*values* were obtained by linear regression (Spearman's correlation). (**B**) Number of single-nucleotide-variation (SNV) per Mb and cell types per donor, of age-matched controls (n=27) and AD patients (n=45). Each dot represents mean value for a donor. Statistics: *p-values* are calculated with unpaired two-tailed Mann-Whitney U. Note: non-parametric tests were used when data did not follow a normal distribution (D'Agostino-Pearson normality test). (**C**) Number of SNV per Mb in PU.1+ samples across brain regions, of age-matched controls (n=27) and AD patients (n=45). Each dot represents a sample. Statistics: *p-values* are calculated with Kruskal–Wallis, multiple comparisons. Note: non-parametric tests were used when data did not follow a normal distribution (D'Agostino-Pearson normality test). (**D**) Correlation plot represents the mean number of pathogenic variants (P-SNV) as determined by ClinVar and/or OncoKB, per cell type and donor (n=89) (Y-axis), as a function of age (X-axis). Each dot represents mean value for a donor. Statistics: fitted lines, the correlation coefficients (rs) and associated *p values* were obtained by linear regression (Spearman's correlation). (**E**) Number of P-SNV per Mb and cell types per sample, of age-matched controls (n=27) and AD patients (n=45). Each dot represents a sample. Statistics: *p-values* are calculated with unpaired two-tailed Mann-Whitney U. Note: non-parametric tests were used when data did not follow a normal distribution (D'Agostino-Pearson normality test). (**F**) Number of P-SNV per Mb in PU.1 samples across brain regions, of age-matched controls (n=27) and AD patients (n=45). Each dot represents a sample. Statistics: *p-values* for comparison within each group (controls and patients) are calculated with Kruskal–Wallis test and Dunn's test for multiple comparisons. *p-values* for the comparison of P-SNV between the cortex of C and the cortex of AD (0.01) was calculated with unpaired two-tailed Mann-Whitney U. Note: non-parametric tests were used when data did not follow a normal distribution (D'Agostino-Pearson normality test). (**G**) Number of P-SNV per Mb and cell types per donor for age-matched controls (n=27) and AD patients (n=45). Each dot represents mean value for a donor. Statistics: *p-values* are calculated with unpaired two-tailed Mann-Whitney U test. Odds ratio (95% CI, 2.049–29.02) and *p values* for the association between AD and the presence of pathogenic variants are calculated by multivariate logistic regression, with age and sex as covariates. (**H**) P-SNV burden as a function of age and disease status (age-matched controls and AD). Linear lines

Figure 2 continued on next page

Figure 2 continued

represent trend lines from mixed-effects linear regression that incorporates individual donor as a random effect (blue, control: p=0.0025, R^2=0.13; red, NDD: $P=9.1 \times 10^{-16}$, R^2=0.50 by Pearson's correlation). The model's total explanatory power is substantial (conditional R^2=0.48). Both age and AD are associated with a significant increase in SNV burden in this model ($P<1 \times 10^{-4}$ and $P=1 \times 10^{-4}$, respectively, by likelihood ratio test). Anatomical regions of the brain specimen and originating brain banks were not incorporated because the models incorporating those parameters did not significantly improve the overall model fitting by likelihood ratio test (see Methods). Graph depicts SNV burden corrected by the mixed-effects model (See *Figure 2—figure supplement 1E* for observed P-SNV burden).

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Data corresponding to panels A, B, C, D, E, F, G, and H.

Figure supplement 1. Analysis of pathogenic variants.

Figure supplement 1—source data 1. Data corresponding to panels A, B, C, D, E, and G.

Figure supplement 2. Summary of Alzheimer's disease (AD) patients characteristics and pathogenic variants.

and DN samples from AD patients and age-matched controls (*Figure 2B*)**,** and between PU.1+ nuclei from the cortex, hippocampus, and brainstem/cerebellum samples (*Figure 2C*). These data altogether indicate that the clonal diversity of microglia and blood both increase with age, and that the clonal diversity of blood cells is lower in AD than in age-matched controls who died of other causes including cancer and cardiovascular diseases (see Materials and methods). This is consistent with recent studies showing that clonal hematopoiesis is associated with a higher risk of several diseases related to ageing such as cardiovascular diseases, but is inversely associated with the risk of AD (*Jaiswal and Ebert, 2019*; *Bouzid et al., 2023*).

Microglia clones carrying pathogenic variants are enriched in AD patients

In contrast to the global SNV burden, increased P-SNV burden was correlated not only with age (*Figure 2D*), but also with the disease status (AD) (*Figure 2E–H*). Within the control group, the SNV and P-SNV burden was higher in the blood of controls treated for cancer (*Figure 2—figure supplement 1B*). The P-SNV burden per Mb was selectively and highly enriched in PU.1+ samples from AD patients in comparison to age-matched controls (p=0.0003, *Figure 2E*). Analysis of PU1+ P-SNV/Mb burden per brains region indicated that the P-SNV/Mb burden was similar between brain regions within each group (*Figure 2F* and *Figure 2—figure supplement 1C*), and therefore attributable to AD status rather than sampling bias. Analysis of mutational load per donor confirmed that microglial clones carrying P-SNV were enriched in the brain of AD patients in comparison to age-matched controls (*Figure 2G*). Despite the relatively modest cohort size, a logistic regression analysis confirmed the association between the presence of P-SNVs in PU.1+ nuclei and AD after adjusting for sex and age (*OR = 7*; p=0.0035, *Figure 2G* and *Figure 2—figure supplement 1D*). A mixed-effects linear regression model analysis also showed an excess of P-SNVs in AD independently of the effect of age (p=0.0215) (*Figure 2H* and *Figure 2—figure supplement 1E*).

In addition, genes targeted by P-SNV were all expressed in microglia (*Figure 2—figure supplement 1F* and *Supplementary file 3*) and the analysis of P-SNV/Mb mutational load restricted to genes that are not expressed in microglia did not show an enrichment of candidate pathogenic variants in AD patients (*Figure 2—figure supplement 1G* and *Supplementary file 3*). Altogether, these results show an association between microglia clones carrying P-SNV and AD in this series.

AD patients carry microglial clones with MAP-Kinase pathway variants including recurrent CBL variants

Pathways analysis of genes carrying P-SNV in microglia from AD patients, against the background of the 716 genes sequenced, showed that the most significant pathways enriched were the receptor tyrosine kinase/MAP-Kinase pathways (Reactome, GO, and canonical pathways, *Figure 3A* and *Supplementary file 4*), corresponding to pathogenic/oncogenic variants in 6 of the 15 genes of the classical MAPK pathway (*Rauen, 2013*) (CBL, BRAF, RIT1, NF1, PTPN11, KRAS), TEK**,** and the KEGG Chronic Myeloid Leukemia (CML) pathway, which includes the former plus SMAD5 and TP53 (*Figure 3B, C*, *Figure 2—figure supplement 2*). Mutational load for MAPK genes was significantly higher in AD patients in comparison to age-matched control (*Figure 3C*). Other enriched pathways, albeit less
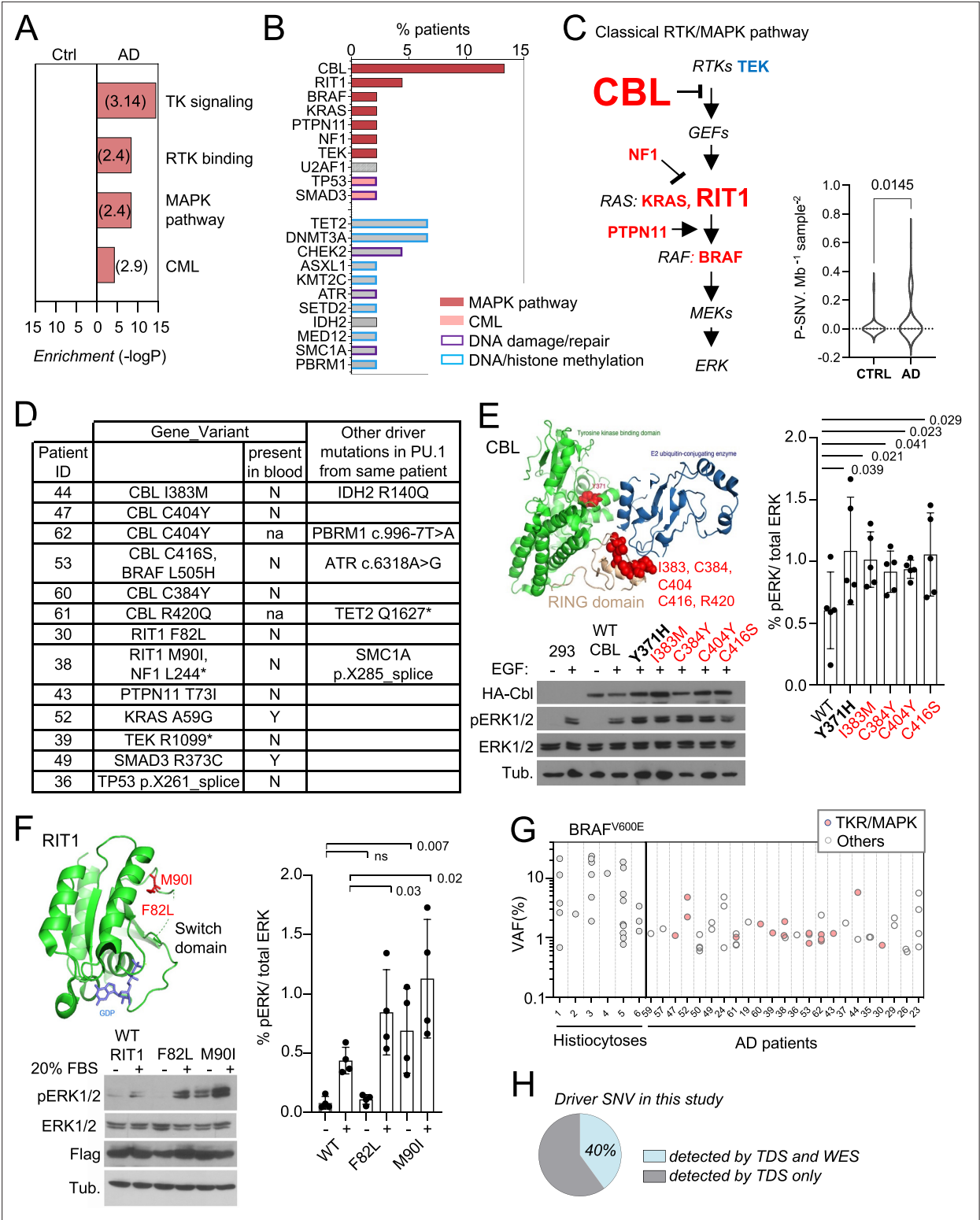
Figure 3. Somatic microglial clones with multiple and recurrent CBL and MAP-Kinase pathway activating variants. (**A**) Pathway enrichment analysis for the genes target of pathogenic variants (P-SNVs) using the panel of 716 genes as background set. Graph shows the most enriched pathways by: Reactome Gene Sets, GO Molecular Functions, Canonical Pathways and KEGG Pathway (see complete list in **Supplementary file 4**). (**B**) Bar plot indicates the genes carrying P-SNV (y-axis) and the % of Alzheimer's disease (AD) patients carrying P-SNV for each gene (x-axis). Genes are color-coded by pathway. (**C**) Representation of the classical MAPK pathway, the six genes mutated in AD patients are labeled in red, TEK is labeled in blue, and larger font size indicate reccurence of variants in a given gene. Violin plot shows distribution of P-SNV in genes from classical MAPK pathway per Mb

Figure 3 continued on next page

Figure 3 continued

sequenced and per sample in patients and controls, *p-value*: unpaired two-tailed Mann-Whitney U test. (**D**) Summary Table showing patients carrying P-SNV in the classical RTK/MAPK pathway and Chronic Myeloid Leukemia (CML)-associated genes (see *Supplementary file 3*) and indicating the detection of variants in blood, and their association with other variants in microglia. (**E**) Recurrent variants in the ring-like domain of CBL are indicated in red on the diagram structure of gene, above representative Western blot from cell lysates from HEK293T cells expressing WT, positive control (Y371H), or CBL variants alleles found in patients, and stimulated with EGF or control, probed with antibodies against Phospho-p44/42 MAPK (Erk 1/2, Thr202/Tyr204), total p44/42 MAPK (Erk1/2), HA-tag, and tubulin (BOTTOM). Histogram (RIGHT) represents quantification of the increase of the Phospho-ERK1/2/total ERK1/2 ratio in western blots in n=5 independent experiments, statistics: unpaired one-tailed t-test. (**F**) RIT1 M90I and F82L are represented on the 3D structure of the gene (pdb code: 4klz, F82 is within a segment whose structure was not resolved) and representative western blot from HEK293T cells expressing Flag-RIT1 (WT and mutants) and treated -/+ 20% FBS before harvesting. Lysates were probed with antibodies against Phospho-p44/42 MAPK (Erk 1/2, Thr202/Tyr204), total p44/42 MAPK (Erk1/2, (MAPK)), Flag, and tubulin. Histogram (RIGHT) represents quantification of the increase of the Phospho-ERK1/2/total ERK1/2 ratio in western blots in n=4 independent experiments, statistics: unpaired one-tailed t-test. (**G**) Variant allelic frequency (VAF, %) for the BRAF[V600E] allele in PU.1[+] nuclei from brain samples from histiocytosis patients (each dot represents a sample) and for P-SNVs in in PU.1+ nuclei from brain of AD patients (each dot represent a variant). Note: non-parametric tests were used when data did not follow a normal distribution (D'Agostino-Pearson normality test). (**H**) Percentage of P-SNVs detected by targeted deep sequencing (TDS) which were also detected by Whole-Exome-Sequencing (WES).

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. Data corresponding to panels A, B, C, E, F, and G.

Source data 2. Original PNG files for western blot analysis, indicating the relevant bands and treatments, displayed in panel E and F.

Source data 3. Unedited western blot JPEGs.

Figure supplement 1. Functional analysis of variants in HEK293 and BV2 cell lines.

Figure supplement 1—source data 1. Data corresponding to panel B.

Figure supplement 1—source data 2. Original PNG files for western blot analysis, indicating the relevant bands and treatments, displayed in panel A and B.

Figure supplement 1—source data 3. Unedited western blot JPEGs.

---

significant, included genes involved in DNA repair and chromatin binding/methyltransferase activity (*Figure 3B*; *Supplementary file 4*). No pathway was enriched in age-matched controls. Of note, we did not observe microglia P-SNVs within genes reported to be associated with neurological disorders (*Supplementary file 2*) in patients (*Supplementary file 3*). P-SNV targeting genes of the classical RTK/MAPK pathway (*Figure 3C*) were detected in the PU.1[+] samples from ~25% of the AD patients tested (p=0.0145 vs age-matched controls, *Figure 3D*, *Figure 2—figure supplement 2*). Strikingly, half of these patients (six patients, 13% of AD patients in this series) carried recurrent P-SNV in the RING domain of CBL (*Schnittger et al., 2012*; *Fernandes et al., 2010*; *Sargin et al., 2007*; *Dunbar et al., 2008*; *Bernard et al., 2014*; *Loh et al., 2009*; *Grand et al., 2009*; *Klampfl et al., 2013*; *Niemeyer et al., 2010*; *Javadi et al., 2013*; *Ogawa, 2019*; *Figure 3B–E*). Two additional patients presented with P-SNV in the Switch II domain of RIT1 (*Gómez-Seguí et al., 2013*; *Figure 3B–F*). Microglia from the three other patients carried activating KRAS (p.A59G), PTPN11 (p.T73I), and TEK (p.R1099*) oncogenic variants previously described in cancer and sporadic venous malformations (*Kim et al., 2016*; *Niihori et al., 2005*; *Soblet et al., 2013*; *Figure 3B and D*). In addition, a 12th patient carried a gain of function (GOF) U2AF1 (p.S34F) variant (*Okeyo-Owuor et al., 2015*), which is not a 'classical MAPK gene' but activates the MAPK pathway in myeloid malignancies (*Smith et al., 2019*; *Figure 3B*, *Figure 2—figure supplement 2*). Two patients carried two different MAPK activating variants: microglia from one patient carried an activating BRAF (p.L505H) variant *Choi et al., 2014* in addition to loss of function (LOF) variant CBL (p.C416S), and another patient carried the NF1 (p.L2442*) LOF variant *Heim et al., 1995*; *Bollag et al., 1996* in addition to the activating RIT1 (p.M90I) variant (*Figure 3D*, *Figure 2—figure supplement 2*). Five patients also carried additional P-SNV targeting genes involved in DNA repair with tumor suppressor function *Song et al., 2019*; *Zannini et al., 2014*, including the loss of function variants in ATR (c.6318A>G) *Fang et al., 2004* and SMC1A (p.X285_splice) (*Figure 3D*, *Figure 2—figure supplement 2*), and in DNA/histone methylation including TET2 (p.Q1627*) (*Schnittger et al., 2012*; *Haferlach et al., 2014*), IDH2 (p.R140Q) (*Ward et al., 2010*), and PBRM1 (c.996–7T>A) (*Brownlee et al., 2014*; *Figure 3D*, *Figure 2—figure supplement 2*). Finally, two patients carried oncogenic variants in genes from the KEGG Chronic Myeloid Leukemia (CML) pathway, SMAD3 (p.R373C) *Ku et al., 2007*, and TP53 (pX261_splice) *Bougeard et al., 2015* (*Figure 3D* and *Figure 2—figure supplement 2*). The detection of multiple

oncogenic variants in the same patients is reminiscent of the features observed in myeloproliferative disorders described outside the brain (*Ogawa, 2019*; *Haferlach et al., 2014*).

Recurrent CBL and RIT1 variants activate the MAPK pathway

CBL is an E3 ubiquitin-protein ligase that negatively regulates RTK signaling via MAPK (*Liyasova et al., 2015*). CBL somatic and germ-line LOF variants such as R420Q have been previously associated with tumoral diseases including clonal myeloproliferative disorders *Schnittger et al., 2012*; *Fernandes et al., 2010*; *Sargin et al., 2007*; *Dunbar et al., 2008*; *Bernard et al., 2014*; *Loh et al., 2009*; *Grand et al., 2009*; *Klampfl et al., 2013*; *Niemeyer et al., 2010*; *Javadi et al., 2013*; *Ogawa, 2019* and RASopathies (*Brand et al., 2014*), respectively. We confirmed that CBL RING-domain variants found in AD patients increased MAPK phosphorylation in response to EGF upon expression of HA-tagged WT or mutant alleles in HEK293T cells (*Figure 3E*). RIT1 is a RAS GTPase, and somatic or germ-line GOF variants such as RIT1 F82L and RIT1 M90I, also enhance MAPK signaling in malignancies *Gómez-Seguí et al., 2013* and RASopathies (*Meyer Zum Büschenfelde et al., 2018*; *Aoki et al., 2013*), respectively. As in the case of CBL variants, the 2 RIT1 variants found in AD patients increased MAPK phosphorylation in response to FBS in HEK293T cells expressing these mutant alleles (*Figure 3F*). These data altogether indicate that a subset of AD patients (12/45, ~27% of this series) present with microglial clones carrying one or several oncogenic variants that activate the RTK/MAPK pathway, and are characterized by recurrent oncogenic variants in CBL and RIT1.

Allelic frequency of the patients' MAPK activating variants

The allelic frequencies at which MAPK activating variants are detected in brain samples from AD patients range from ~1–6% in microglia (*Figure 3G*), which correspond to mutant clones representing 2 to 12% of all microglia in these samples, assuming heterozygosity. This range of allelic frequency is frequently observed for the MAPK-activating BRAF^{V600E} variant in microglia isolated from brain samples of six patients diagnosed with *BRAF^{V600E+}* histiocytosis, a rare clonal myeloid disorder associated with neurodegeneration (*Mass et al., 2017*; *Boyd et al., 2020*; *Bhatia et al., 2020*; *Diamond et al., 2016*; *Héritier et al., 2018*; *Figure 3G* and *Supplementary file 5*). These data suggested that the size of the mutant microglial clones in AD patients was compatible with a role in a neuro-inflammatory/neuro-degeneration process.

Other variants found in microglia from AD patients

Pathogenic variants that did not involve the MAPK pathway included LOF variants in the DNA repair gene CHEK2 including CHEK2 c.319+1 G>A *Cybulski et al., 2011* and CHEK2 R346H (*Figure 2—figure supplement 2* and *Figure 3—figure supplement 1A*), Mediator Complex gene MED12 (*Graham and Schwartz, 2013*), Histone methyltransferases SETD2 *Yang et al., 2016* and KMT2C/MLL3, the DNA methyltransferase DNMT3A (*Haferlach et al., 2014*; *Walter et al., 2011*), DNA demethylating enzymes TET2 and the Polycomb proteins ASXL1 (*Haferlach et al., 2014*). Of note, TET2, DNMT3, and KMT2C variants when present, were frequently detectable in the patients' matching blood at low allelic frequency (*Figure 2—figure supplement 2*). TET2, DNMT3, and KMT2C are frequently mutated in clonal hematopoiesis (*Jaiswal et al., 2014*; *Genovese et al., 2014*), suggesting that in contrast to other variants, the presence of TET2, DNMT3, and KMT2C/MLL3 in the brain of patients may reflect the entry of blood clones in the brain.

In half of the AD patients, no microglia pathogenic variants were identified. Targeted deep sequencing (TDS) cannot identify variants located outside of the BRAIN-PACT panel, such as other potential additional variants that would activate the MAPK pathway. Therefore, we performed whole exome sequencing (WES) of PU.1$^+$ nuclei at an average depth ~400 x, in selected samples from 48 donors, including samples from most of the patients negative for pathogenic variants by TDS (n=17 out of 22), a selection of patients with variants identified by TDS (n=16 out of 23), and 15 controls, followed by a curated Mutect analysis. Only 6/15 (40%) of the pathogenic SNVs previously identified by TDS and confirmed by ddPCR were detectable by WES in these samples (*Figure 3H*), indicating a lower sensitivity of WES. Nevertheless, after annotation by four modeling predictors Polyphen, SIFT, CADD/MSC, and FATHMM-XF (*Xi et al., 2004*; *Shihab et al., 2015*; *Adzhubei et al., 2010*; *Ng and Henikoff, 2001*; *Kircher et al., 2014*; *Itan et al., 2016*) additional SNVs predicted to be deleterious with high confidence were identified in 8/22 patients without pathogenic variants identified by

TDS (*Figure 2—figure supplement 2* and *Supplementary file 6*). Interestingly, four of the predicted deleterious variants identified by WES targeted genes that regulate the MAPK pathway ARHGAP9, ARHGEF26, CHD8, and DIXDC1 (*Figure 2—figure supplement 2* and *Supplementary file 6*).

The patients' MAPK activating variants increase ERK phosphorylation, proliferation, inflammatory, and mTOR pathways in murine microglia and macrophages

CBL variants increased ERK phosphorylation upon lentiviral transduction in BV2 murine microglial cells (*Blasi et al., 1990*; *Henn et al., 2009*; *Figure 3—figure supplement 1B*). However, as this line was immortalized by v-Raf, which might interfere with the study of the MAPK pathway, we also stably expressed WT and variant CBL, RIT1, KRAS, PTPN11 alleles in SV-U19–5 transformed mouse 'MAC' lines (*Yu et al., 2008*; *Xiong et al., 2011*) (see Methods and *Figure 4—figure supplement 1A and B*). MAC lines expressing CBL, RIT1, KRAS, and PTPN11 variants presented with increased ERK phosphorylation and/or increontrols, as measured by Western immunoblotting and EdU incorporation (*Figure 4A* and *Figure 4—figure supplement 1A and B*). In addition, Hallmark and KEGG pathway analysis of RNAseq data from control and mutant lines showed increased RAS, TNF, IL6, and JAK STAT signaling, complement, inflammatory responses, and mTOR pathway activation signatures in mutants (*Figure 4B* and *Supplementary file 7*). These data indicated that microglia variants from patient's activate murine microglial cells and growth factor-dependent macrophages with proliferative and inflammatory responses in vitro. However, overexpression of mutant alleles in mouse cell lines does not necessarily recapitulate or predict the effects of a heterozygous genetic variant in physiological conditions. Thus, we investigated the role of CBL^C404Y allele in heterozygous human primary microglia-like cells.

Heterozygosity for a CBL variant allele activates human microglia-like cells

We used prime editing *Anzalone et al., 2019* of human induced pluripotent stem cells (hiPSCs, see Methods) to generate isogenic hiPSCs clones heterozygous for the patients' variants (*Figure 4C* and *Figure 4—figure supplement 1*). We focused our analysis on CBL^404C/Y mutant lines because CBL was mutated in six patients and two of them carried the same *CBL c.1211G>A* p.C404Y variant (*Figure 2—figure supplement 2*). Microglia-like cells were differentiated from two independent hiPSC-derived CBL^404C/Y lines and their isogenic CBL^404C/C controls (*Figure 4C* and *Figure 4—figure supplement 1C*). CBL^404C/Y and isogenic CBL^404C/C microglia-like cells expressed similar amount of CBL total mRNA and protein, and CBL^404C/Y cells expressed wt and mutant mRNA in similar amounts, as expected assuming bi-allelic expression of CBL (*Figure 4—figure supplement 1D-5F*). CBL^404C/Y cells presented with a phenotype comparable to isogenic CBL^404C/C microglia-like cells for expression of IBA1, CSF1R, NGFR, EGFR, CD11b, MRC1, CD36, CD11c, Tim4, CD45, and MHC Class II (*Figure 4—figure supplement 1G*). Their viability was also comparable to control (*Supplementary file 5H*). However, CBL^404C/Y cells were larger and presented with more lamellipodia, resulting in an amoeboid morphology less frequently observed in isogenic controls (*Figure 4D and E*), and their proliferation rate was slightly increased, as measured by EdU incorporation (*Figure 4F*). Moreover CBL^404C/Y cells cultured in CSF1-supplemented medium also presented with a higher basal pERK level than control when restimulated with CSF1 (*Figure 4—figure supplement 1I*), and ERK phosphorylation after stimulation of starved microglia-like cells with CSF-1 was increased by ~ twofold in comparison to isogenic WT (*Figure 4G*). Altogether, these results showed that heterozygosity for a CBL^C404Y allele is sufficient to activate human microglia-like cells increasing their proliferation and ERK activation.

Heterozygosity for a CBL^C404Y allele drives a microglial neuroinflammatory/AD associated signature

Gene Set Enrichment Analyses (GSEA) of RNAseq comparing CBL^404C/Y and isogenic CBL^404C/C microglia-like cells showed upregulation of Glycolysis, Oxidative Phosphorylation, and mTORC1 signatures, indicating increased metabolism and energy consumption by the mutant cells (*Figure 5A* and *Supplementary file 8*). In addition, as observed in MAC lines, CBL^404C/Y cells upregulated complement, TNF, and JAK STAT signaling and inflammatory signatures (*Figure 5A*; *Supplementary file 8*; *Ghosh et al., 2018*). Increased production of TNF, IL-6, IFN-γ, IL-1β, C3, and complement Factor
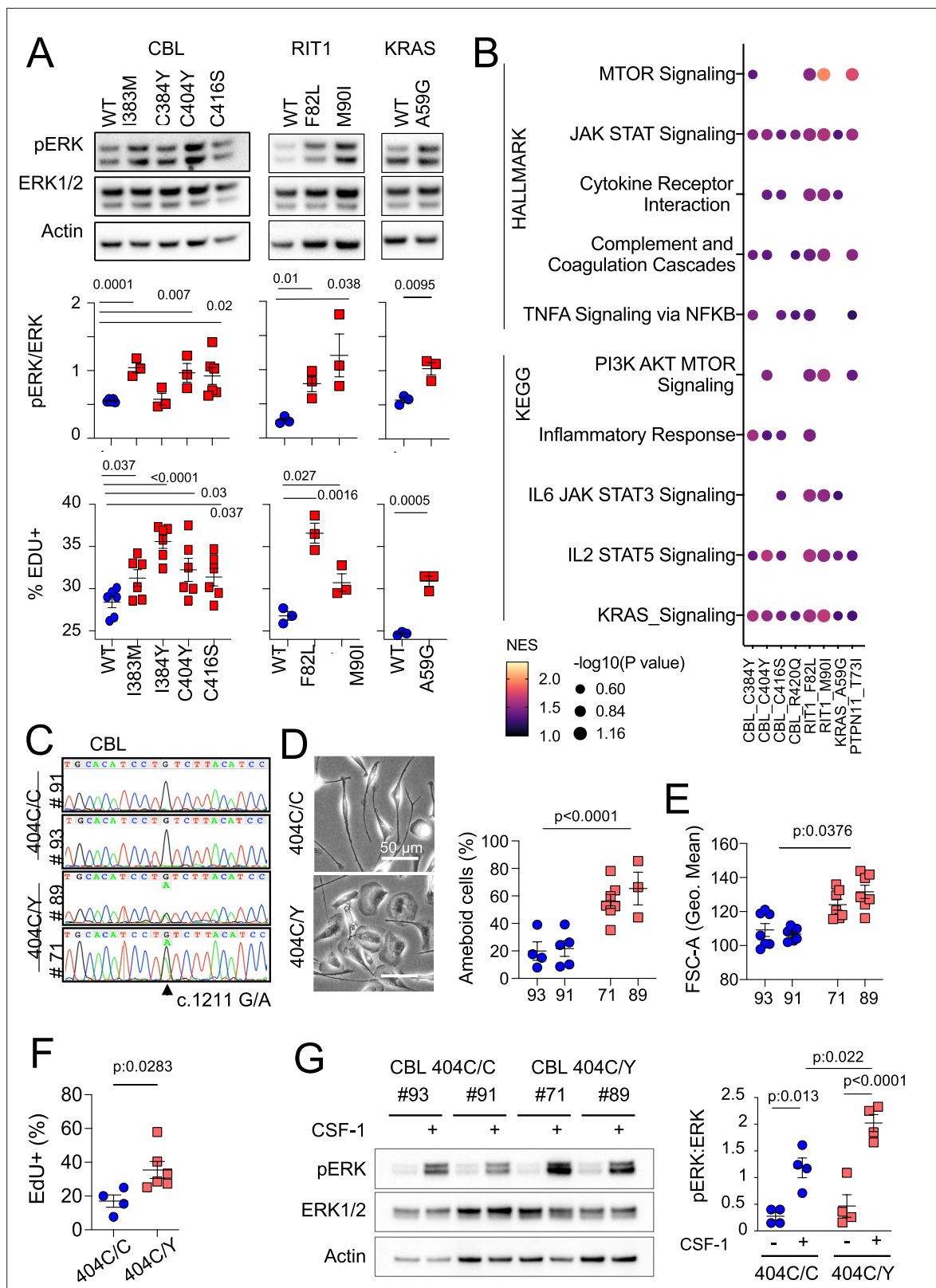
Figure 4. MAPK pathway activating variants in mouse macrophages and human induced Pluripotent Stem Cells (iPSC)-derived microglia-like cells. (**A**) Representative western-blot analysis (Top panels) and quantification (Middle panels) of phospho- and total-ERK in lysates from a murine CSF-1 dependent macrophage cell line expressing CBL^{WT}, CBL^{I383M}, CBL^{C384Y}, CBL^{C404Y}, CBL^{C416S} (n=3–6), and RIT1^{WT}, RIT1^{F82L} and RIT1^{M90I} (n=3), KRAS^{WT}, and KRAS^{A59G} (n=3). Bottom panels depicts flow cytometry analysis of EdU incorporation in the same lines. Statistics, Unpaired t-test. (**B**) HALLMARK and

Figure 4 continued on next page

Figure 4 continued

KEGG pathways (FDR/adj.p value <0.25, selected from **Supplementary file 7**) enriched in gene set enrichment analysis (GSEA) of RNAseq from mutant CSF-1 dependent macrophages lines CBL^{I383M}, CBL^{C384Y}, CBL^{C404Y}, CBL^{C416S}, CBL^{R420Q}, RIT1^{F82L} RIT1^{M90I}, KRAS^{A59G}, and PTPN11^{T73I} (n=3–6) in comparison with their wt controls. NES: normalized enrichment score. (C) Sanger sequencing of 2 independent hiPSC clones (#93 and #91) of CBL$^{404C/Y}$ heterozygous mutant carrying the c.1211G/A transition on one allele and 2 independent isogenic control CBL$^{404C/C}$ clones (#71 and #89) all obtained by prime editing. (D) Photomicrographs in CBL$^{404C/C}$ and CBL$^{404C/Y}$ iPSC-derived microglia-like cells.(E) Quantification of leading edge and lateral lamellipodia in CBL$^{404C/C}$ and CBL$^{404C/Y}$ iPSC-derived microglia-like cells. n=3–7, statistics: p-value are obtained by nested one-way ANOVA. (F) Flow cytometry analysis of cell size for the same lines (n>3) statistics: p-value are obtained with nested one-way ANOVA. (G) Flow cytometry analysis of EdU incorporation in CBL$^{404C/C}$ and CBL$^{404C/Y}$ microglia-like cells after a 2 hr EdU pulse. n=3, unpaired t-test. (H) Western-blot analysis (left) and quantification (right) of phospho- and total-ERK proteins in lysates from CBL$^{404C/C}$ and CBL$^{404C/Y}$ microglia-like cells starved of CSF-1 for 4 hr and stimulated with CSF-1 (5 min, 100 ng/mL) (n=4), statistics: p-value are obtained with two-way ANOVA.

The online version of this article includes the following source data and figure supplement(s) for figure 4:

Source data 1. Data corresponding to panels A, D, E, F, and G.

Source data 2. Original PNG files for western blot analysis, indicating the relevant bands and treatments, displayed in A and G.

Source data 3. Unedited western blot JPEGs.

Figure supplement 1. Analysis of mouse and human microglia-like cells.

Figure supplement 1—source data 1. Data corresponding to panels B, D, E, H, I.

Figure supplement 1—source data 2. Original PNG files for western blot analysis, indicating the relevant bands and treatments, displayed in panels A, B, F, and I.

Figure supplement 1—source data 3. Unedited western blot JPEGs.

H (CFH) by CBL$^{404C/Y}$ cells was confirmed by ELISA (**Figure 5B**). In addition, CBL$^{404C/Y}$ microglia-like cells also presented with signatures from the KEGG database associated with neurodegenerative disorders (**Figure 5A**; **Supplementary file 8**), and for the recently published human microglia AD scRNA-seq signature, obtained by analysis of 24 sporadic AD patients and 24 controls (**Mathys et al., 2019**; **Figure 5C**). These data indicated that heterozygosity for the CBL^{C404Y} allele is sufficient to drive expression of a neuroinflammatory/AD signature in a human microglia-like cell type, characterized by increased metabolism and the production of neurotoxic cytokines known to interfere with normal brain homeostasis.

The MAPK variant neuroinflammatory microglial signature is detectable in patients

Analysis of the snRNA-seq data from five samples of purified microglia nuclei from four donors control, AD without and with pathogenic variants (**Figure 1C**, **Figure 1—figure supplement 1D–H** and **Supplementary file 9**) using unsupervised Louvain clustering and GSEA showed that microglia samples from patients carrying variants were enriched for the signatures observed in the MAC lines and CBL$^{404C/Y}$ cells (**Figure 5D and E** and **Figure 5—figure supplement 1**). In particular, microglia cluster 2 and 2B, were most enriched for the inflammatory, TNF, mTOR, and oxidative phosphorylation and glycolysis signatures in patients carrying variants (AD52, AD53) but not the controls (C11, AD34) (**Figure 5D and E** and **Figure 5—figure supplement 1**). Despite the small size of the mutant clones and the low sensitivity of scRNA-seq to detect rare allelic variants, KRAS A59G variant reads were detected in cluster 2/2B from patient AD52.

Altogether, the above results support the hypothesis that patients' microglial clones carrying pathogenic mutations are associated with a metabolic and neuroinflammatory signature that includes the production of neurotoxic cytokines in vitro and in vivo.

Discussion

We report here that microglia from a cohort of 45 AD patients with intermediate-onset sporadic AD (mean age 65 y.o) is enriched for clones carrying pathogenic/oncogenic variants in genes associated with clonal proliferative disorders (**Supplementary file 2**) in comparison to 44 controls. Of note, we did not observe microglia P-SNVs within genes reported to be associated with neurological disorders in the patients.
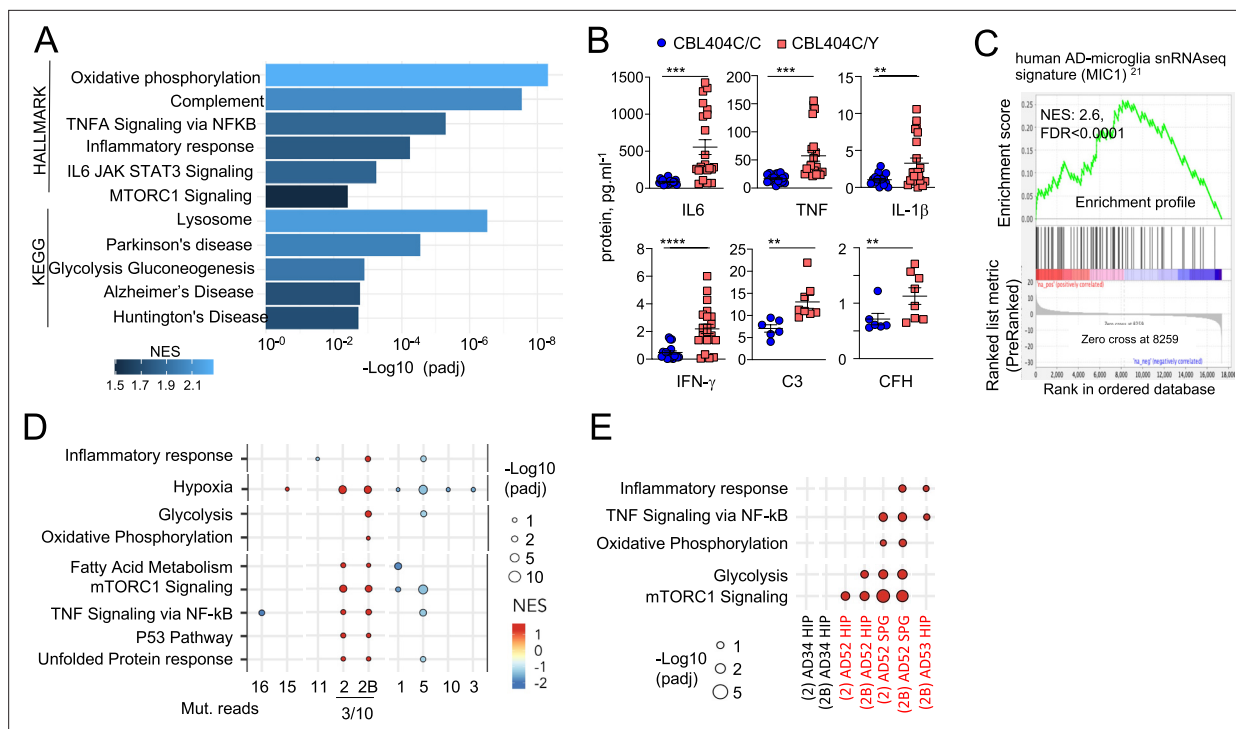
Figure 5. CBL[404C/Y] microglia signature. (**A**) HALLMARK and KEGG pathways (FDR /adj.p value <0.25, selected from *Supplementary file 8*) enriched in gene set enrichment analysis (GSEA) of RNAseq from from CBL[404C/Y] induced Pluripotent Stem Cells (iPSC)-derived macrophages and isogenic controls NES, normalized enrichment score. (**B**) ELISA for pro-inflammatory cytokines (n=3) and complement proteins (n=2) in the supernatant from CBL[404C/Y] iPSC-derived microglial-like cells and isogenic controls. Statistics: *p-value* are obtained by nonparametric Mann-Whitney U test,* *0.05, ** 0.01, *** 0.001, **** 0.0001*. (**C**) GSEA analysis for enrichment of the human AD-microglia snRNA-seq signature (MIC1) *Mathys et al., 2019* in differentially expressed genes between CBL[404Y/C] microglial-like cells and isogenic controls. (**D**) Dot plot represents the GSEA analysis of HALLMARK and KEGG pathways enriched in snRNAseq microglia clusters (samples from all donors). Genes are pre-ranked per cluster using differential expression analysis with SCANPY and the Wilcoxon rank-sum method. Statistical analyses were performed using the fgseaMultilevel function in fgsea R package for HALLMARK and KEGG pathways. Selected gene-sets with p-value <0.05 and adjusted p-value <0.25 are visualized using ggpubr and ggplot2 R package (gene sets/pathways are selected from *Figure 5—figure supplement 1B*, *Supplementary file 9*). (**E**) Dot plot represents the GSEA analysis (as in (**E**)) of HALLMARK and KEGG pathways enriched in cluster 2/2 B and deconvoluted by donor samples (selected from *Figure 5—figure supplement 1A*).

The online version of this article includes the following source data and figure supplement(s) for figure 5:

Source data 1. Data corresponding to panel B.

Figure supplement 1. snRNA-seq analysis of microglia.

These pathogenic variants are absent from blood, glia, or neurons in most cases. They are found predominantly in the MAPK pathway and include reccurrent variants (CBL RING domain variants in six patients), which promote microglial proliferation, activation, and expression of a neuroinflammatory/neurodegereration-associated transcriptional program in vitro and in vivo, and the production of neurotoxic cytokines IL1b, TNF, and IFNg (*Liu and Quan, 2018*; *Roy et al., 2020*; *Jayaraman et al., 2021*; *Ou et al., 2021*). Heterozygous expression of pathogenic CBL variant in human microglia-like cells was sufficient to drive a transcriptional program that associates with increased metabolic activity and a neurotoxic inflammatory response, also observed in microglia from patients with MAPK-activating variants.

The association between AD and MAPK pathway variants is consistent with a previous study where WES performed on unseparated brain tissue from AD patients showed that putative pathogenic somatic variants were enriched for the MAPK pathway, despite the lower sensitivity of the approach and the lack of cellular specificity (*Park et al., 2019*). The pathogenic role of the somatic pathogenic variants in the MAPK pathway associated with the microglia of AD patients is supported by several lines of evidence. We show here that they promote a neuroinflammatory/neurodegereration-associated transcriptional program in microglia-like cells. In addition, somatic variants that activate the MAPK pathway in tissue macrophages cause a clonal proliferative and inflammatory disease called

Histiocytosis, strongly associated with neurodegeneration (*Mass et al., 2017*; *Boyd et al., 2020*; *Bhatia et al., 2020*; *Diamond et al., 2016*), and introduction in mouse microglia of the variant allele most frequently associated with histiocytosis (BRAF^{V600E}) causes neurodegeneration in mice (*Mass et al., 2017*). The allelic frequencies of pathogenic variants found in AD patients are lower than values classically observed in solid tumors or leukemia, but within the range of the clonal frequency of pathogenic T cells observed in auto-immune diseases (*Thapa et al., 2015*), and we found that they were in the range of the allelic frequencies observed for the BRAF^{V600E} variant in microglia in the brain of Histiocytosis patients. Moreover, the RAS/MAPK signaling pathway is involved in microglia proliferation, activation, and inflammatory response (*Lindberg et al., 2014*; *Qu et al., 2012*; *Coniglio et al., 2012*), neuronal death, neurodegeneration, and AD pathogenesis (*McQuade and Blurton-Jones, 2019*; *Nott et al., 2019*; *Mass et al., 2017*; *Scheltens et al., 2018*), and its activation has been proposed to be an early event in the pathophysiology of AD in human (*Lachén-Montes et al., 2016*). Neuroinflammation is an early event in AD pathogenesis, increasingly considered as critical in pathogenesis initiation and progression (*Arnaud et al., 2022*; *Schöll et al., 2015*; *Kinney et al., 2018*). This is underscored by the observation that the main known genetic risk factor for sporadic AD is the APOE4 allele, responsible for an increased inflammatory response in the brain of APOE4 carriers (*Arnaud et al., 2022*). In this regard, the contributing role of MAPK activating variants could be comparable to that of the APOE4 allele, and we noted that the allelic frequency of APOE4 allele is lower in patients with pathogenic variants (16/46 alleles, 34%) than in patients without detected variant (23/44 alleles, 53%) although the difference did not reach significance in this series.

Variants targeting the DNA-repair and DNA/histone methylation pathways are also enriched among AD patients, sometimes associated with the same patients, albeit their functional significance was not investigated here. Of note, however, germline variants of the DNA-repair transcription factor TP53, and DNA damage sensors ATR and CHEK2 were shown to promote accelerated neurodegeneration in human (*Song et al., 2019*; *Zannini et al., 2014*).

Microglia variants are frequently absent from blood, and our DNA sequencing barcoding approach does not support a model where blood cells massively infiltrate the brain or replace the microglia pool in patients from our series, but instead consistent with the local maintenance and proliferation of microglia (*Askew et al., 2017*; *Réu et al., 2017*). In addition, our results are consistent with a recent study showing that clonal hematopoiesis was inversely associated with the risk of AD (*Bouzid et al., 2023*).

The association of microglia clones carrying pathogenic variants with AD in a subset of patients is not a consequence of an overall increase in microglia mutational load (SNV) in AD. Together with evidence that pathogenic variants drive neuroinflammation, these data suggest that these clones could contribute to AD pathogenesis, together with other genetic and environmental factors. Lewy bodies, amyloid angiopathy, tauopathy, or alpha synucleinopathy, were equally distributed among AD patients with or without microglia clones carrying MAPK activating variants. The natural history of the microglial clones is difficult to study in human. It is possible that microglial clones with proliferative and activation advantages and a neuroinflammatory and neurotoxic profile may be present at the onset and contribute to the early stages of the disease. Alternatively, it is also possible that the microglial clones carrying the pathogenic mutations appear or are selected later during the course of the disease in the inflammatory milieu of the AD brain. In the latter case, pathogenic microglial clones may contribute to disease progression, i.e., neuroinflammation and neurodegeneration.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line (*Homo sapiens*)	293T cell line	ATCC	RRID:CVCL_0063	
Cell line (*Mus musculus*)	BV2 microglial cell line	ARP American Research Products, Inc	RRID:CVCL_0182	inmortalized with v-raf

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line (*Mus musculus*)	Mouse primary CSF-1	gift of Dr. E. R. Stanley (Albert Einstein College of Medicine, Bronx, NY).		immortalized with the SVU19-5 retrovirus
Cell line (*Homo sapiens*)	C12 induced Pluripotent Stem Cells (iPSCs) (female)	Derived from peripheral blood mononuclear cells (PBMCs)		Used as a healthy wildtype control iPSC line (WT CBL)
Cell line (*Homo sapiens*)	293T- Flag-tagged WT and 293 T-R346H CHK2	This paper		generated by transfection
Cell line (*Homo sapiens*)	293T-Flag-tagged WT, 293T- RIT1 F82L, 293T-RIT1 M90I	This paper		generated by transfection
Cell line (*Homo sapiens*)	293T-CBL, 293T-CBLI383M, 293T-CBLC404Y, 293T-CBLC416S, 293T-CBLC384Y, 293T-CBLY371H	This paper		generated by transfection
Cell line (*Mus musculus*)	BV2-empty vector, BV2-CBL, BV2-CBLI383M, BV2-CBLC404Y,BV2-CBLC416S, BV2-CBLC384Y, BV2-CBLY371H	This paper		generated by viral transduction
Cell line (*Mus musculus*)	MAC-RIT1, MAC-RIT1 F82L, MAC-RIT1 M90I	This paper		generated by viral transduction
Cell line (*Homo sapiens*)	MAC-CBL, MAC-CBLI383M, MAC-CBLC404Y, MAC-CBLC416S, MAC-CBLC384Y, MAC-CBLY371H	This paper		generated by viral transduction
Cell line (*Homo sapiens*)	MAC-KRAS, MAC-KRASA59G	This paper		generated by viral transduction
Cell line (*Homo sapiens*)	MAC-PTPN11, MAC- PTPN11T73I	This paper		generated by viral transduction
Cell line (*Homo sapiens*)	IPSC-CBLC404Y	This paper		C12 line genetically modified to contain CBLC404Y
Transfected construct (human)	Flag-tagged CHK2	Sino Biological		
Transfected construct (human)	Flag-tagged RIT1	Origene		
Transfected construct (human)	Flag-tagged RIT1M90I and Flag-tagged RIT1F82L	This paper		generated by site-directed mutagenesis using the QuikChange Kit
Transfected construct (human)	pcDNA3-HA-tagged c-Cbl	gift from Dr. Nicholas Carpino (Stony Brook)		
Transfected construct (human)	pcDNA3-HA-tagged- CBLI383M, CBLC404Y, CBLC416S, CBLC384Y, CBLY371H	This paper		generated by site-directed mutagenesis using the QuikChange Kit
Transfected construct (human)	pHAGE_puro	gift from Christopher Vakoc	Addgene plasmid # 118692; http://n2t.net/addgene:118692; RRID:Addgene_118692	
Transfected construct (human)	pHAGE-KRAS	gift from Gordon Mills & Kenneth Scott	Addgene plasmid # 116755; http://n2t.net/addgene:116755; RRID:Addgene_116755	
Transfected construct (human)	pHAGE-PTPN11	gift from Gordon Mills & Kenneth Scott	Addgene plasmid # 116782; http://n2t.net/addgene:116782; RRID:Addgene_116782	
Transfected construct (human)	pHAGE-PTPN11-T73I	gift from Gordon Mills & Kenneth Scott	Addgene plasmid # 116647; http://n2t.net/addgene:116647; RRID:Addgene_116647	

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Transfected constructs (human)	Phage-CBL, Phage-CBLI383M, Phage-CBLC404Y, Phage-CBLC416S, Phage-RIT1, Phage-RIT1M90I, Phage-RIT1F82L, phage-KRASA59G and pHAGE-CBLC384Y	This paper		generated by Azenta Life Sciences via a PCR cloning approach and targeted mutagenesis
Transfected construct (human)	pDONR223_KRAS_p.A59G	gift from Jesse Boehm & William Hahn & David Root	Addgene plasmid # 81662; http://n2t.net/addgene:81662; RRID:Addgene_81662	
Biological samples (Homo-sapiens)	Biological samples (brain tissue and blood) from patients and controls	See 'Human sample collection and consent information' in methods for details		
Biological samples (*Mus musculus*)	CF1 Mouse Embryonic Fibroblasts, irradiated	Thermo Fisher Scientific	A34181	
Antibody	NeuN-PE. Mouse monoclonal	Milli-Mark	Cat# FCMAB317PE, RRID:AB_11212465	used at 1:500 dilution
Antibody	Pu.1-AlexaFluor 647. Rabbit monoclonal	Cell Signaling	Cat# 2240, RRID:AB_2186911	used at 1:50 dilution
Antibody	anti-Cdc42, rabbit polyclonal	Santa Cruz	Cat# sc-87, RRID:AB_631213	1 µg
Antibody	Phospho-p44/42 MAPK (Thr202/Tyr204) mouse monoclinal	Cell Signaling	Cat# 4370, RRID:AB_2315112	used at 1:1000 dilution
Antibody	total p44/42 MAPK, rabbit polyclonal	Cell Signaling	Cat# 9102, RRID:AB_330744	used at 1:1000 dilution
Antibody	HA, mouse monoclonal	Millipore	Cat# 05–904, RRID:AB_417380	used at 1:1000 dilution
Antibody	Flag, mouse monoclonal	Sigma	Cat# A8592, RRID:AB_439702	used at 1:1000 dilution
Antibody	pCHEK2 (T383), Rabbit Polyclona	Abcam	Cat# ab59408, RRID:AB_942224	used at 1:500 dilution
Antibody	Anti-γ-Tubulin, mouse monoclonal	Sigma	Cat# T6557, RRID:AB_477584	used at 1:10000 dilution
Antibody	anti-c-CBL, Rabbit Polyclonal	Cell Signaling	Cat# 2747, RRID:AB_2275284	used at 1:1000 dilution
Antibody	anti-RIT1, Rabbit Polyclonal	Abcam	Cat# ab53720, RRID:AB_882379	used at 1:1000 dilution
Antibody	anti-KRAS, Rabbit Polyclonal	Sigma	Cat# H00003845-M01, RRID:AB_540078	1 µg/mL
Antibody	anti-Actin, mouse monoclonal	Sigma	Cat# MAB1501, RRID:AB_2223041	used at 1:10,000 dilution
Antibody	anti-rabbit IgG HRP-linked, goat polyclonal	Cell Signaling	Cat# 7074, RRID:AB_2099233	used at 1:1000 dilution
Antibody	anti-mouse IgG HRP-linked, horse polyclonal	Cell Signaling	Cat# 7076, RRID:AB_330924	used at 1:1000 dilution
Antibody	PE- anti-CD115 (CSF1-R), rat monoclonal	BD	Cat# 565368, RRID:AB_2739206	used at 1:100 dilution
Antibody	PE/Cy7-conjugated anti-CD11b, mouse monoclonal	BD Biosciences	at# 557743, RRID:AB_396849	used at 1:100 dilution
Antibody	Alexa Fluor 488-conjugated-anti-CD206, mouse monoclonal	Thermo Fisher Scientific	Cat# 53-2069-42, RRID:AB_2574416	used at 1:100 dilution
Antibody	PE-conjugated anti-integrin, mouse monoclonal	R&D systems	Cat# FAB3050P, RRID:AB_920540	used at 1:100 dilution
Antibody	PE/Cy5-conjugated anti-CD11c, mouse monoclonal	BD Biosciences	Cat# 551077, RRID:AB_394034	used at 1:100 dilution
Antibody	APC-conjugated anti-Tim4, mouse monoclonal	Biolegend	Cat# 354007, RRID:AB_2564543	used at 1:100 dilution
Antibody	PE/Cy7-conjugated anti-HLA-DR, mouse monoclonal	BD Biosciences	Cat# 560651, RRID:AB_1727528	used at 1:100 dilution
Antibody	BV650-conjugated anti-CD45, mouse monoclonal	Thermo Fisher Scientific	Cat# 416-0459-42, RRID:AB_2925684	used at 1:100 dilution

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	APC/Cy7-conjugated anti-CD14, rat monoclonal	Biolegend	Cat# 123317, RRID:AB_10900813	used at 1:100 dilution
Antibody	PE-conjugated anti-NGFR, mouse monoclonal	eBioscience	Cat# 12-9400-42, RRID:AB_2572710	used at 1:100 dilution
Antibody	Alexa Fluor 647-conjugated anti-EGFR, mouse monoclonal	BD Pharmigen	Cat# 563577, RRID:AB_2738288	used at 1:100 dilution
Antibody	APC/Cy7-conjugated anti- CD36, mouse monoclonal	Biolegend	Cat# 336213, RRID:AB_2072512	used at 1:100 dilution
Antibody	APC-conjugated anti-CD172a (SIRPa), mouse monoclonal	Thermo Fisher Scientific	Cat# 17-1729-42, RRID:AB_1944409	used at 1:100 dilution
Antibody	Alexa Fluor 555-conjugated anti-Iba1 antibody, Rabbit monoclonal	Cell Signaling	Cat# 36618, RRID:AB_2943227	used at 1:100 dilution
Commercial assay or kit	KAPA Hyper Prep Kit	Kapa Biosystems	KK8504	
Software, algorithm	muTect 1	https://github.com/soccin/BIC-variants_pipeline and https://github.com/soccin/Variant-PostProcess; doi:10.1016/j.jmoldx.2014.12.006	v1.1.7	
Software, algorithm	ShearwaterML	Martincorena, I. et al.		
Software, algorithm	FlowJo	BD	10.6.2	
Commercial assay or kit	QIAamp DNA Micro Kit	Qiagen	Cat#56304	
Commercial assay or kit	HiSeq 3000/4000 SBS Kit	Illumina		
Sequence-based reagent	KRAS_G12D, ddPCR	Bio-Rad	Unique Assay ID: dHsaMDV2510596	
Sequence-based reagent	MTOR_Arg1616His_c.4847G>A	Bio-Rad	Unique Assay ID: dHsaMDV2510596	
Commercial assay or kit	10 X genomics Reagent Kit 3' v3.1	10 X genomics		
Software, algorithm	Seurat v4.0.3	https://github.com/satijalab; *Hao et al., 2021*		
Commercial assay or kit	NovaSeq 6000 S4 Reagent Kit (200 Cycles)	Illumina		
Commercial assay or kit	RNeasy Mini kit	Qiagen		
Sequence-based reagent	CBL_I383M	Thermo Fisher Scientific	dHsaMDS675699482	
Sequence-based reagent	CBL_C384Y	Thermo Fisher Scientific	dHsaMDS386449640	
Sequence-based reagent	CBL_C404Y	Thermo Fisher Scientific	dHsaMDS437459772	
Sequence-based reagent	CBL_mRNA_C404Y	Thermo Fisher Scientific	dMDS334857054	
Sequence-based reagent	CBL_C416S	Thermo Fisher Scientific	dHsaMDS613275900	
Sequence-based reagent	RIT1_ F82L	Thermo Fisher Scientific	dMDS959028273	
Sequence-based reagent	RIT1_M90I	Thermo Fisher Scientific	dHsaMDS133045056	

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	c-CBL FAM	Thermo Fisher Scientific	Hs01011446_m1	
Sequence-based reagent	CBLb FAM	Thermo Fisher Scientific	Hs00180288_m1	
Sequence-based reagent	GAPDH VIC	Thermo Fisher Scientific	Hs02786624_g1	
Commercial assay or kit	Zombie Violet Viability	Biolegend		
Commercial assay or kit	Cytofix/Cytoperm solution	BD Pharmingen		
Commercial assay or kit	Click-iT Plus EdU Alexa Fluor 647 Flow Cytometry Assay Kit	Thermo Fisher Scientific		
Commercial assay or kit	miRNeasy Mini Kit	Qiagen		
Commercial assay or kit	MagMAX mirVana Total RNA Isolation Kit	Thermo Fisher Scientific		
Commercial assay or kit	KingFisher Flex Magnetic Particle Processor	Thermo Fisher Scientific		
Commercial assay or kit	TruSeq Stranded mRNA LT Kit	Illumina		
Software, algorithm	R/Bioconductor package DESeq	EMBL Heidelberg	https://bioconductor.org/packages//2.10/bioc/html/DESeq.html	

Tissue samples

The study was conducted according to the Declaration of Helsinki. Human tissues were obtained with patient-informed consent and used under approval by the Institutional Review Boards from Memorial Sloan Kettering Cancer Center (IRB protocols #X19-027). Snap-frozen human brain and matched blood were provided by the Netherlands Brain Bank (NBB), the Human Brain Collection Core (HBCC, NIH), Hospital Sant Joan de Déu, and the Rapid Autopsy Program (MSKCC, IRB #15–021). Samples were neuropathologically evaluated and classified by the collaborating institutions as AD (*Dubois et al., 2007*; *Braak and Braak, 1991*; *Braak and Braak, 1995*; *McKhann et al., 1984*; *McKhann et al., 2011*) or non-dementia controls. The mean age of AD patients is 65 y old (55.5% female, 44.5% male). The mean age of all controls is 54 y old (60% female, 40% male), and the mean age of AD age-matched controls was 70 y old (60% female, 40% male). The overall mean of the post-mortem delay interval was 9.8 hr. Patients did not present with germline pathogenic PSEN1/2/3 or APP AD's associated variants. For additional information on donor's brain regions, sex, age, cause of death, Apoe status, Braak status see *Supplementary file 1*. To avoid possible contamination of sequencing data with mutations associated with donor's tumoral disease in the group of non-dementia controls, we refrained from selecting cases with blood malignancies or with brain tumors. Samples from histiocytosis patients were collected under GENE HISTIO study (approved by CNIL and CPP Ile-de France) from Pitié-Salpêtrière Hospital and Hospital Trousseau and from Memorial Sloan Kettering Cancer Center.

Nuclei isolation from frozen brain samples, FACS-sorting, and DNA extraction

All samples were handled and processed under Air Clean PCR Workstation. An average of 400 mg of frozen brain tissues were homogenized with a sterile Dounce tissue grinder using a sterile non-ionic surfactant-based buffer to isolate cell nuclei ('homogenization buffer:' 250 mM Sucrose, 25 mM KCL, 5 mM MgCl2, 10 mM Tris buffer pH 8.0, 0.1% (v/v) Triton X-100, 3 µM DAPI, Nuclease Free Water). Homogenate was filtered in a 40 µm cell strainer and centrifuged 800 g 8 min 4 °C. To clean-up the homogenate, we performed a iodixanol density gradient centrifugation as follows: pellet was gently mixed 1:1 with iodixanol medium at 50% (50% Iodixanol, 250 mM Sucrose, 150 mM KCL, 30 mM

MgCl2, 60 mM Tris buffer pH 8.0, Nuclease Free Water) and homogenization buffer. This solution layered to a new tube containing equal volume of iodixanol medium at 29% and centrifuged 13.500 g for 20 min at 4 °C. Nuclei pellet was gently resuspended in 200 µl of FACS buffer (0.5% BSA, 2 mM EDTA) and incubated on ice for 10 min. After centrifugation 800 g 5 min 4 °C, sample was incubated with anti-NeuN (neuronal marker, 1:500, Anti-NeuN-PE, clone A60 Milli-Mark) for 40 min. After centrifugation 800 g 5 min 4 °C, sample was washed with 1 X Permeabilization buffer (Foxp3 /Transcription Factor Staining Buffer Set, eBioscience) and centrifuged 1300 g for 5 min, without breaks to improve nuclei recovery. Staining with anti-Pu.1 antibody in 1 X Permeabilization buffer (myeloid marker 1:50, Pu.1-AlexaFluor 647, 9G7 Cell Signaling) was performed for 40 min. After a wash with FACS buffer samples were prepared for FACS. Nuclei were FACS-sorted in a BD FACS Aria with a 100 µm nozzle and a sheath pressure 20 psi, operating at ~1000 events per second. Nuclei were sorted into 1.5 ml certified RNAse, DNAse DNA, ATP, and Endotoxins tubes containing 100 µl of sterile PBS. For detailes on sorted samples see *Supplementary file 1*. Sorting purity was >95%. Sorting strategy is depicted in *Figure 1—figure supplement 1*. Of note, the Double-negative gate is restricted to prevent cross-contamination between cell types. Nuclei suspensions were centrifuged 20 min at 6000 g and processed immediately for gDNA extraction with QIAamp DNA Micro Kit (Qiagen) following manufacture instructions. DNA from whole-blood samples was extracted with QIAamp DNA Micro Kit (Qiagen) following manufacture instructions. Flow cytometry data was collected using DiVa 8.0.1 Software. Subsequent analysis was performed with FlowJo_10.6.2. For sorting strategy, see *Figure 1—figure supplement 1*.

DNA library preparation and sequencing

DNA samples were submitted to the Integrated Genomics Operation (IGO) at MSKCC for quality and quantity analysis, library preparation and sequencing. DNA quality mas measured with Tapestation 2200. All samples had a DNA Integrity Number (DIN) >6. After PicoGreen quantification,~200 ng of genomic DNA were used for library construction using the KAPA Hyper Prep Kit (Kapa Biosystems KK8504) with eight cycles of PCR. After sample barcoding, 2.5 ng-1µg of each library were pooled and captured by hybridization with baits specific to either the HEME-PACT (Integrated Mutation Profiling of Actionable Cancer Targets related to Hematological Malignancies) assay, designed to capture all protein-coding exons and select introns of 576 (2.88Mb) commonly implicated oncogenes, tumor suppressor genes *Cheng et al., 2015* and/or HEME/BRAIN-PACT (716 genes, 3.44 Mb, *Supplementary file 2*) an expanded panel that included additional custom targets related to neurological diseases including, Alzheimer's Disease, Parkinson's Disease, Amyotrophic Lateral Sclerosis (ALS), and others (*Supplementary file 1*; *Bras et al., 2012*; *Renton et al., 2014*; *Karch et al., 2014*; *Karch and Goate, 2015*; *Turner et al., 2013*; *Ferrari et al., 2015*; *Kouri et al., 2015*; *Scholz and Bras, 2015*; *Nalls et al., 2014*). To simplify, in the manuscript, the combined panel is referred to as 'BRAIN-PACT.' In *Supplementary file 3*, 'Heme-only' or 'Brain-only' is indicated in the cases for which only one or the other panels were used. Capture pools were sequenced on the HiSeq 4000, using the HiSeq 3000/4000 SBS Kit (Illumina) for PE100 reads. Samples were sequenced to a mean depth of coverage of 1106 x (Control samples: 1071 x, AD samples 1100 x). For detailed information on the sample quality control checks used to avoid potential sample and/or barcode mix-ups and contamination from external DNA, see *Cheng et al., 2015*.

Mutation data analysis

The data processing pipeline for detecting variants in Illumina HiSeq data is as follows. First, the FASTQ files are processed to remove any adapter sequences at the end of the reads using cutadapt (v1.6). The files are then mapped using the BWA mapper (bwa mem v0.7.12). After mapping the SAM files are sorted and read group tags are added using the PICARD tools. After sorting in coordinate order the BAM's are processed with PICARD MarkDuplicates. The marked BAM files are then processed using the GATK toolkit (v 3.2) according to best practices for tumor normal pairs. They are first realigned using ABRA (v 0.92) and then the base quality values are recalibrated with the BaseQRecalibrator. Somatic variants are then called in the processed BAMs using MuTect (v1.1.7) for SNV and ShearwaterML (*Martincorena et al., 2015*; *Martincorena and Campbell, 2015*; *Martincorena et al., 2018*). **muTect (v1.1.7):** to identify somatic variants and eliminate germline variants, we run the pipeline as follows: PU.1, DN and Blood samples against matching-NeuN samples, and NeuN samples

against matching-PU.1. In addition, we ran all samples against a Frozen-Pool of 10 random genomes. We selected Single Nucleotide Variations (SNVs) [Missense, Nonsense, Splice Site, Splice Regions] that were supported by at least four or more mutant reads and with coverage of 50 x or more. Fill-out file for each project (~27 samples per sequencing pool), were used to exclude by manual curation, variants with high background noise. This resulted in 428 variants (Missense, Nonsense, Splice_site, Splice_Region).

ShearwaterML, was used to look for low allelic frequency somatic mutations as it has been shown to efficiently call variants present in a small fraction of cells with true positives being ~90%. Briefly, the basis of this algorithm is that is uses a collection of deep-sequenced samples to learn for each site a base-specific error model, by fitting a beta-binomial distribution to each site combining the error rates across all normal samples both the mean error rate at the site and the variation across samples, and comparing the observed variant rate in the sample of interest against this background model using a likelihood-ratio test. For detailed description of this algorithm please refer to *Martincorena et al., 2015*; *Martincorena et al., 2018*. In our data set, for each cell type (NeuN, DN, PU.1) we used as 'normal' a combination of the other cell types, i.e., PU.1 vs NeuN +DN, DN vs NeuN +PU.1, NEUN vs PU.1+DN, Blood vs NeuN +DN. Since all samples were processed and sequenced using the same protocol, we expect the background error to be even across samples. More than 400 samples were used as background leading to an average background coverage >400.000 x. Resulting variants for each cell type were filtered out as germline if they were present in more than 20% of all reads across samples. Additionally, variants with coverage of less than 50 x and more than 35% variant allelic frequency (VAF) were removed from downstream analysis. p-values were corrected for multiple testing using Benjamini & Hochberg's False Discovery Rate (FDR) *Reiner et al., 2003* and a q-value of cutoff of 0.01 was used to call somatic variants. Variants were required to have a least one supporting read in each strand. Somatic variants within 10 bp of an indel were filtered out as they typically reflect mapping errors. We selected Single Nucleotide Variations (SNVs) [Intronic, Intergenic, Missense, Nonsense, Splice Site, Splice Regions] that were supported by at least 4 or more mutant reads and annotated them using VEP. Finally, to reduce the risk of SNP contamination, we excluded variants with a MAF (minor allelic frequency) cutoff of 0.01 using the gnomeAD database. This resulted in 509 SNVs.

We compared the final mutant calls from Muetct1 and ShearwaterML and found that 30% of the events (111 variants) that were called by MuTect1 were also called by ShearwaterML. Overall a total of 826 variants (*Supplementary file 3*) were found, with a mean coverage at the mutant site of 668.3 X (10% percentile: 276 X, 90% percentile: 1181 X) and a mean of 29.1 mutant reads (10% percentile: 4, 90% percentile: 52), with 84% of mutated supported by at least five mutant reads (*Supplementary file 3*). The median allelic frequency was ~1.34% (*Supplementary file 3*). Negative results for matching brain negative samples were confirmed in 100% of samples at a mean depth of ~5000 x (range 648–23.000 x) (*Supplementary file 3*), confirming nuclei sorting purity of >95% for PU.1$^+$, DN, and NEUN$^+$ populations.

Validation of variants by droplet-digital-PCR (ddPCR)

We performed validation of ~11% of unique variants (69/760) by droplet-digital PCR (ddPCR) on pre-amplified DNA or on libraries (in the cases where DNA was not sufficient). Around 15% (15/69) of the variants analyzed by ddPCR were called by ShearwaterML, ~44% (34/69) were called by Mutect1 and 40% (24/69) by both ShearwaterML +Mutect1. Altogether we confirmed 62/69 of variants tested (~90%). In addition, 61 assays (from variants detected in PU.1+nuclei) were tested in paired cell types isolated from the same brain region. Assays were also run in matching blood when available. The mean depth of ddPCR was ~5000 x and mutant counts of three or more were considered positive. VAF obtained by ddPCR correlated with original VAF by sequencing (R2 0.93, p<0.0001). For **KRAS_G12D:** Bio-Rad validated assay (Unique Assay ID: dHsaMDV2510596) and **MTOR_Arg1616His_c.4847G>A:** Bio-Rad validated assay (Unique Assay ID: dHsaMDV2510596) were used. The remaining assays were designed and ordered through Bio-Rad. For setting-up the right conditions for newly designed assays, cycling conditions were tested to ensure optimal annealing/extension temperature as well as optimal separation of positive from empty droplets. All reactions were performed on a QX200 ddPCR system (Bio-Rad catalog # 1864001). When possible, each sample was evaluated in technical duplicates or quartets. Reactions contained 10 ng gDNA, primers and probes, and digital PCR Supermix for probes

(no dUTP). Reactions were partitioned into a median of ~31,000 droplets per well using the QX200 droplet generator. Emulsified PCRs were run on a 96-well thermal cycler using cycling conditions identified during the optimization step (95 °C 10'; 40–50 cycles of 94 °C 30' and 52–56°C 1'; 98 °C 10'; 4 °C hold). Plates were read and analyzed with the QuantaSoft software to assess the number of droplets positive for mutant DNA, wild-type DNA, both, or neither. ddPCR results are listed in *Supplementary file 3*.

Classification of variants

To classify somatic variants according to their pathogenicity we did as follows: Variants were classified as 'pathogenic (P-SNV)' if reported as 'pathogenic/likely pathogenic' by ClinVar *Landrum et al., 2014* and/or 'oncogenic/predicted oncogenic/likely oncogenic' by OncoKb (*Chakravarty et al., 2017*; *Supplementary file 3*). These two databases report pathogenicity in cancer and other diseases, based on supporting evidence from curated literature (see corresponding citations in *Supplementary file 3*). We considered classical-MAPK-pathway genes those reported to be mutated in RASopathies: *BRAF, CBL, KRAS, MAP2K1, NF1, PTPN11, SOS1, RIT1, SHOC2, NRAS, RAF1, RASA1, HRAS, MAP2K2, SPRED1* (*Rauen, 2013*; *Tidyman and Rauen, 2016* ; *Supplementary file 3*).

Quantification of mutational load and statistics

We defined mutational load or mutational burden as the number of synonymous and non-synonymous somatic single-nucleotide-variant (SNV) per megabase of genome examined (*Zehir et al., 2017*). Overall, a total of 826 SNVs were detected resulting in 0.3 mutations/Mb sequenced. As detailed in the manuscript, the mutational load varies considerably across cell types and patients. To quantify mutational load we took into consideration the panel used for sequencing each sample: HEME-PACT (2.88 Mb) or the extended panel BRAIN-PACT (3.44 Mb) (see *Supplementary file 2*). Therefore, the number of mutations was normalized by the number of Mb sequenced for that specific sample. In the cases where we calculated mutational load per patient, we averaged the mutational load of each sample from that patient for a given cell type (i.e. if for one patient, 2 PU.1 samples were sequenced, one from hippocampus and one from superior parietal cortex (with BRAIN-PACT) then the mutational load for PU.1 for that patient is the mean of the mutational load of the 2 PU.1 samples analyzed). For the quantification of 'pathogenic' variants, the same analysis is performed, quantifying only variants that are reported as pathogenic by ClinVar and/or OncoKb. Statistical significance was analyzed with GraphPad Prism (v9) and R (3.6.3). Non-parametric tests were used when data did not follow a normal distribution (Normality test: D'Agostino-Pearson and Shapiro-Wilk test). For normally distributed data, unpaired t-test was used to compare two groups and one-way, nested one-way or two-way analyses of variance (ANOVA) were used for comparing more than two groups, as indicated in the Figure legends. For data that did not have a normal distribution, the tests performed were unpaired two-tailed Mann-Whitney U test and Kruskal–Wallis test and Dunn's test for multiple comparisons. Pearson and Spearman were used for correlation analysis. In *Figure 2G*, we used multivariate logistic regression analysis to test if there was an association between Alzheimer's disease and the presence of pathogenic variants in PU.1$^+$ nuclei. We used Alzheimer's disease as a dependent variable, and age, sex, and the presence of pathogenic variant/s (Yes/No) as co-variates. In all the statistical tests, significance was considered at $p<0.05$. For Venn Diagram plots, we used (*Bardou et al., 2014*).

Mixed-effects modeling of somatic P-SNV burden

To evaluate the correlation between P-SNV burden and disease status (non-dementia controls and AD) after adjusting for other factors such as individual donor and age, we performed linear mixed-effects regression modeling using the nlme package in R. In the subsequent analysis, we also tested another framework implemented in the lme4 package and confirmed similar findings. We estimated the linear mixed model via maximum likelihood method with nlminb optimizer. The model included individual donor as a random effect. We also tested if inclusion of other co-variates (i.e. sex, anatomical location of the brain, and source biobank of brain samples) as random effects improved the overall model fitting via likelihood ratio test. However, none of these co-variates improved the model fitting ($p>0.99$). Thus, we used a relatively simple model that incorporated disease status and age as fixed effects and donor as random effect. The total explanatory power of the final model is substantial (conditional $R^2=0.48$). To assess the significance of age or disease status in predicting P-SNV burden,

we constructed another model that does not incorporate the variable as fixed effect, and compared the two models via likelihood ratio test. The variable was considered significantly associated with P-SNV burden when the p-value was below 0.05 and the AIC increased after removing the variable.

Pathway enrichment analysis of genes target of variants was performed using Metascape *Zhou et al., 2019* and the following ontology sources: KEGG Pathway (*Huang et al., 2009a*; *Huang et al., 2009b*), GO Molecular function (*Ashburner et al., 2000*; *The Gene Ontology, 2019*), Reactome Gene Sets *Ashburner et al., 2000*; *The Gene Ontology, 2019*, and Canonical Pathways (*Schaefer et al., 2009*). The list of 716 genes from the targeted panel were used as the enrichment background. Terms with a p-value <0.05, a minimum count of 3, and an enrichment factor >1.5 (the enrichment factor is the ratio between the observed counts and the counts expected by chance) are shown. p-values are calculated based on the cumulative hypergeometric distribution *Zar, 2010*.

Expression of target genes in microglia

To evaluate the expression levels of the genes identified in this study as target of somatic variants, we consulted a publicly available database (https://www.proteinatlas.org/), and also plotted their expression as determined by RNAseq in two studies (Galatro et al. GSE99074, *Galatro et al., 2017*, and *Gosselin et al., 2017*) (Supplementary Fig S3 and Figure S2). For data from Galatro et al. (GSE99074, *Galatro et al., 2017*), normalized gene expression data and associated clinical information of isolated human microglia (N=39) and whole brain (N=16) from healthy controls were downloaded from GEO. For data from *Gosselin et al., 2017* , raw gene expression data and associated clinical information of isolated microglia (N=3) and whole brain (N=1) from healthy controls were extracted from the original dataset. Raw counts were normalized using the DESeq2 package in R (*Love et al., 2014*).

Nuclei isolation from frozen brain samples for snRNA-seq

For snRNA-seq studies, we only selected samples with a RIN score in whole tissue of six or more. All samples were handled and processed under Air Clean PCR Workstation. About 250–400 mg of frozen brain tissues were homogenized with a sterile Dounce tissue grinder using a sterile homogenization buffer to isolate cell nuclei (250 mM Sucrose, 25 mM KCL, 5 mM MgCl2, 10 mM Tris buffer pH 8.0, 0.1% (v/v) Triton X-100, 3 μM DAPI, Nuclease Free Water and 20 U/ml of Superase-In RNase inhibitor, and 40 U/ml RNasin ribonuclease inhibitor). Homogenate was filtered in a 40 μm cell strainer and centrifuged 800 g 8 min 4 °C. To clean-up the homogenate, we performed a iodixanol density gradient centrifugation as follows: pellet was gently mixed 1:1 with iodixanol medium at 50% (50% Iodixanol, 250 mM Sucrose, 150 mM KCL, 30 mM MgCl2, 60 mM Tris buffer pH 8.0, Nuclease Free Water) and homogenization buffer. This solution layered to a new tube containing equal volume of iodixanol medium at 29% and centrifuged 13.500 g for 20 min at 4 °C. Nuclei pellet was resuspended in FACS buffer with RNAse inhibitors (0.5% BSA, 2 mM EDTA, Superase-In RNase inhibitor and 40 U/ml RNasin ribonuclease inhibitor) and centrifuged 800 g 5 min, 4 °C. Nuclei pellet was fixed with 90% ice-cold methanol and incubated for 10 min on ice, followed by a centrifugation at 1300 g (without brakes, which improves with nuclei recovery after fixation). The pellet was resuspended in permeabilization buffer (6% BSA, Superase-In RNase inhibitor 20 U/mL, RNasin ribonuclease inhibitor 40 U/mL and 0.05% Triton) followed by a centrifugation at 1300 g. Sample was incubated with anti-Pu.1 antibody (microglia marker 1:50, Pu.1-AlexaFluor 647, 9G7 Cell Signaling) in permeabilization buffer. After a wash with FACS buffer sample were ready for sorting. Nuclei are FACS-sorted in a BD FACS Aria with a 100 μm nozzle and a sheath pressure 20 psi, operating at ~1000 events per second. Nuclei were sorted into 1.5 ml certified RNAse, DNAse DNA, ATP, and Endotoxins tubes containing 100 μl of sterile PBS. For each population, we sorted >10^5 nuclei into FACS buffer.

SnRNA-seq library preparation and sequencing

The single-nuclei RNA-seq of FACS-sorted nuclei suspensions was performed on Chromium instrument (10 X genomics) following the user guide manual (Reagent Kit 3' v3.1). Each sample, containing approximately 10,000 nuclei at a final dilution of ~1000 cells/μl was loaded onto the cartridge following the manual. The individual transcriptomes of encapsulated cells were barcoded during RT step and resulting cDNA purified with DynaBeads followed by amplification per manual guidelines. Next, PCR-amplified product was fragmented, A-tailed, purified with 1.2 X SPRI beads, ligated to the sequencing adapters, and indexed by PCR. The indexed DNA libraries were double-size purified (0.6–0.8 X) with

SPRI beads and sequenced on Illumina NovaSeq S4 platform (R1 – 26 cycles, i7 – 8 cycles, R2 – 70 cycles or higher). Sequencing depth was ~200 million reads per sample on average. FASQ files were processed using SEQC pipeline *Azizi et al., 2018* for quality control, mapping to GRCH38 reference genome, and log2 transformation of the data with the default SEQC parameters to obtain the gene-cell count matrix.

SnRNA-seq analysis

Seurat v4.0.3 with default parameters was used to perform sctransform (SCT) normalization, integration and Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction. The FindClusters function was used for cell clustering. To improve clustering, all samples were analyzed in an integrated analysis, based on canonical correlation analysis (CCA). Cell types were annotated using the top 500 DEGs of each cell type in a human cortex database. Data can be accessed at https://weillcornellmed.shinyapps.io/Human_brain/. The removal of doublets using DoubletFinder and cells with high mitochondrial content (>10% mitochondrial RNA) yielded between 6437 and 9241 nuclei per patient and sample. Microglia represented 94 ± 3% of total cells. Unique Molecular Identifiers (UMIs) per nucleus and gene count per nucleus were comparable between donors. Integrated_snn at resolution 0.2 outlined 16 microglia clusters. Except for cluster 13 consisting of 97% of cells from the healthy control Control 11_AG, all donors and samples were represented in every cluster. One cluster contained few cells (0.84% of total microglia, for an average of 6.20 ± 1.60% for other clusters) and was marked by a low number of cluster-enriched genes and was excluded from further analyses. For pathway enrichment analysis, genes were pre-ranked using differential expression analysis in SCANPY *Wolf et al., 2018* with Wilcoxon rank-sum method. Statistical analysis were performed using the fgsea-Multilevel function in fgsea R package *Korotkevich et al., 2019* for HALLMARK and KEGG pathways. Gene sets with p-value <0.05 and adjusted p-value <0.25 were selected and visualized using ggpubr and ggplot2 *Hadley, 2016* R package. For the variant analysis of AD52_HIP harboring a KRAS^{A59G} (c.176C>G) clone, Integrative Genomics Viewer (IGV) software was used to display sequencing reads at KRAS c.176C (exon 3; GRCh38 chr12:25,227,348). Cells within each cluster were identified based on the 16-digit barcodes from SEQC-aligned reads. Barcodes were converted to the 10 X Genomics format and used to sample reads from each cluster within the original BAM file. BAM subsets for each cluster were read with IGV and reads with identical UMIs were filtered out to account for amplification bias.

Whole-exome-sequencing and analysis

Remaining libraries from a selected group of PU.1 and NEUN samples sequenced with BRAIN-PACT (see above) were sequenced by Whole-Exome-Sequencing (WES). Matching NEUN samples were sequenced to extract the germline variants. Around 100 ng of library were captured by hybridization using the xGen Exome Research Panel v2.0 (IDT) according to the manufacturer's protocol. PCR amplification of the post-capture libraries was carried out for 12 cycles. Samples were run on a NovaSeq 6000 in a PE100 run, using the NovaSeq 6000 S4 Reagent Kit (200 Cycles) (Illumina). Samples were covered to an average of 419 X. The data processing pipeline for detecting variants in Novaseq data is as follows. First the FASTQ files are processed to remove any adapter sequences at the end of the reads using cutadapt (v1.6). The files are then mapped using the BWA mapper (bwa mem v0.7.12). After mapping the SAM files are sorted and read group tags are added using the PICARD tools. After sorting in coordinate order the BAM's are processed with PICARD MarkDuplicates. The marked BAM files are then processed using the GATK toolkit (v 3.2) according to the best practices for tumor normal pairs. They are first realigned using ABRA (v 0.92) and then the base quality values are recalibrated with the BaseQRecalibrator. Somatic variants are then called in the processed BAMs using muTect (v1.1.7) for SNV and the Haplotype caller from GATK with a custom post-processing script to call somatic indels. The full pipeline is available here https://github.com/soccin/BIC-variants_pipeline (copy archived at *Socci, 2022*) and the post processing code is at https://github.com/soccin/Variant-PostProcess, (copy archived at *Socci, 2025*). We selected Single Nucleotide Variants (SNVs) [Missense, Nonsense, Splice Site, Splice Regions] that were supported by at least 8 or more mutant reads, variant allelic frequency above 5% and with coverage of 50 x. Annotation was performed using VEP. Finally, to reduce the risk of SNP contamination, we excluded variants with a MAF (minor allelic frequency) cutoff of 0.01 using the genomeAD database. Variants were classified as 'candidate pathogenic' when

SNV is predicted to affect the protein as determined by PolyPhen-2 (possibly and probably damaging) and SIFT (deleterious) and CADD-MSC (high) and FATHMM-XF Functional Analysis through Hidden Markov Models (pathogenic) (*Rogers et al., 2018*; *Supplementary file 5*).

Cell lines

HEK293T cell culture and transfection

HEK 293T cells (ATCC) were maintained in Dulbecco's modified Eagle's medium (Mediatech, Inc) supplemented with 10% fetal bovine serum (Sigma) and 1000 IU/ml penicillin, 1000 IU/ml streptomycin.

BV2 microglial cell line

BV2 murine microglial cells were cultured in Dulbecco's modified Eagle's medium (DMEM) High Glucose medium (Gibco), Glutamax (Gibco), sodium pyruvate, 1% non-essential amino acids (Invitrogen), and 10% heat-inactivated fetal bovine serum (FBS, EMD Millipore). For MAPK activation experiments, cells were treated with M-CSF1 100 ng/ml for 5 min.

MAC cell lines

Mouse primary CSF-1 dependent macrophages immortalized with the SV-U19-5 retrovirus *Xiong et al., 2011* were a gift of Dr. E. R. Stanley (Albert Einstein College of Medicine, Bronx, NY). They were cultured in RPMI 1640 medium with Glutamax (Gibco), 10% heat-inactivated fetal bovine serum (FBS, EMD Millipore) and 100 ng/mL recombinant CSF-1 (gift from Dr. E. R. Stanley). Growth medium was renewed every second day. When confluency reached 80%, cells were passaged by cell scraping and plated at 5×10^4 cells/cm² in tissue culture treated plates. For signaling pathway analyses, cell proliferation assays or collection for RNA sequencing, cells were plated 1 d prior at 5×10^4 cells/cm² in medium containing 10 ng/mL CSF-1 for lines expressing wild-type (WT) and mutant CBL, RIT1, and KRAS proteins and 100 ng/mL CSF-1 for lines expressing WT and mutant PTPN11 proteins. Cells were grown at 37 °C and 5% CO2.

Human-induced pluripotent stem cell (hiPSC) culture

hiPSC lines were derived from peripheral blood mononuclear cells (PBMCs) of a healthy donor. Written informed consent was obtained according to the Helsinki convention. The study was approved by the Institutional Review Board of St Thomas' Hospital; Guy's hospital; the King's College London University; the Memorial Sloan Kettering Cancer Center and by the Tri- institutional (MSKCC, Weill-Cornell, Rockefeller University) Embryonic Stem Cell Research Oversight (ESCRO) Committee. hiPSC were derived using Sendai viral vectors (Thermo Fisher Scientific; A16517). Newly derived hiPSC clones were maintained in culture for 10 passages (2–3 mo) to remove any traces of Sendai viral particles. Over 90% of hiPSCs in the derived lines expressed high levels of the pluripotency markers NANOG and OCT4 by flow cytometry. The C12 hiPSC WT line was engineered to carry a CBL p.C404Y, c.1211G>A heterozygous variant at the endogenous CBL locus. HiPSCs of passage 25–35 were cultured on confluent irradiated CF1 mouse embryo fibroblasts (MEFs, Gibco) in hiPSC medium consisting of knock-out DMEM (Invitrogen), 10% knock-out-Serum Replacement (Invitrogen), 2 mM L-glutamine (Gibco), 100 U/mL penicillin-streptomycin (Invitrogen), 1% non-essential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (R&D). hiPSC medium was supplement with 10 ng/mL bFGF (PeproTech) and changed every second day. Two days before culture with hiPSCs, MEFs were plated at 20,000 cells/cm² in DMEM supplemented with 10% heat-inactivated fetal bovine serum (FBS, EMD Millipore), 100 U/mL penicillin- streptomycin (Invitrogen), 1% non-essential amino acids (Invitrogen) and 0.1 mM β-mercaptoethanol (R&D Systems) on 150 mm tissue culture plates coated with 0.1% gelatin (Sigma). hiPSCs were passaged weekly with 250 U/mL collagenase type IV (Thermo Fisher Scientific) at a 1:4 to 1:6 ratio onto MEF cells in hiPSC medium supplemented with 10 µM Rock inhibitor (Y-27632 dihydrochloride, Sigma). Cells were maintained at 37 °C and 5% CO2 and they were routinely tested for mycoplasma and periodically assessed for genomic integrity by karyotyping. Microglia-like cells were obtained from hiPSCs using an embryoid body (EB)-based protocol as previously described *Lachmann et al., 2015*. Briefly, hiPSC were loosened with 250 U/mL collagenase type IV (Thermo Fisher Scientific) and lifted with cell scraping. For EBs formation, hiPSC colonies were transferred to suspension plates on an orbital shaker in hiPSC medium supplemented with 10 µM Rock

inhibitor (Y-27632 dihydrochloride, Sigma). After 6 d, EBs were transferred to 6-wells tissue culture treated plates in STEMdiff APEL 2 medium (Stem Cell Technology) with 5% Protein Free Hybridoma Media (Gibco), 100 U/mL penicillin-streptomycin (Invitrogen), 25 ng/mL IL-3 (Peprotech) and 50 ng/mL CSF-1 (Peprotech). Microglia-like cells were harvested every week from the supernatant of EBs cultures. Collected microglia-like cells were used immediately for signaling pathway analyses or plated for 6–7 d in RPMI 1640 medium with Glutamax supplement (Gibco), 10% heat-inactivated fetal bovine serum (FBS, EMD Millipore) and 100 ng/mL human recombinant CSF-1 (Peprotech) in tissue culture plates for cytology, flow cytometry, RNA sequencing, and supernatant analyses of cytokines release. Microglia like cells differentiation was monitored by May-Grunwald Giemsa staining and flow cytometry analyses of myeloid markers.

All cells used in the study were routinely tested for mycoplasma and were negative.

Plasmids used in in-vitro studies (HEK293, BV2, and MAC lines)

The expression vectors for Flag-tagged CHK2 kinase and RIT1 were from Sino Biological and Origene, respectively. The vector encoding pcDNA3-HA-tagged c-Cbl was a kind gift from Dr. Nicholas Carpino (Stony Brook). RIT1^{M90I,} RIT1^{F82L}, CBL^{I383M}, CBL^{C404Y}, CBL^{C416S}, CBL^{C384Y}, CBL^{Y371H} were generated by site-directed mutagenesis using the QuikChange Kit (Agilent). pHAGE_puro was a gift from Christopher Vakoc (Addgene plasmid # 118692; http://n2t.net/addgene:118692; RRID:Addgene_118692) (*Lu et al., 2018*). pHAGE-KRAS was a gift from Gordon Mills & Kenneth Scott (Addgene plasmid # 116755; http://n2t.net/addgene:116755; RRID:Addgene_116755) (*Ng et al., 2018*) *Cheng et al., 2015*. pHAGE-PTPN11 was a gift from Gordon Mills & Kenneth Scott (Addgene plasmid # 116782; http://n2t.net/addgene:116782; RRID:Addgene_116782) (*Ng et al., 2018*). pHAGE-PTPN11-T73I was a gift from Gordon Mills & Kenneth Scott (Addgene plasmid # 116647; http://n2t.net/addgene: 116647; RRID:Addgene_116647) (*Ng et al., 2018*). pDONR223_KRAS_p.A59G was a gift from Jesse Boehm & William Hahn & David Root (Addgene plasmid # 81662; http://n2t.net/addgene:81662; RRID:Addgene_81662) (*Kim et al., 2016*), Phage-CBL, Phage-CBL^{I383M}, Phage-CBL^{C404Y}, Phage-CBL^{C416S}, Phage-RIT1, Phage-RIT1^{M90I} and Phage-RIT1^{F82L} and Phage-KRAS^{A59G} were generated by Azenta Life Sciences via a PCR cloning approach. pHAGE-CBL^{C384Y} plasmid was generated at Azenta Life Science by targeted mutagenesis of pHAGE-CBL.

Generation of mutant lines

HEK293 were transfected 24 hr after plating with 2.5 µL of Mirus Transit LT1 per µg of DNA. Cells were harvested and lysed 48 hr after transfection using a buffer containing 25 mM Tris, pH 7.5, 1 mM EDTA, 100 mM NaCl, 1% NP-40, 10 µg/ml leupeptin, 10 µg/ml aprotinin, 200 µM PMSF, and 0.2 mM Na$_3$VO$_4$. For EGF stimulation, the media was replaced 24 hr after transfection with DMEM containing 1% FBS and antibiotics. After a further 24 hr in this starvation media, the cells were stimulated with 50 ng/ml EGF for 5 min at 37 °C.

Lentiviral production and transduction of BV2 and MAC cell lines

For BV2 cell line, cells were transduced for 24 hr without the presence of Vpx VLPs and selected with 2.5 µg/mL puromycin (Fisher Scientific). For 'MAC' lines, Vpx-containing virus-like particles (Vpx VLPs) were produced by transfection of HEK293T cells with 4.8 µg VSV-g plasmid and 31.2 µg pSIV3/Vpx plasmid, a gift from Dr. M. Menager (Imagine Institute, Paris, France) using TransIT-293 Transfection Reagent (Mirus Bio, Fisher Scientific). Forty-eight hours after transfection, the supernatant containing Vpx VLPs was collected and used immediately for lentiviral transduction of macrophages. Viral supernatants were obtained by transfection of HEK293T cells using X-tremeGENE HP DNA Transfection Reagent (Sigma). Packaging vectors used were psPAX2 (gift from Didier Trono Addgene plasmid # 12260; http://n2t.net/addgene:12260; RRID:Addgene_12260) and pMD2.G (gift from Didier Trono, Addgene plasmid # 12259; http://n2t.net/addgene:12259; RRID:Addgene_12259). Cells were transduced for 24 hr in presence of Vpx VLPs. Transduced macrophages were selected with 5 µg/mL puromycin (Fisher Scientific).

Generation of the CBL[+/C404Y] and isogenic WT hiPSC lines

The CBL[C404Y] (c.1211 G>A) variant was inserted at the endogenous locus in the C12 WT hiPSC using Cytidine base editing (CBE) with CBE enzyme BE3-FNLS (*Zafra et al., 2018*). Briefly, the sgRNA for CBE was designed to target the non-coding strand and introduce the position 6 'C-to-T' conversion, to create the G-to-A conversion on the coding strand. The sgRNA target sequence was cloned into the pSPgRNA (Addgene plasmid # 47108) *Perez-Pinera et al., 2013* to make the gene targeting construct. To introduce the CBL C404Y variants, the WT hiPSC (C12) were dissociated using Accutase (Innovative Cell Technologies) and electroporated (1×10^6 cells per reaction) with 4 µg sgRNA-construct plasmid and 4 µg CBE enzyme coding vector BE3-FNLS (Addgene plasmid # 112671) (*Zafra et al., 2018*) using Lonza 4D-Nucleofector and the Nucleofector solution (Lonza V4XP-3034) following our previously reported protocol (*Zhong et al., 2020*). The cells were then seeded, and 4 d later, the hiPSC were dissociated into single cells by Accutase and re-plated at a low density (four per well in 96-well plates) to get the single-cell clones. Ten days later, individual colonies were picked, expanded and analyzed by PCR and DNA sequencing to identify the clones carried the desired CBL[C404Y] heterozygous variant and the isogenic WT control clones. The sgRNA target, PCR, and sequencing primers are listed below.

	sgRNA target	PCR-Forward primer	PCR-Reverse primer (used for sequencing)
CBL[C404Y]	TAAGACAGGATGTGCACATG	TGGGCTCCACATTCCAACTA	GCCCTGACCTTCTGATTCCT

Western blotting

For HEK293 cells, lysates were resolved by SDS-PAGE, transferred to PVDF membranes, and probed with the appropriate antibodies. Horseradish peroxidase-conjugated secondary antibodies (GE Healthcare) and Western blotting substrate (Thermo) were used for detection. For anti-Cdc42 immunoprecipitation experiments, cell lysates (1 mg total protein) were incubated overnight with 1 µg of anti-Cdc42 antibody (Santa Cruz) and 25 µL of protein A agarose (Roche) at 4°C. Anti-Flag immunoprecipitations were done with anti-Flag M2 affinity resin (Sigma). The beads were washed three times with lysis buffer, then eluted with SDS-PAGE buffer, and resolved by SDS-PAGE. The proteins were transferred to PVDF membrane for Western blot analysis. Antibodies used are Phospho-p44/42 MAPK (pErk 1/2) (Thr202/Tyr204) is from Cell Signaling #4370, total p44/42 MAPK (Erk1/2) is from Cell Signaling #9102, HA tag from Millipore # 05–904, Flag antibody is from Sigma (#A8592), pCHEK2 (T383) antibody is from Abcam, #ab59408, Cdc42 antibody is from Santa Cruz (#sc87) and Anti-γ-Tubulin antibody (Sigma T6557). **For Immunoprecipitation Kinase assay in HEK293T cells,** cell lysates (1 mg protein) were incubated overnight with 30 µL of anti-Flag M2 affinity resin on a rotator at 4°C, then washed three times with Tris-buffered saline (TBS). A portion of each sample was eluted with SDS-PAGE sample buffer and analyzed by anti-Flag Western blotting. The remaining sample was used for a radioactive kinase assay. The immunoprecipitated proteins were incubated with 25 µL of reaction buffer (30 mM Tris, pH 7.5, 20 mM MgCl$_2$, 1 mg/mL BSA, 400 µM ATP), 650 µM CHKtide peptide (KKKVRSGLYRSPSMPENLNRPR, SignalChem), and 50–100 cpm/pmol of [γ *Zar, 2010*-P] ATP at 30 °C for 15 min. The reactions were quenched using 45 µL of 10% trichloroacetic acid. The samples were centrifuged and 30 µL of the reaction was spotted onto Whatman P81 cellulose phosphate paper. After washing with 0.5% phosphoric acid, incorporation of radioactive phosphate into the peptide was measured by scintillation counting. **For MAC lines and hiPSC-derived cells,** cell lysates obtained with RIPA buffer +1:1000 Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific) were sonicated three times for 30 s at 4 °C (Bioruptor, Diagenode). Protein quantification of supernatant was done with Precision Red Advanced Protein Assay (Cytoskeleton). Proteins were boiled for 5 min at 95 °C in NuPAGE LDS sample buffer (Invitrogen) and separated in NuPAGE 4–12% Bis-Tris Protein Gel (Invitrogen) in NuPAGE MES SDS Running Buffer (Invitrogen). Electrophoretic transfer to a nitrocellulose membrane (Thermo Fisher Scientific) was done in NuPAGE Transfer Buffer (Invitrogen). Blocking was performed for 60 min in TBS-T +5% nonfat milk (Cell Signaling) and incubated with primary antibodies at 4°C: rabbit anti-p44/42 MAPK (ERK1/2) (Cell Signaling; 1:1000); rabbit anti-P-p44/42 MAPK (Cell Signaling, 1:1000); rabbit anti-c-CBL (Cell signaling, 1:1000); rabbit anti-RIT1 (Abcam, 1:1000); mouse anti-KRAS (clone 3B10-2F2, Sigma, 1 µg/mL); mouse anti-Actin (clone MAB1501, Sigma, 1:10,000). Primary antibodies were detected using the secondary anti-rabbit

IgG HRP-linked (Cell Signaling, 1:1000) or the anti-mouse IgG HRP-linked (Cell Signaling, 1:1000) were used to detect primary antibodies, with SuperSignal West Femto Chemiluminescent Substrate (Thermo Fisher Scientific) using a ChemiDoc MP Imaging System (Bio-Rad). pERK/ERK ratios were measured with ImageJ software.

DNA/RNA isolation, dd-PCR, and RTqPCR in MAC lines and hiPSC-derived cells

Genomic DNA was extracted using QIAamp DNA Micro Kit (50) (Qiagen), following the manufacturer's instructions. Total RNA was extracted using RNeasy Mini kit (Qiagen), following the manufacturer's instructions. cDNA was generated by reverse transcription using Invitrogen SuperScript IV Reverse Transcriptase (Invitrogen) with oligo(dT) primers. The TaqMan gene expression assays used were c-CBL FAM (Hs01011446_m1), CBLb FAM (Hs00180288_m1) and GAPDH VIC (Hs02786624_g1) (Thermo Fisher Scientific). RT-qPCR was performed using Applied Biosystems TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific) and a QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific). The results were normalized to GAPDH. For droplet PCR analyses, assays specific for the detection of I383M, C384Y, C404Y and C416S in CBL and F82L and M90I in RIT1, A59G in KRAS, and corresponding WT sequences (listed below) were obtained from Bio-Rad. Cycling conditions were tested to ensure optimal annealing/extension temperature as well as optimal separation of positive from empty droplets. Optimization was done with a known positive control. After PicoGreen quantification, 2.6–9 ng gDNA or cDNA were combined with locus-specific primers, FAM- and HEX-labeled probes, HaeIII, and digital PCR Supermix for probes (no dUTP). All reactions were performed on a QX200 ddPCR system (Bio-Rad catalog # 1864001) and each sample was evaluated in technical duplicates. Reactions were partitioned into a median of ~19,000 droplets per well using the QX200 droplet generator. Emulsified PCRs were run on a 96-well thermal cycler using cycling conditions identified during the optimization step (95 °C 10'; 40 cycles of 94 °C 30' and 52–55°C 1'; 98 °C 10'; 4 °C hold). Plates were read and analyzed with the QuantaSoft software to assess the number of droplets positive for mutant or wild-type DNA.

Assay name	Assay ID
CBL_I383M	dHsaMDS675699482
CBL_C384Y	dHsaMDS386449640
CBL_C404Y	dHsaMDS437459772
CBL_mRNA_C404Y	dMDS334857054
CBL_C416S	dHsaMDS613275900
RIT1_ F82L	dMDS959028273
RIT1_M90I	dHsaMDS133045056
KRAS_A59G	dHsaMDS581417660

Flow cytometry analyses for surface antigens CSF1-R, CD11b, MRC1, α5β3, CD11c, Tim4, HLA-DR, CD45, CD14, NGFR, EGFR, CD36, and SIRPα were performed using PE-conjugated anti-CD115 (CSF1-R) (clone 9-4D2, BD Pharmingen), PE/Cy7-conjugated anti-CD11b (clone ICRF44, Biolegend), Alexa Fluor 488-conjugated anti-CD206 (MRC1) (clone 19.2, Thermo Fisher Scientific), PE-conjugated anti-integrin α5β3 (clone 23C6, R&D systems), PE/Cy5-conjugated anti-CD11c (Clone B-ly6, BD Pharmigen), APC-conjugated anti-Tim4 (Clone 9F4, BioLegend), PE/Cy7-conjugated anti-HLA-DR (clone G46-6, BD Pharmigen), BV650-conjugated anti-CD45 (clone HI30, BD Horizon), APC/Cy7-conjugated anti-CD14 (clone M5E2, Biolegend), PE-conjugated anti-NGFR (clone ME20.4, eBioscience), Alexa Fluor 647-conjugated anti-EGFR (clone EGFR.1, BD Pharmigen), APC/Cy7-conjugated anti-CD36 (clone 5-271, BioLegend), and APC-conjugated anti-CD172a (SIRPα) (Clone: 15 414, Thermo Fisher Scientific) antibodies. Iba1 expression was detected following fixation and permeabilization of macrophages using BD Cytofix/Cytoperm solution (BD Pharmigen). Cells were marked with Zombie Violet Viability (Biolegend). After incubation with FcR Blocking Reagent (Miltenyi Biotec), cells were stained with Alexa Fluor 555-conjugated anti-Iba1 antibody (clone E4O4W, Cell Signaling). Flow cytometry was performed using a BD Biosciences

LSR Fortessa flow cytometer with Diva software. Data were analyzed using FlowJo (BD Biosciences LLC).

Cell proliferation analyses

For hiPSC-derived cells, cell suspension was filtered through a 100 µm nylon mesh (Corning) and marked with Zombie Violet Viability (Biolegend). After incubation with FcR Blocking Reagent (Miltenyi Biotec), surface receptors were labeled with PE/Cy7-conjugated anti-CD11b (clone ICRF44, Biolegend), Alexa Fluor 488-conjugated anti-CD206 (MRC1) (clone 19.2, Thermo Fisher Scientific), BV650-conjugated anti-CD45 (clone HI30, BD Horizon), APC/Cy7-conjugated anti-CD14 (clone M5E2, Biolegend) prior to EdU detection. For proliferation studies in the mouse macrophage cell lines, macrophages were incubated with 10 µM EdU (Thermo Fisher Scientific) for 2 hr at 37 °C and collected by cell scraping and marked with Zombie Violet Viability (Biolegend) prior to EdU detection. EdU detection was performed using the Click-iT Plus EdU Alexa Fluor 647 Flow Cytometry Assay Kit (Thermo Fisher Scientific), following the manufacturer's instructions. hiPSC-derived macrophages were analyzed using a BD Biosciences Aria III cell sorter and macrophages were identified as CD11b$^+$CD45$^+$CD14$^+$MRC1$^+$. The macrophage cell lines were analyzed using a BD Biosciences LSR Fortessa flow cytometer. Data were analyzed using FlowJo 10.6 (BD Biosciences LLC).

Enzyme-linked immunosorbent assay

Supernatants of iPSC-derived microglia-like cells were analyzed for human inflammatory cytokines IL-6, TNFα, IL-1β, IFNγ and for the complement C3 and complement Factor H by Enzyme-linked immunosorbent assay (ELISA) at Eve Technologies (Calgary, AB).

Bulk RNA sequencing (RNAseq)

Three biological replicates were processed for each condition/cell line. In view of RNA sequencing, phase separation in cells lysed in 1 mL TRIzol Reagent (Thermo Fisher Scientific) was induced with 200 µL chloroform and RNA was extracted from the aqueous phase using the miRNeasy Mini Kit (Qiagen) on the QIAcube Connect (Qiagen) according to the manufacturer's protocol with 350 µL input, or using the MagMAX mirVana Total RNA Isolation Kit (Thermo Fisher Scientific) on the King-Fisher Flex Magnetic Particle Processor (Thermo Fisher Scientific) according to the manufacturer's protocol with 350 µL input. Samples were eluted in 30 µL RNase-free water. After RiboGreen quantification and quality control by Agilent BioAnalyzer, 231–500 ng of total RNA with RIN values of 9.4–10 underwent polyA selection and TruSeq library preparation according to instructions provided by Illumina (TruSeq Stranded mRNA LT Kit, Illumina), with 8 cycles of PCR. Samples were barcoded and run on a NovaSeq 6000 in a PE100 run, using the NovaSeq 6000 S4 Reagent Kit (200 Cycles) (Illumina). An average of 90 million paired reads was generated per sample. Ribosomal reads represented 0–1.6% of the total reads generated and the percent of mRNA bases averaged 79%.

Bulk RNAseq analysis

FastQ files of 2×100 bp paired-end reads were quality checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2012). Samples with high quality reads (Phred score ≥ 30) were aligned to the *Mus musculus* genome (GRCm38.80) for the MAC lines or *Homo sapiens* (assembly GRCh38.p14) for the IPSCs lines using STAR aligner. For the MAC lines, we computed the expression count matrix from the mapped reads using HTSeq (www-huber.embl.de/users/anders/HTSeq) and one of several possible gene model databases. The raw count matrix generated by HTSeq are then be processed using the R/Bioconductor package DESeq (www-huber.embl.de/users/anders/DESeq) which is used to both normalize the full dataset and analyze differential expression between sample groups. For the ISPCs line dataset, gene quantification was performed using feature counts from the Subread package in R. Gene expression levels were normalized and log2 transformed using the Trimmed Mean of M-values (TMM) method and differential expression analysis was performed using the edgeR package in R. For hiPSC derived cells, GSEA (Hallmark, KEGG, GO, REACTOME) were performed using the fgsea package in R on a pre-ranked list (formula: sign(ogFC) * -log10(PValue)) on all expressed genes in the dataset. For the MAC cell lines dataset, GSEA was performed using gsea4.3.2 for KEGG and HALLMARK canonical pathways in MSigDB v 7.5.1. Significant genesets were selected based on an FDR ≤ 0.25.

For lists of differentially expressed genes, genes were selected with controlled False Positive Rate (B&H method) at 5% (FDR ≤ 0.05). Genes were considered upregulated/downregulated for log2 fold change >1.5 or <−1.5.

Statistical analysis

Statistical methods are detailed in the corresponding sections above (Quantification of mutational load and statistics, Bulk RNAseq analysis, SnRNA-seq analysis) and in the Fig. legends. p-values of 0.05 and adj. p-values (FDR) of 0.25 are considered significant unless otherwise specified.

Code availability

All code used in this study has been previously published as referenced in the method section above.

Acknowledgements

Additional information

Author contributions

Rocio Vicario, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Methodology, Writing – original draft, Project administration, Writing – review and editing; Stamatina Fragkogianni, Nicholas D Socci, Data curation, Formal analysis, Methodology; Leslie Weber, Data curation, Formal analysis, Investigation; Tomi Lazarov, Conceptualization, Data curation, Formal analysis, Investigation; Yang Hu, Samantha Y Hayashi, Barbara

Craddock, Masato Ogishi, Bertrand Boisson, Olivier Elemento, Jean-Laurent Casanova, Data curation, Formal analysis; Araitz Alberdi, Formal analysis, Project administration; Ann Baako, Oyku Ay, Ting Zhou, Methodology; Estibaliz Lopez-Rodrigo, W Todd Miller, Formal analysis, Investigation, Methodology; Rajya Kappagantula, Christine A Iacobuzio-Donahue, Netherlands Brain Bank, Resources; Agnes Viale, Resources, Methodology; Richard M Ransohoff, Richard Chesworth, Funding acquisition, Investigation; Omar Abdel-Wahab, Investigation; Frédéric Geissmann, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Funding acquisition, Investigation, Methodology, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Rocio Vicario ![ORCID] https://orcid.org/0000-0002-7894-5261
Tomi Lazarov ![ORCID] https://orcid.org/0000-0002-6312-0080
Richard M Ransohoff ![ORCID] https://orcid.org/0000-0003-0175-6910
Omar Abdel-Wahab ![ORCID] https://orcid.org/0000-0002-3907-6171
Frédéric Geissmann ![ORCID] https://orcid.org/0000-0001-5029-2468

Ethics

Tissue samples. The study was conducted according to the Declaration of Helsinki. Human tissues were obtained with patient-informed consent and used under approval by the Institutional Review Boards from Memorial Sloan Kettering Cancer Center (IRB protocols #X19-027). Snap-frozen human brain and matched blood were provided by the Netherlands Brain Bank (NBB), the Human Brain Collection Core (HBCC, NIH), Hospital Sant Joan de Déu and the Rapid Autopsy Program (MSKCC, IRB #15-021). Samples were neuropathologically evaluated and classified by the collaborating institutions as Alzheimer's disease (AD) 1-5 or non-dementia controls. The mean age of AD patients is 65 years old (55.5% female, 44.5% male). The mean age of all controls is 54 years old (60% female, 40% male), and the mean age of AD age-matched controls was 70 years old (60% female , 40% male). The overall mean of the post-mortem delay interval was 9.8 hours. Patients did not present with germline pathogenic PSEN1/2/3 or APP AD's associated variants. For additional information on donor's brain regions, sex, age, cause of death, Apoe status, Braak status, see Supplementary file 1. To avoid possible contamination of sequencing data with mutations associated with donor's tumoral disease in the group of non-dementia controls, we refrained from selecting cases with blood malignances or with brain tumors. Samples from histiocytosis patients were collected under GENE HISTIO study (approved by CNIL and CPP Ile-de France) from Pitié-Salpêtrière Hospital and Hospital Trousseau and from Memorial Sloan Kettering Cancer Center.

Peer review material

Reviewer #1 (Public review): https://doi.org/10.7554/eLife.96519.3.sa1
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.96519.3.sa2
Author response https://doi.org/10.7554/eLife.96519.3.sa3

Additional files

Supplementary files

Supplementary file 1. Characteristics of Alzheimer's disease (AD) and control donors and samples.

Supplementary file 2. Targeted-sequencing gene panel.

Supplementary file 3. Variants identified in Alzheimer's disease and control brain samples.

Supplementary file 4. Pathway enrichment analysis for genes target of pathogenic variants in PU.1 samples.

Supplementary file 5. BRAFV600E in brain PU.1+ cells from Histiocytosis patients.

Supplementary file 6. Predicted deleterious variants by whole exome sequencing (WES).

Supplementary file 7. RNAseq analysis of mouse cell lines: Differential expressed genes and GSEA analysis.

Supplementary file 8. RNAseq analysis of human induced Pluripotent Stem Cells (hIPSC) derived microglial-like cells: Differential expressed genes and gene set enrichment analysis (GSEA).

Supplementary file 9. Single nuclei RNAseq analysis of control and Alzheimer's disease (AD) microglia: Differential expressed genes per clusters and gene set enrichment analysis (GSEA).

MDAR checklist

Data availability

DNA sequencing data processed for selection of somatic variants are available for all patients and samples in *Supplementary file 3*. Raw DNA sequencing data (FASTQ files) from targeted-deep sequencing are deposited in dbGaP under project accession number phs002213.v1.p1, for samples where patient-informed consent for public deposition of DNA sequencing data was obtained. Bulk RNAseq raw data from IPSCs and MAC lines are deposited in GEO (GSE274037). Sn-RNAseq raw data are deposited in GEO (GSE286627) and as an interactive analysis web tool accessible at https://weillcornellmed.shinyapps.io/Human_brain/.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Vicario R	2025	DNA sequencing of postmortem brain samples processed for selection of somatic variants	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002213.v1.p1	NCBI dbGaP, phs002213.v1.p1
Geissmann F, Vicario R	2025	A microglia clonal inflammatory disorder in Alzheimer's Disease	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE274037	NCBI Gene Expression Omnibus, GSE274037
Geissmann F, Vicario R	2025	A microglia clonal inflammatory disorder in Alzheimer's Disease II	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE286627	NCBI Gene Expression Omnibus, GSE286627

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nature Methods* **7**:248–249. DOI: https://doi.org/10.1038/nmeth0410-248, PMID: 20354512

Alzheimer's Association. 2019. Alzheimer's disease facts and figures. alzheimers dement. https://www.alz.org/media/documents/alzheimers-facts-and-figures-2019-r.pdf [Accessed June 1, 2024].

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**:149–157. DOI: https://doi.org/10.1038/s41586-019-1711-4, PMID: 31634902

Aoki Y, Niihori T, Banjo T, Okamoto N, Mizuno S, Kurosawa K, Ogata T, Takada F, Yano M, Ando T, Hoshika T, Barnett C, Ohashi H, Kawame H, Hasegawa T, Okutani T, Nagashima T, Hasegawa S, Funayama R, Nagashima T, et al. 2013. Gain-of-function mutations in RIT1 cause Noonan syndrome, a RAS/MAPK pathway syndrome. *American Journal of Human Genetics* **93**:173–180. DOI: https://doi.org/10.1016/j.ajhg.2013.05.021, PMID: 23791108

Arnaud L, Benech P, Greetham L, Stephan D, Jimenez A, Jullien N, García-González L, Tsvetkov PO, Devred F, Sancho-Martinez I, Izpisua Belmonte JC, Baranger K, Rivera S, Nivet E. 2022. APOE4 drives inflammation in human astrocytes via TAGLN3 repression and NF-κB activation. *Cell Reports* **40**:111200. DOI: https://doi.org/10.1016/j.celrep.2022.111200, PMID: 35977506

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**:25–29. DOI: https://doi.org/10.1038/75556, PMID: 10802651

Askew K, Li K, Olmos-Alonso A, Garcia-Moreno F, Liang Y, Richardson P, Tipton T, Chapman MA, Riecken K, Beccari S, Sierra A, Molnár Z, Cragg MS, Garaschuk O, Perry VH, Gomez-Nicola D. 2017. Coupled proliferation and apoptosis maintain the rapid turnover of microglia in the adult brain. *Cell Reports* **18**:391–405. DOI: https://doi.org/10.1016/j.celrep.2016.12.041, PMID: 28076784

Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, Nainys J, Wu K, Kiseliovas V, Setty M, Choi K, Fromme RM, Dao P, McKenney PT, Wasti RC, Kadaveru K, Mazutis L, Rudensky AY, Pe'er D. 2018. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**:1293–1308. DOI: https://doi.org/10.1016/j.cell.2018.05.060, PMID: 29961579

Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. 2014. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**:293. DOI: https://doi.org/10.1186/1471-2105-15-293, PMID: 25176396

Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, Martincorena I, Petljak M, Alexandrov LB, Gundem G, Tarpey PS, Roerink S, Blokker J, Maddison M, Mudie L, Robinson B, Nik-Zainal S, Campbell P,

Goldman N, van de Wetering M, et al. 2014. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**:422–425. DOI: https://doi.org/10.1038/nature13448, PMID: 25043003

Bernard V, Gebauer N, Dinh T, Stegemann J, Feller AC, Merz H. 2014. Applicability of next-generation sequencing to decalcified formalin-fixed and paraffin-embedded chronic myelomonocytic leukaemia samples. *International Journal of Clinical and Experimental Pathology* **7**:1667–1676 PMID: 24817963.

Bhatia A, Hatzoglou V, Ulaner G, Rampal R, Hyman DM, Abdel-Wahab O, Durham BH, Dogan A, Ozkaya N, Yabe M, Petrova-Drus K, Panageas KS, Reiner A, Rosenblum M, Diamond EL. 2020. Neurologic and oncologic features of Erdheim-Chester disease: a 30-patient series. *Neuro-Oncology* **22**:979–992. DOI: https://doi.org/10.1093/neuonc/noaa008, PMID: 31950179

Bian Z, Gong Y, Huang T, Lee CZW, Bian L, Bai Z, Shi H, Zeng Y, Liu C, He J, Zhou J, Li X, Li Z, Ni Y, Ma C, Cui L, Zhang R, Chan JKY, Ng LG, Lan Y, et al. 2020. Deciphering human macrophage development at single-cell resolution. *Nature* **582**:571–576. DOI: https://doi.org/10.1038/s41586-020-2316-7, PMID: 32499656

Blasi E, Barluzzi R, Bocchini V, Mazzolla R, Bistoni F. 1990. Immortalization of murine microglial cells by a v-raf/v-myc carrying retrovirus. *Journal of Neuroimmunology* **27**:229–237. DOI: https://doi.org/10.1016/0165-5728(90)90073-v, PMID: 2110186

Bollag G, Clapp DW, Shih S, Adler F, Zhang YY, Thompson P, Lange BJ, Freedman MH, McCormick F, Jacks T, Shannon K. 1996. Loss of NF1 results in activation of the Ras signaling pathway and leads to aberrant growth in haematopoietic cells. *Nature Genetics* **12**:144–148. DOI: https://doi.org/10.1038/ng0296-144, PMID: 8563751

Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermey P, Belotti M, Gauthier-Villars M, Stoppa-Lyonnet D, Consolino E, Brugières L, Caron O, Benusiglio PR, Bressac-de Paillerets B, Bonadona V, Bonaïti-Pellié C, Tinat J, Baert-Desurmont S, Frebourg T. 2015. Revisiting li-fraumeni syndrome from TP53 mutation carriers. *Journal of Clinical Oncology* **33**:2345–2352. DOI: https://doi.org/10.1200/JCO.2014.59.5728, PMID: 26014290

Bouzid H, Belk JA, Jan M, Qi Y, Sarnowski C, Wirth S, Ma L, Chrostek MR, Ahmad H, Nachun D, Yao W, NHLBI Trans-Omics for Precision Medicine Consortium, Beiser A, Bick AG, Bis JC, Fornage M, Longstreth WT Jr, Lopez OL, Natarajan P, Psaty BM, et al. 2023. Clonal hematopoiesis is associated with protection from Alzheimer's disease. *Nature Medicine* **29**:1662–1670. DOI: https://doi.org/10.1038/s41591-023-02397-2, PMID: 37322115

Boyd LC, O'Brien KJ, Ozkaya N, Lehky T, Meoded A, Gochuico BR, Hannah-Shmouni F, Nath A, Toro C, Gahl WA, Estrada-Veras JI, Dave RH. 2020. Neurological manifestations of Erdheim-chester disease. *Annals of Clinical and Translational Neurology* **7**:497–506. DOI: https://doi.org/10.1002/acn3.51014, PMID: 32227455

Braak H, Braak E. 1991. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica* **82**:239–259. DOI: https://doi.org/10.1007/BF00308809, PMID: 1759558

Braak H, Braak E. 1995. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging* **16**:271–278. DOI: https://doi.org/10.1016/0197-4580(95)00021-6, PMID: 7566337

Brand K, Kentsch H, Glashoff C, Rosenberger G. 2014. RASopathy-associated CBL germline mutations cause aberrant ubiquitylation and trafficking of EGFR. *Human Mutation* **35**:1372–1381. DOI: https://doi.org/10.1002/humu.22682, PMID: 25178484

Bras J, Guerreiro R, Hardy J. 2012. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nature Reviews. Neuroscience* **13**:453–464. DOI: https://doi.org/10.1038/nrn3271, PMID: 22714018

Brownlee PM, Chambers AL, Cloney R, Bianchi A, Downs JA. 2014. BAF180 promotes cohesion and prevents genome instability and aneuploidy. *Cell Reports* **6**:973–981. DOI: https://doi.org/10.1016/j.celrep.2014.02.012, PMID: 24613357

Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, Rudolph JE, Yaeger R, Soumerai T, Nissan MH, Chang MT, Chandarlapaty S, Traina TA, Paik PK, Ho AL, Hantash FM, Grupe A, Baxi SS, Callahan MK, Snyder A, et al. 2017. OncoKB: a precision oncology knowledge base. *JCO Precision Oncology* **2017**:PO.17.00011. DOI: https://doi.org/10.1200/PO.17.00011, PMID: 28890946

Chartier-Harlin MC, Crawford F, Houlden H, Warren A, Hughes D, Fidani L, Goate A, Rossor M, Roques P, Hardy J. 1991. Early-onset Alzheimer's disease caused by mutations at codon 717 of the beta-amyloid precursor protein gene. *Nature* **353**:844–846. DOI: https://doi.org/10.1038/353844a0, PMID: 1944558

Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Brannon AR, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, et al. 2015. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *The Journal of Molecular Diagnostics* **17**:251–264. DOI: https://doi.org/10.1016/j.jmoldx.2014.12.006, PMID: 25801821

Choi J, Landrette SF, Wang T, Evans P, Bacchiocchi A, Bjorson R, Cheng E, Stiegler AL, Gathiaka S, Acevedo O, Boggon TJ, Krauthammer M, Halaban R, Xu T. 2014. Identification of PLX4032-resistance mechanisms and implications for novel RAF inhibitors. *Pigment Cell & Melanoma Research* **27**:253–262. DOI: https://doi.org/10.1111/pcmr.12197, PMID: 24283590

Coniglio SJ, Eugenin E, Dobrenis K, Stanley ER, West BL, Symons MH, Segall JE. 2012. Microglial stimulation of glioblastoma invasion involves epidermal growth factor receptor (EGFR) and colony stimulating factor 1 receptor (CSF-1R) signaling. *Molecular Medicine* **18**:519–527. DOI: https://doi.org/10.2119/molmed.2011.00217, PMID: 22294205

Cybulski C, Wokołorczyk D, Jakubowska A, Huzarski T, Byrski T, Gronwald J, Masojć B, Deebniak T, Górski B, Blecharz P, Narod SA, Lubiński J. 2011. Risk of breast cancer in women with a CHEK2 mutation with and without a family history of breast cancer. *Journal of Clinical Oncology* **29**:3747–3752. DOI: https://doi.org/10.1200/JCO.2010.34.0778, PMID: 21876083

D'Gama AM, Woodworth MB, Hossain AA, Bizzotto S, Hatem NE, LaCoursiere CM, Najm I, Ying Z, Yang E, Barkovich AJ, Kwiatkowski DJ, Vinters HV, Madsen JR, Mathern GW, Blümcke I, Poduri A, Walsh CA. 2017. Somatic mutations activating the mTOR pathway in dorsal telencephalic progenitors cause a continuum of cortical dysplasias. *Cell Reports* **21**:3754–3766. DOI: https://doi.org/10.1016/j.celrep.2017.11.106, PMID: 29281825

Diamond EL, Durham BH, Haroche J, Yao Z, Ma J, Parikh SA, Wang Z, Choi J, Kim E, Cohen-Aubart F, Lee SCW, Gao Y, Micol JB, Campbell P, Walsh MP, Sylvester B, Dolgalev I, Aminova O, Heguy A, Zappile P, et al. 2016. Diverse and targetable kinase alterations drive histiocytic neoplasms. *Cancer Discovery* **6**:154–165. DOI: https://doi.org/10.1158/2159-8290.CD-15-0913, PMID: 26566875

Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P. 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *The Lancet. Neurology* **6**:734–746. DOI: https://doi.org/10.1016/S1474-4422(07)70178-3, PMID: 17616482

Dunbar AJ, Gondek LP, O'Keefe CL, Makishima H, Rataul MS, Szpurka H, Sekeres MA, Wang XF, McDevitt MA, Maciejewski JP. 2008. 250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies. *Cancer Research* **68**:10349–10357. DOI: https://doi.org/10.1158/0008-5472.CAN-08-2754, PMID: 19074904

Durham BH, Lopez Rodrigo E, Picarsic J, Abramson D, Rotemberg V, De Munck S, Pannecoucke E, Lu SX, Pastore A, Yoshimi A, Mandelker D, Ceyhan-Birsoy O, Ulaner GA, Walsh M, Yabe M, Petrova-Drus K, Arcila ME, Ladanyi M, Solit DB, Berger MF, et al. 2019. Activating mutations in CSF1R and additional receptor tyrosine kinases in histiocytic neoplasms. *Nature Medicine* **25**:1839–1842. DOI: https://doi.org/10.1038/s41591-019-0653-6, PMID: 31768065

Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, Parker JJ, Atabay KD, Gilmore EC, Poduri A, Park PJ, Walsh CA. 2012. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151**:483–496. DOI: https://doi.org/10.1016/j.cell.2012.09.035, PMID: 23101622

Fang Y, Tsao CC, Goodman BK, Furumai R, Tirado CA, Abraham RT, Wang XF. 2004. ATR functions as a gene dosage-dependent tumor suppressor on a mismatch repair-deficient background. *The EMBO Journal* **23**:3164–3174. DOI: https://doi.org/10.1038/sj.emboj.7600315, PMID: 15282542

Fernandes MS, Reddy MM, Croteau NJ, Walz C, Weisbach H, Podar K, Band H, Carroll M, Reiter A, Larson RA, Salgia R, Griffin JD, Sattler M. 2010. Novel oncogenic mutations of CBL in human acute myeloid leukemia that activate growth and survival pathways depend on increased metabolism. *The Journal of Biological Chemistry* **285**:32596–32605. DOI: https://doi.org/10.1074/jbc.M110.106161, PMID: 20622007

Ferrari R, Grassi M, Salvi E, Borroni B, Palluzzi F, Pepe D, D'Avila F, Padovani A, Archetti S, Rainero I, Rubino E, Pinessi L, Benussi L, Binetti G, Ghidoni R, Galimberti D, Scarpini E, Serpente M, Rossi G, Giaccone G, et al. 2015. A genome-wide screening and SNPs-to-genes approach to identify novel genetic risk factors associated with frontotemporal dementia. *Neurobiology of Aging* **36**:2904. DOI: https://doi.org/10.1016/j.neurobiolaging.2015.06.005, PMID: 26154020

Frank SA. 2010. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *PNAS* **107 Suppl 1**:1725–1730. DOI: https://doi.org/10.1073/pnas.0909343106, PMID: 19805033

Galatro TF, Holtman IR, Lerario AM, Vainchtein ID, Brouwer N, Sola PR, Veras MM, Pereira TF, Leite REP, Möller T, Wes PD, Sogayar MC, Laman JD, den Dunnen W, Pasqualucci CA, Oba-Shinjo SM, Boddeke E, Marie SKN, Eggen BJL. 2017. Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nature Neuroscience* **20**:1162–1171. DOI: https://doi.org/10.1038/nn.4597, PMID: 28671693

Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, Purcell SM, Svantesson O, Landén M, Höglund M, Lehmann S, Gabriel SB, Moran JL, Lander ES, Sullivan PF, Sklar P, et al. 2014. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England Journal of Medicine* **371**:2477–2487. DOI: https://doi.org/10.1056/NEJMoa1409405, PMID: 25426838

Ghosh S, Castillo E, Frias ES, Swanson RA. 2018. Bioenergetic regulation of microglia. *Glia* **66**:1200–1212. DOI: https://doi.org/10.1002/glia.23271, PMID: 29219210

Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L. 1991. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**:704–706. DOI: https://doi.org/10.1038/349704a0, PMID: 1671712

Gómez-Seguí I, Makishima H, Jerez A, Yoshida K, Przychodzen B, Miyano S, Shiraishi Y, Husseinzadeh HD, Guinta K, Clemente M, Hosono N, McDevitt MA, Moliterno AR, Sekeres MA, Ogawa S, Maciejewski JP. 2013. Novel recurrent mutations in the RAS-like GTP-binding gene RIT1 in myeloid malignancies. *Leukemia* **27**:1943–1946. DOI: https://doi.org/10.1038/leu.2013.179, PMID: 23765226

Gosselin D, Skola D, Coufal NG, Holtman IR, Schlachetzki JCM, Sajti E, Jaeger BN, O'Connor C, Fitzpatrick C, Pasillas MP, Pena M, Adair A, Gonda DD, Levy ML, Ransohoff RM, Gage FH, Glass CK. 2017. An environment-

dependent transcriptional network specifies human microglia identity. *Science* **356**:eaal3222. DOI: https://doi.org/10.1126/science.aal3222, PMID: 28546318

Graham JM Jr, Schwartz CE. 2013. MED12 related disorders. *American Journal of Medical Genetics. Part A* **161A**:2734–2740. DOI: https://doi.org/10.1002/ajmg.a.36183, PMID: 24123922

Grand FH, Hidalgo-Curtis CE, Ernst T, Zoi K, Zoi C, McGuire C, Kreil S, Jones A, Score J, Metzgeroth G, Oscier D, Hall A, Brandts C, Serve H, Reiter A, Chase AJ, Cross NCP. 2009. Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. *Blood* **113**:6182–6192. DOI: https://doi.org/10.1182/blood-2008-12-194548, PMID: 19387008

Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, Cruchaga C, Sassi C, Kauwe JSK, Younkin S, Hazrati L, Collinge J, Pocock J, Lashley T, Williams J, Lambert J-C, Amouyel P, Goate A, Rademakers R, Morgan K, et al. 2013. TREM2 variants in Alzheimer's disease. *The New England Journal of Medicine* **368**:117–127. DOI: https://doi.org/10.1056/NEJMoa1211851, PMID: 23150934

Hadley W. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. DOI: https://doi.org/10.1007/978-3-319-24277-4

Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, Schnittger S, Sanada M, Kon A, Alpermann T, Yoshida K, Roller A, Nadarajah N, Shiraishi Y, Shiozawa Y, Chiba K, Tanaka H, Koeffler HP, Klein HU, Dugas M, et al. 2014. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**:241–247. DOI: https://doi.org/10.1038/leu.2013.336, PMID: 24220272

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* **184**:3573–3587. DOI: https://doi.org/10.1016/j.cell.2021.04.048, PMID: 34062119

Hebert LE, Weuve J, Scherr PA, Evans DA. 2013. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* **80**:1778–1783. DOI: https://doi.org/10.1212/WNL.0b013e31828726f5, PMID: 23390181

Heim RA, Kam-Morgan LN, Binnie CG, Corns DD, Cayouette MC, Farber RA, Aylsworth AS, Silverman LM, Luce MC. 1995. Distribution of 13 truncating mutations in the neurofibromatosis 1 gene. *Human Molecular Genetics* **4**:975–981. DOI: https://doi.org/10.1093/hmg/4.6.975, PMID: 7655472

Henn A, Lund S, Hedtjärn M, Schrattenholz A, Pörzgen P, Leist M. 2009. The suitability of BV2 cells as alternative model system for primary microglia cultures or for animal experiments examining brain inflammation. *ALTEX* **26**:83–94. DOI: https://doi.org/10.14573/altex.2009.2.83, PMID: 19565166

Héritier S, Barkaoui M-A, Miron J, Thomas C, Moshous D, Lambilliotte A, Mazingue F, Kebaili K, Jeziorski E, Plat G, Aladjidi N, Pacquement H, Galambrun C, Brugières L, Leverger G, Mansuy L, Paillard C, Deville A, Pagnier A, Lutun A, et al. 2018. Incidence and risk factors for clinical neurodegenerative Langerhans cell histiocytosis: a longitudinal cohort study. *British Journal of Haematology* **183**:608–617. DOI: https://doi.org/10.1111/bjh.15577, PMID: 30421536

Huang DW, Sherman BT, Lempicki RA. 2009a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**:44–57. DOI: https://doi.org/10.1038/nprot.2008.211

Huang DW, Sherman BT, Lempicki RA. 2009b. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**:1–13. DOI: https://doi.org/10.1093/nar/gkn923, PMID: 19033363

Itan Y, Shang L, Boisson B, Ciancanelli MJ, Markle JG, Martinez-Barricarte R, Scott E, Shah I, Stenson PD, Gleeson J, Cooper DN, Quintana-Murci L, Zhang SY, Abel L, Casanova JL. 2016. The mutation significance cutoff: gene-level thresholds for variant predictions. *Nature Methods* **13**:109–110. DOI: https://doi.org/10.1038/nmeth.3739, PMID: 26820543

Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C, Getz G, Correa A, et al. 2014. Age-related clonal hematopoiesis associated with adverse outcomes. *The New England Journal of Medicine* **371**:2488–2498. DOI: https://doi.org/10.1056/NEJMoa1408617, PMID: 25426837

Jaiswal S, Ebert BL. 2019. Clonal hematopoiesis in human aging and disease. *Science* **366**:eaan4673. DOI: https://doi.org/10.1126/science.aan4673, PMID: 31672865

Javadi M, Richmond TD, Huang K, Barber DL. 2013. CBL linker region and RING finger mutations lead to enhanced granulocyte-macrophage colony-stimulating factor (GM-CSF) signaling via elevated levels of JAK2 and LYN. *The Journal of Biological Chemistry* **288**:19459–19470. DOI: https://doi.org/10.1074/jbc.M113.475087, PMID: 23696637

Jayaraman A, Htike TT, James R, Picon C, Reynolds R. 2021. TNF-mediated neuroinflammation is linked to neuronal necroptosis in Alzheimer's disease hippocampus. *Acta Neuropathologica Communications* **9**:159. DOI: https://doi.org/10.1186/s40478-021-01264-w, PMID: 34625123

Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, Bjornsson S, Huttenlocher J, Levey AI, Lah JJ, Rujescu D, Hampel H, Giegling I, Andreassen OA, Engedal K, Ulstein I, Djurovic S, Ibrahim-Verbaas C, Hofman A, Ikram MA, et al. 2013. Variant of TREM2 associated with the risk of Alzheimer's disease. *The New England Journal of Medicine* **368**:107–116. DOI: https://doi.org/10.1056/NEJMoa1211103, PMID: 23150908

Karch CM, Cruchaga C, Goate AM. 2014. Alzheimer's disease genetics: from the bench to the clinic. *Neuron* **83**:11–26. DOI: https://doi.org/10.1016/j.neuron.2014.05.041, PMID: 24991952

Karch CM, Goate AM. 2015. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biological Psychiatry* **77**:43–51. DOI: https://doi.org/10.1016/j.biopsych.2014.05.006, PMID: 24951455

Keogh MJ, Wei W, Aryaman J, Walker L, van den Ameele J, Coxhead J, Wilson I, Bashton M, Beck J, West J, Chen R, Haudenschild C, Bartha G, Luo S, Morris CM, Jones NS, Attems J, Chinnery PF. 2018. High prevalence of focal and multi-focal somatic genetic variants in the human brain. *Nature Communications* **9**:4257. DOI: https://doi.org/10.1038/s41467-018-06331-w, PMID: 30323172

Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, David E, Baruch K, Lara-Astaiso D, Toth B, Itzkovitz S, Colonna M, Schwartz M, Amit I. 2017. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**:1276–1290.. DOI: https://doi.org/10.1016/j.cell. 2017.05.018, PMID: 28602351

Khoshkhoo S, Wang Y, Chahine Y, Erson-Omay EZ, Robert SM, Kiziltug E, Damisah EC, Nelson-Williams C, Zhu G, Kong W, Huang AY, Stronge E, Phillips HW, Chhouk BH, Bizzotto S, Chen MH, Adikari TN, Ye Z, Witkowski T, Lai D, et al. 2023. Contribution of somatic ras/raf/mitogen-activated protein kinase variants in the hippocampus in drug-resistant mesial temporal lobe epilepsy. *JAMA Neurology* **80**:578–587. DOI: https://doi. org/10.1001/jamaneurol.2023.0473, PMID: 37126322

Kim E, Ilic N, Shrestha Y, Zou L, Kamburov A, Zhu C, Yang X, Lubonja R, Tran N, Nguyen C, Lawrence MS, Piccioni F, Bagul M, Doench JG, Chouinard CR, Wu X, Hogstrom L, Natoli T, Tamayo P, Horn H, et al. 2016. Systematic functional interrogation of rare cancer variants identifies oncogenic alleles. *Cancer Discovery* **6**:714–726. DOI: https://doi.org/10.1158/2159-8290.CD-16-0160, PMID: 27147599

Kim JS. 2023. Monocyte-Derived Microglia With. [*bioRxiv*]. DOI: https://doi.org/10.1101/2023.11.16.567402

Kinney JW, Bemiller SM, Murtishaw AS, Leisgang AM, Salazar AM, Lamb BT. 2018. Inflammation as a central mechanism in Alzheimer's disease. *Alzheimer's & Dementia* **4**:575–590. DOI: https://doi.org/10.1016/j.trci. 2018.06.014, PMID: 30406177

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**:310–315. DOI: https://doi.org/10. 1038/ng.2892, PMID: 24487276

Klampfl T, Milosevic JD, Puda A, Schönegger A, Bagienski K, Berg T, Harutyunyan AS, Gisslinger B, Rumi E, Malcovati L, Pietra D, Elena C, Della Porta MG, Pieri L, Guglielmelli P, Bock C, Doubek M, Dvorakova D, Suvajdzic N, Tomin D, et al. 2013. Complex patterns of chromosome 11 aberrations in myeloid malignancies target CBL, MLL, DDB1 and LMO2. *PLOS ONE* **8**:e77819. DOI: https://doi.org/10.1371/journal.pone.0077819, PMID: 24147083

Koh HY, Kim SH, Jang J, Kim H, Han S, Lim JS, Son G, Choi J, Park BO, Heo WD, Han J, Lee HJ, Lee D, Kang HC, Shong M, Paik SB, Kim DS, Lee JH. 2018. BRAF somatic mutation contributes to intrinsic epileptogenicity in pediatric brain tumors. *Nature Medicine* **24**:1662–1668. DOI: https://doi.org/10.1038/ s41591-018-0172-x, PMID: 30224756

Korotkevich G, Sukhov V, Sergushichev A. 2019. Fast gene set enrichment analysis. *bioRxiv*. DOI: https://doi. org/10.1101/060012

Kouri N, Ross OA, Dombroski B, Younkin CS, Serie DJ, Soto-Ortolaza A, Baker M, Finch NCA, Yoon H, Kim J, Fujioka S, McLean CA, Ghetti B, Spina S, Cantwell LB, Farlow MR, Grafman J, Huey ED, Ryung Han M, Beecher S, et al. 2015. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nature Communications* **6**:7247. DOI: https://doi.org/10.1038/ ncomms8247, PMID: 26077951

Krasemann S, Madore C, Cialic R, Baufeld C, Calcagno N, El Fatimy R, Beckers L, O'Loughlin E, Xu Y, Fanek Z, Greco DJ, Smith ST, Tweet G, Humulock Z, Zrzavy T, Conde-Sanroman P, Gacias M, Weng Z, Chen H, Tjon E, et al. 2017. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity* **47**:566–581. DOI: https://doi.org/10.1016/j.immuni.2017.08.008, PMID: 28930663

Ku JL, Park SH, Yoon KA, Shin YK, Kim KH, Choi JS, Kang HC, Kim IJ, Han IO, Park JG. 2007. Genetic alterations of the TGF-beta signaling pathway in colorectal cancer cell lines: a novel mutation in Smad3 associated with the inactivation of TGF-beta-induced transcriptional activation. *Cancer Letters* **247**:283–292. DOI: https://doi. org/10.1016/j.canlet.2006.05.008, PMID: 16828225

Lachén-Montes M, González-Morales A, de Morentin XM, Pérez-Valderrama E, Ausín K, Zelaya MV, Serna A, Aso E, Ferrer I, Fernández-Irigoyen J, Santamaría E. 2016. An early dysregulation of FAK and MEK/ERK signaling pathways precedes the β-amyloid deposition in the olfactory bulb of APP/PS1 mouse model of Alzheimer's disease. *Journal of Proteomics* **148**:149–158. DOI: https://doi.org/10.1016/j.jprot.2016.07.032, PMID: 27498392

Lachmann N, Ackermann M, Frenzel E, Liebhaber S, Brennig S, Happle C, Hoffmann D, Klimenkova O, Lüttge D, Buchegger T, Kühnel MP, Schambach A, Janciauskiene S, Figueiredo C, Hansen G, Skokowa J, Moritz T. 2015. Large-scale hematopoietic differentiation of human induced pluripotent stem cells provides granulocytes or macrophages for cell replacement therapies. *Stem Cell Reports* **4**:282–296. DOI: https://doi.org/10.1016/j. stemcr.2015.01.005, PMID: 25680479

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research* **42**:D980–D985. DOI: https://doi.org/10.1093/nar/gkt1113, PMID: 24234437

Lanoiselée HM, Nicolas G, Wallon D, Rovelet-Lecrux A, Lacour M, Rousseau S, Richard AC, Pasquier F, Rollin-Sillaire A, Martinaud O, Quillard-Muraine M, de la Sayette V, Boutoleau-Bretonniere C, Etcharry-Bouyx F, Chauviré V, Sarazin M, le Ber I, Epelbaum S, Jonveaux T, Rouaud O, et al. 2017. APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: a genetic screening study of familial and sporadic cases. *PLOS Medicine* **14**:e1002270. DOI: https://doi.org/10.1371/journal.pmed.1002270, PMID: 28350801

Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K. 1995a. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**:973–977. DOI: https://doi.org/10.1126/science.7638622, PMID: 7638622

Levy-Lahad E, Wijsman EM, Nemens E, Anderson L, Goddard KA, Weber JL, Bird TD, Schellenberg GD. 1995b. A familial Alzheimer's disease locus on chromosome 1. *Science* **269**:970–973. DOI: https://doi.org/10.1126/science.7638621, PMID: 7638621

Lim JS, Kim W, Kang HC, Kim SH, Park AH, Park EK, Cho YW, Kim S, Kim HM, Kim JA, Kim J, Rhee H, Kang SG, Kim HD, Kim D, Kim DS, Lee JH. 2015. Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. *Nature Medicine* **21**:395–400. DOI: https://doi.org/10.1038/nm.3824, PMID: 25799227

Lindberg OR, Brederlau A, Kuhn HG. 2014. Epidermal growth factor treatment of the adult brain subventricular zone leads to focal microglia/macrophage accumulation and angiogenesis. *Stem Cell Reports* **2**:440–448. DOI: https://doi.org/10.1016/j.stemcr.2014.02.003, PMID: 24749069

Liu X, Quan N. 2018. Microglia and CNS Interleukin-1: beyond immunological concepts. *Frontiers in Neurology* **9**:8. DOI: https://doi.org/10.3389/fneur.2018.00008, PMID: 29410649

Liyasova MS, Ma K, Lipkowitz S. 2015. Molecular pathways: cbl proteins in tumorigenesis and antitumor immunity-opportunities for cancer treatment. *Clinical Cancer Research* **21**:1789–1794. DOI: https://doi.org/10.1158/1078-0432.CCR-13-2490, PMID: 25477533

Loh ML, Sakai DS, Flotho C, Kang M, Fliegauf M, Archambeault S, Mullighan CG, Chen L, Bergstraesser E, Bueso-Ramos CE, Emanuel PD, Hasle H, Issa JP, van den Heuvel-Eibrink MM, Locatelli F, Stary J, Trebo M, Wlodarski M, Zecca M, Shannon KM, et al. 2009. Mutations in CBL occur frequently in juvenile myelomonocytic leukemia. *Blood* **114**:1859–1863. DOI: https://doi.org/10.1182/blood-2009-01-198416, PMID: 19571318

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. DOI: https://doi.org/10.1186/s13059-014-0550-8, PMID: 25516281

Lu B, Klingbeil O, Tarumoto Y, Somerville TDD, Huang YH, Wei Y, Wai DC, Low JKK, Milazzo JP, Wu XS, Cao Z, Yan X, Demerdash OE, Huang G, Mackay JP, Kinney JB, Shi J, Vakoc CR. 2018. A Transcription factor addiction in leukemia imposed by the MLL promoter sequence. *Cancer Cell* **34**:970–981. DOI: https://doi.org/10.1016/j.ccell.2018.10.015, PMID: 30503706

Lumsden AL, Mulugeta A, Zhou A, Hyppönen E. 2020. Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank. *EBioMedicine* **59**:102954. DOI: https://doi.org/10.1016/j.ebiom.2020.102954, PMID: 32818802

Martin PM, Stanley RE, Ross AP, Freitas AE, Moyer CE, Brumback AC, Iafrati J, Stapornwongkul KS, Dominguez S, Kivimäe S, Mulligan KA, Pirooznia M, McCombie WR, Potash JB, Zandi PP, Purcell SM, Sanders SJ, Zuo Y, Sohal VS, Cheyette BNR. 2018. DIXDC1 contributes to psychiatric susceptibility by regulating dendritic spine and glutamatergic synapse density via GSK3 and Wnt/β-catenin signaling. *Molecular Psychiatry* **23**:467–475. DOI: https://doi.org/10.1038/mp.2016.184, PMID: 27752079

Martincorena I, Campbell PJ. 2015. Somatic mutation in cancer and normal cells. *Science* **349**:1483–1489. DOI: https://doi.org/10.1126/science.aab4082, PMID: 26404825

Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, Wedge DC, Fullam A, Alexandrov LB, Tubio JM, Stebbings L, Menzies A, Widaa S, Stratton MR, Jones PH, Campbell PJ. 2015. Tumor evolution: High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**:880–886. DOI: https://doi.org/10.1126/science.aaa6806, PMID: 25999502

Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, Cagan A, Murai K, Mahbubani K, Stratton MR, Fitzgerald RC, Handford PA, Campbell PJ, Saeb-Parsy K, Jones PH. 2018. Somatic mutant clones colonize the human esophagus with age. *Science* **362**:911–917. DOI: https://doi.org/10.1126/science.aau3879, PMID: 30337457

Mass E, Jacome-Galarza CE, Blank T, Lazarov T, Durham BH, Ozkaya N, Pastore A, Schwabenland M, Chung YR, Rosenblum MK, Prinz M, Abdel-Wahab O, Geissmann F. 2017. A somatic mutation in erythro-myeloid progenitors causes neurodegenerative disease. *Nature* **549**:389–393. DOI: https://doi.org/10.1038/nature23672, PMID: 28854169

Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, Menon M, He L, Abdurrob F, Jiang X, Martorell AJ, Ransohoff RM, Hafler BP, Bennett DA, Kellis M, Tsai LH. 2019. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**:332–337. DOI: https://doi.org/10.1038/s41586-019-1195-2, PMID: 31042697

McConnell MJ, Moran JV, Abyzov A, Akbarian S, Bae T, Cortes-Ciriano I, Erwin JA, Fasching L, Flasch DA, Freed D, Ganz J, Jaffe AE, Kwan KY, Kwon M, Lodato MA, Mills RE, Paquola ACM, Rodin RE, Rosenbluh C, Sestan N, et al. 2017. Intersection of diverse neuronal genomes and neuropsychiatric disease: the brain somatic mosaicism network. *Science* **356**:eaal1641. DOI: https://doi.org/10.1126/science.aal1641, PMID: 28450582

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* **34**:939–944. DOI: https://doi.org/10.1212/wnl.34.7.939, PMID: 6610841

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH. 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7**:263–269. DOI: https://doi.org/10.1016/j.jalz.2011.03.005, PMID: 21514250

McQuade A, Blurton-Jones M. 2019. Microglia in alzheimer's disease: exploring how genetics and phenotype influence risk. *Journal of Molecular Biology* **431**:1805–1817. DOI: https://doi.org/10.1016/j.jmb.2019.01.045, PMID: 30738892

Meyer Zum Büschenfelde U, Brandenstein LI, von Elsner L, Flato K, Holling T, Zenker M, Rosenberger G, Kutsche K. 2018. RIT1 controls actin dynamics via complex formation with RAC1/CDC42 and PAK1. *PLOS Genetics* **14**:e1007370. DOI: https://doi.org/10.1371/journal.pgen.1007370, PMID: 29734338

Miller MB, Reed HC, Walsh CA. 2021. Brain somatic mutation in aging and Alzheimer's Disease. *Annual Review of Genomics and Human Genetics* **22**:239–256. DOI: https://doi.org/10.1146/annurev-genom-121520-081242, PMID: 33979534

Murrell JR, Price B, Lane KA, Baiyewu O, Gureje O, Ogunniyi A, Unverzagt FW, Smith-Gamble V, Gao S, Hendrie HC, Hall KS. 2006. Association of apolipoprotein E genotype and Alzheimer disease in African Americans. *Archives of Neurology* **63**:431–434. DOI: https://doi.org/10.1001/archneur.63.3.431, PMID: 16533971

Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, DeStefano AL, Kara E, Bras J, Sharma M, Schulte C, Keller MF, Arepalli S, Letson C, Edsall C, Stefansson H, Liu X, Pliner H, Lee JH, Cheng R, et al. 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature Genetics* **46**:989–993. DOI: https://doi.org/10.1038/ng.3043, PMID: 25064009

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. *Genome Research* **11**:863–874. DOI: https://doi.org/10.1101/gr.176601, PMID: 11337480

Ng PKS, Li J, Jeong KJ, Shao S, Chen H, Tsang YH, Sengupta S, Wang Z, Bhavana VH, Tran R, Soewito S, Minussi DC, Moreno D, Kong K, Dogruluk T, Lu H, Gao J, Tokheim C, Zhou DC, Johnson AM, et al. 2018. Systematic functional annotation of somatic mutations in cancer. *Cancer Cell* **33**:450–462. DOI: https://doi.org/10.1016/j.ccell.2018.01.021, PMID: 29533785

Niemeyer CM, Kang MW, Shin DH, Furlan I, Erlacher M, Bunin NJ, Bunda S, Finklestein JZ, Gorr TA, Mehta P, Schmid I, Kropshofer G, Corbacioglu S, Lang PJ, Klein C, Schlegel PG, Heinzmann A, Schneider M, Starý J, van den Heuvel-Eibrink MM, et al. 2010. Germline CBL mutations cause developmental abnormalities and predispose to juvenile myelomonocytic leukemia. *Nature Genetics* **42**:794–800. DOI: https://doi.org/10.1038/ng.641, PMID: 20694012

Niihori T, Aoki Y, Ohashi H, Kurosawa K, Kondoh T, Ishikiriyama S, Kawame H, Kamasaki H, Yamanaka T, Takada F, Nishio K, Sakurai M, Tamai H, Nagashima T, Suzuki Y, Kure S, Fujii K, Imaizumi M, Matsubara Y. 2005. Functional analysis of PTPN11/SHP-2 mutants identified in Noonan syndrome and childhood leukemia. *Journal of Human Genetics* **50**:192–202. DOI: https://doi.org/10.1007/s10038-005-0239-7, PMID: 15834506

Nikolaev SI, Vetiska S, Bonilla X, Boudreau E, Jauhiainen S, Rezai Jahromi B, Khyzha N, DiStefano PV, Suutarinen S, Kiehl TR, Mendes Pereira V, Herman AM, Krings T, Andrade-Barazarte H, Tung T, Valiante T, Zadeh G, Tymianski M, Rauramaa T, Ylä-Herttuala S, et al. 2018. Somatic activating KRAS mutations in arteriovenous malformations of the brain. *The New England Journal of Medicine* **378**:250–261. DOI: https://doi.org/10.1056/NEJMoa1709449, PMID: 29298116

Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, Han CZ, Pena M, Xiao J, Wu Y, Keulen Z, Pasillas MP, O'Connor C, Nickl CK, Schafer ST, Shen Z, Rissman RA, Brewer JB, Gosselin D, Gonda DD, et al. 2019. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**:1134–1139. DOI: https://doi.org/10.1126/science.aay0793, PMID: 31727856

Ogawa S. 2019. Genetics of MDS. *Blood* **133**:1049–1059. DOI: https://doi.org/10.1182/blood-2018-10-844621, PMID: 30670442

Okeyo-Owuor T, White BS, Chatrikhi R, Mohan DR, Kim S, Griffith M, Ding L, Ketkar-Kulkarni S, Hundal J, Laird KM, Kielkopf CL, Ley TJ, Walter MJ, Graubert TA. 2015. U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia* **29**:909–917. DOI: https://doi.org/10.1038/leu.2014.303, PMID: 25311244

Ou W, Yang J, Simanauskaite J, Choi M, Castellanos DM, Chang R, Sun J, Jagadeesan N, Parfitt KD, Cribbs DH, Sumbria RK. 2021. Biologic TNF-α inhibitors reduce microgliosis, neuronal loss, and tau phosphorylation in a transgenic mouse model of tauopathy. *Journal of Neuroinflammation* **18**:312. DOI: https://doi.org/10.1186/s12974-021-02332-7, PMID: 34972522

Park JS, Lee J, Jung ES, Kim MH, Kim IB, Son H, Kim S, Kim S, Park YM, Mook-Jung I, Yu SJ, Lee JH. 2019. Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nature Communications* **10**:3090. DOI: https://doi.org/10.1038/s41467-019-11000-7, PMID: 31300647

Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10**:973–976. DOI: https://doi.org/10.1038/nmeth.2600, PMID: 23892895

Poduri A, Evrony GD, Cai X, Elhosary PC, Beroukhim R, Lehtinen MK, Hills LB, Heinzen EL, Hill A, Hill RS, Barry BJ, Bourgeois BFD, Riviello JJ, Barkovich AJ, Black PM, Ligon KL, Walsh CA. 2012. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron* **74**:41–48. DOI: https://doi.org/10.1016/j.neuron.2012.03.010, PMID: 22500628

Qu WS, Tian DS, Guo ZB, Fang J, Zhang Q, Yu ZY, Xie MJ, Zhang HQ, Lü JG, Wang W. 2012. Inhibition of EGFR/MAPK signaling reduces microglial inflammatory response and the associated secondary damage in rats after spinal cord injury. *Journal of Neuroinflammation* **9**:178. DOI: https://doi.org/10.1186/1742-2094-9-178, PMID: 22824323

Rauen KA. 2013. The RASopathies. *Annual Review of Genomics and Human Genetics* **14**:355–369. DOI: https://doi.org/10.1146/annurev-genom-091212-153523, PMID: 23875798

Reiner A, Yekutieli D, Benjamini Y. 2003. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**:368–375. DOI: https://doi.org/10.1093/bioinformatics/btf877, PMID: 12584122

Renton AE, Chiò A, Traynor BJ. 2014. State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience* **17**:17–23. DOI: https://doi.org/10.1038/nn.3584, PMID: 24369373

Réu P, Khosravi A, Bernard S, Mold JE, Salehpour M, Alkass K, Perl S, Tisdale J, Possnert G, Druid H, Frisén J. 2017. The Lifespan and turnover of microglia in the human brain. *Cell Reports* **20**:779–784. DOI: https://doi.org/10.1016/j.celrep.2017.07.004, PMID: 28746864

Rodriguez GA, Tai LM, LaDu MJ, Rebeck GW. 2014. Human APOE4 increases microglia reactivity at Aβ plaques in A mouse model of Aβ deposition. *Journal of Neuroinflammation* **11**:111. DOI: https://doi.org/10.1186/1742-2094-11-111, PMID: 24948358

Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, Chi H, Lin C, Holman K, Tsuda T. 1995. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **376**:775–778. DOI: https://doi.org/10.1038/376775a0, PMID: 7651536

Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. 2018. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **34**:511–513. DOI: https://doi.org/10.1093/bioinformatics/btx536, PMID: 28968714

Roy ER, Wang B, Wan Y-W, Chiu G, Cole A, Yin Z, Propson NE, Xu Y, Jankowsky JL, Liu Z, Lee VM-Y, Trojanowski JQ, Ginsberg SD, Butovsky O, Zheng H, Cao W. 2020. Type I interferon response drives neuroinflammation and synapse loss in Alzheimer disease. *The Journal of Clinical Investigation* **130**:1912–1930. DOI: https://doi.org/10.1172/JCI133737, PMID: 31917687

Sando SB, Melquist S, Cannon A, Hutton ML, Sletvold O, Saltvedt I, White LR, Lydersen S, Aasly JO. 2008. APOE epsilon 4 lowers age at onset and is a high risk factor for Alzheimer's disease; a case control study from central Norway. *BMC Neurology* **8**:9. DOI: https://doi.org/10.1186/1471-2377-8-9, PMID: 18416843

Sargin B, Choudhary C, Crosetto N, Schmidt MHH, Grundler R, Rensinghoff M, Thiessen C, Tickenbrock L, Schwäble J, Brandts C, August B, Koschmieder S, Bandi SR, Duyster J, Berdel WE, Müller-Tidow C, Dikic I, Serve H. 2007. Flt3-dependent transformation by inactivating c-Cbl mutations in AML. *Blood* **110**:1004–1012. DOI: https://doi.org/10.1182/blood-2007-01-066076, PMID: 17446348

Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, Rosi BL, Gusella JF, Crapper-MacLachlan DR, Alberts MJ. 1993. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43**:1467–1472. DOI: https://doi.org/10.1212/wnl.43.8.1467, PMID: 8350998

Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. 2009. PID: the pathway interaction database. *Nucleic Acids Research* **37**:D674–D679. DOI: https://doi.org/10.1093/nar/gkn653, PMID: 18832364

Scheltens P, Prins N, Lammertsma A, Yaqub M, Gouw A, Wink AM, Chu HM, van Berckel BNM, Alam J. 2018. An exploratory clinical study of p38α kinase inhibition in Alzheimer's disease. *Annals of Clinical and Translational Neurology* **5**:464–473. DOI: https://doi.org/10.1002/acn3.549, PMID: 29687023

Schnittger S, Bacher U, Eder C, Dicker F, Alpermann T, Grossmann V, Kohlmann A, Kern W, Haferlach C, Haferlach T. 2012. Molecular analyses of 15,542 patients with suspected BCR-ABL1-negative myeloproliferative disorders allow to develop a stepwise diagnostic workflow. *Haematologica* **97**:1582–1585. DOI: https://doi.org/10.3324/haematol.2012.064683, PMID: 22511494

Schöll M, Carter SF, Westman E, Rodriguez-Vieitez E, Almkvist O, Thordardottir S, Wall A, Graff C, Långström B, Nordberg A. 2015. Early astrocytosis in autosomal dominant Alzheimer's disease measured in vivo by multi-tracer positron emission tomography. *Scientific Reports* **5**:16404. DOI: https://doi.org/10.1038/srep16404, PMID: 26553227

Scholz SW, Bras J. 2015. Genetics underlying atypical parkinsonism and related neurodegenerative disorders. *International Journal of Molecular Sciences* **16**:24629–24655. DOI: https://doi.org/10.3390/ijms161024629, PMID: 26501269

Serrano-Pozo A, Li Z, Noori A, Nguyen HN, Mezlini A, Li L, Hudry E, Jackson RJ, Hyman BT, Das S. 2021. Effect of *APOE* alleles on the glial transcriptome in normal aging and Alzheimer's disease. *Nature Aging* **1**:919–931. DOI: https://doi.org/10.1038/s43587-021-00123-6, PMID: 36199750

Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day INM, Gaunt TR, Campbell C. 2015. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**:1536–1543. DOI: https://doi.org/10.1093/bioinformatics/btv009, PMID: 25583119

Smith MA, Choudhary GS, Pellagatti A, Choi K, Bolanos LC, Bhagat TD, Gordon-Mitchell S, Von Ahrens D, Pradhan K, Steeples V, Kim S, Steidl U, Walter M, Fraser IDC, Kulkarni A, Salomonis N, Komurov K, Boultwood J, Verma A, Starczynowski DT. 2019. U2AF1 mutations induce oncogenic IRAK4 isoforms and activate innate immune pathways in myeloid malignancies. *Nature Cell Biology* **21**:640–650. DOI: https://doi.org/10.1038/s41556-019-0314-5, PMID: 31011167

Soblet J, Limaye N, Uebelhoer M, Boon LM, Vikkula M. 2013. Variable somatic TIE2 mutations in half of sporadic venous malformations. *Molecular Syndromology* **4**:179–183. DOI: https://doi.org/10.1159/000348327, PMID: 23801934

Socci N. 2022. BIC-variants_pipeline. swh:1:rev:88967b8ddc1af03f95f09458d6c2d2073d5bed62. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:13c637223a69aa12214127c1c0acb2086fc87060;

origin=https://github.com/soccin/BIC-variants_pipeline;visit=swh:1:snp:4023f60335e776fe63506fe5de3b52e2 1570d2c3;anchor=swh:1:rev:88967b8ddc1af03f95f09458d6c2d2073d5bed62

Socci N. 2025. Variant-postprocess. swh:1:rev:8e72acb84500719afde95dcfcfd3b910c59127d2. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:f427243f5c4a23569c26502e8d02705422c092e4; origin=https://github.com/soccin/Variant-PostProcess;visit=swh:1:snp:a833ea5cad143d1c2005a12a599e407d 46d50369;anchor=swh:1:rev:8e72acb84500719afde95dcfcfd3b910c59127d2

Song X, Ma F, Herrup K. 2019. Accumulation of cytoplasmic DNA Due to ATM deficiency activates the microglial viral response system with neurotoxic consequences. *The Journal of Neuroscience* **39**:6378–6394. DOI: https:// doi.org/10.1523/JNEUROSCI.0774-19.2019, PMID: 31189575

Thapa DR, Tonikian R, Sun C, Liu M, Dearth A, Petri M, Pepin F, Emerson RO, Ranger A. 2015. Longitudinal analysis of peripheral blood T cell receptor diversity in patients with systemic lupus erythematosus by next-generation sequencing. *Arthritis Research & Therapy* **17**:132. DOI: https://doi.org/10.1186/s13075-015-0655-9, PMID: 26001779

The Gene Ontology C. 2019. The gene ontology resource: 20 years and still going strong. *Nucleic Acids Research* **47**:D330–D338. DOI: https://doi.org/10.1093/nar/gky1055, PMID: 30395331

Tidyman WE, Rauen KA. 2016. Expansion of the RASopathies. *Current Genetic Medicine Reports* **4**:57–64. DOI: https://doi.org/10.1007/s40142-016-0100-7, PMID: 27942422

Turner MR, Hardiman O, Benatar M, Brooks BR, Chio A, de Carvalho M, Ince PG, Lin C, Miller RG, Mitsumoto H, Nicholson G, Ravits J, Shaw PJ, Swash M, Talbot K, Traynor BJ, Van den Berg LH, Veldink JH, Vucic S, Kiernan MC. 2013. Controversies and priorities in amyotrophic lateral sclerosis. *The Lancet. Neurology* **12**:310–322. DOI: https://doi.org/10.1016/S1474-4422(13)70036-X, PMID: 23415570

Walter MJ, Ding L, Shen D, Shao J, Grillot M, McLellan M, Fulton R, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Kandoth C, Baty J, Westervelt P, DiPersio JF, Mardis ER, Wilson RK, Ley TJ, Graubert TA. 2011. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia* **25**:1153–1158. DOI: https://doi. org/10.1038/leu.2011.44, PMID: 21415852

Ward PS, Patel J, Wise DR, Abdel-Wahab O, Bennett BD, Coller HA, Cross JR, Fantin VR, Hedvat CV, Perl AE, Rabinowitz JD, Carroll M, Su SM, Sharp KA, Levine RL, Thompson CB. 2010. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**:225–234. DOI: https://doi.org/10.1016/j.ccr.2010.01.020, PMID: 20171147

Wei W, Keogh MJ, Aryaman J, Golder Z, Kullar PJ, Wilson I, Talbot K, Turner MR, McKenzie CA, Troakes C, Attems J, Smith C, Sarraj SA, Morris CM, Ansorge O, Jones NS, Ironside JW, Chinnery PF. 2019. Frequency and signature of somatic variants in 1461 human brain exomes. *Genetics in Medicine* **21**:904–912. DOI: https://doi. org/10.1038/s41436-018-0274-3, PMID: 30214067

Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**:15. DOI: https://doi.org/10.1186/s13059-017-1382-0, PMID: 29409532

Xi T, Jones IM, Mohrenweiser HW. 2004. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. *Genomics* **83**:970–979. DOI: https://doi.org/10.1016/j.ygeno.2003.12.016, PMID: 15177551

Xiong Y, Song D, Cai Y, Yu W, Yeung YG, Stanley ER. 2011. A CSF-1 receptor phosphotyrosine 559 signaling pathway regulates receptor ubiquitination and tyrosine phosphorylation. *The Journal of Biological Chemistry* **286**:952–960. DOI: https://doi.org/10.1074/jbc.M110.166702, PMID: 21041311

Yang S, Zheng X, Lu C, Li GM, Allis CD, Li H. 2016. Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. *Genes & Development* **30**:1611–1616. DOI: https://doi.org/10.1101/gad.284323.116, PMID: 27474439

Yu W, Chen J, Xiong Y, Pixley FJ, Dai XM, Yeung YG, Stanley ER. 2008. CSF-1 receptor structure/function in MacCsf1r-/- macrophages: regulation of proliferation, differentiation, and morphology. *Journal of Leukocyte Biology* **84**:852–863. DOI: https://doi.org/10.1189/jlb.0308171, PMID: 18519746

Zafra MP, Schatoff EM, Katti A, Foronda M, Breinig M, Schweitzer AY, Simon A, Han T, Goswami S, Montgomery E, Thibado J, Kastenhuber ER, Sánchez-Rivera FJ, Shi J, Vakoc CR, Lowe SW, Tschaharganeh DF, Dow LE. 2018. Optimized base editors enable efficient editing in cells, organoids and mice. *Nature Biotechnology* **36**:888–893. DOI: https://doi.org/10.1038/nbt.4194, PMID: 29969439

Zannini L, Delia D, Buscemi G. 2014. CHK2 kinase in the DNA damage response and beyond. *Journal of Molecular Cell Biology* **6**:442–457. DOI: https://doi.org/10.1093/jmcb/mju045, PMID: 25404613

Zar J. 2010. *Biostatistical Analysis*. Prentice-Hall.

Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, Hellmann MD, Barron DA, Schram AM, Hameed M, Dogan S, Ross DS, Hechtman JF, DeLair DF, Yao J, Mandelker DL, et al. 2017. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nature Medicine* **23**:703–713. DOI: https://doi.org/10.1038/nm.4333, PMID: 28481359

Zhong A, Li M, Zhou T. 2020. Protocol for the generation of human pluripotent reporter cell lines using CRISPR/ Cas9. *STAR Protocols* **1**:100052. DOI: https://doi.org/10.1016/j.xpro.2020.100052, PMID: 33073252

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **10**:1523. DOI: https://doi.org/10.1038/s41467-019-09234-6, PMID: 30944313