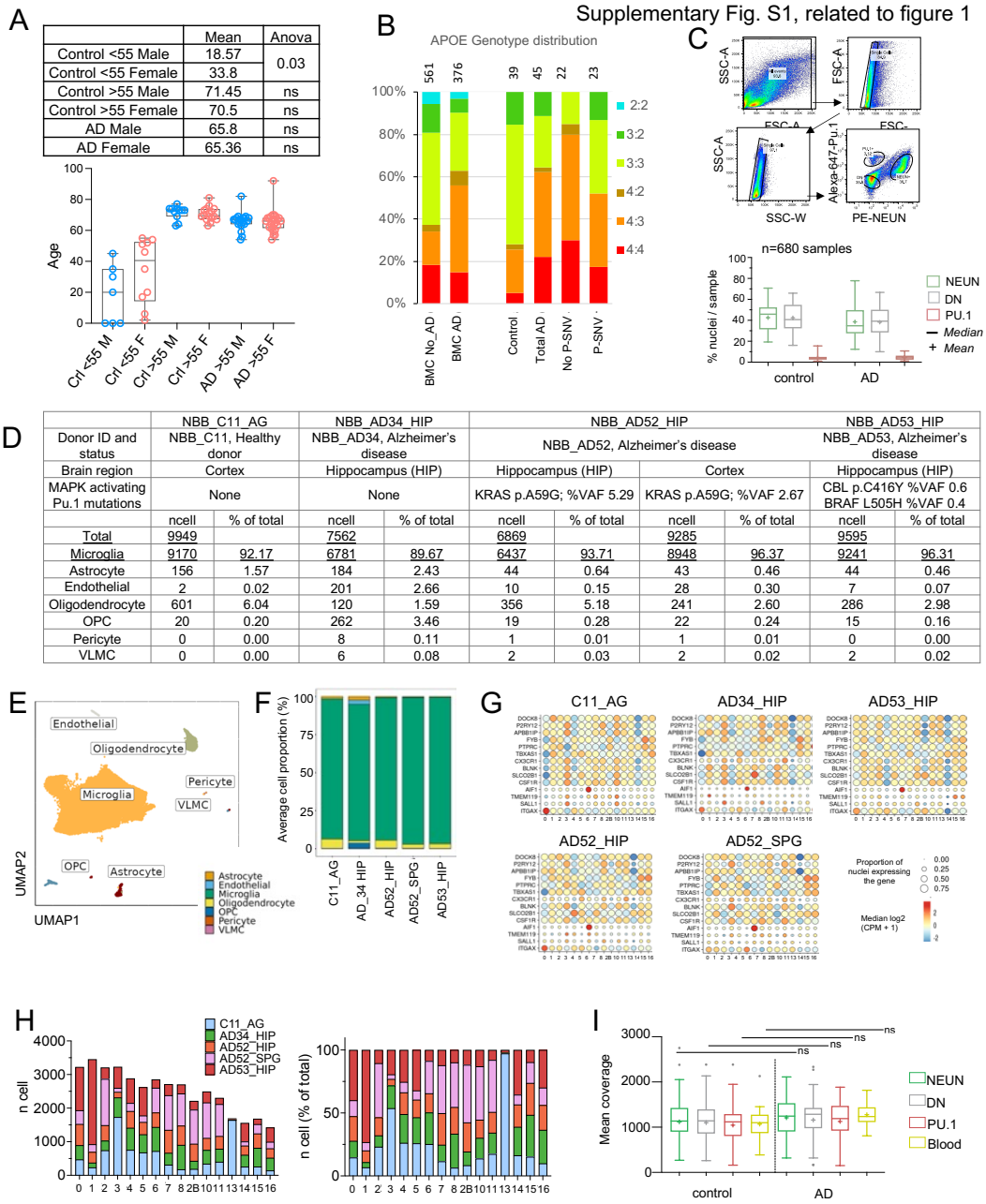
Extended Data

Supplementary Figures 1-6

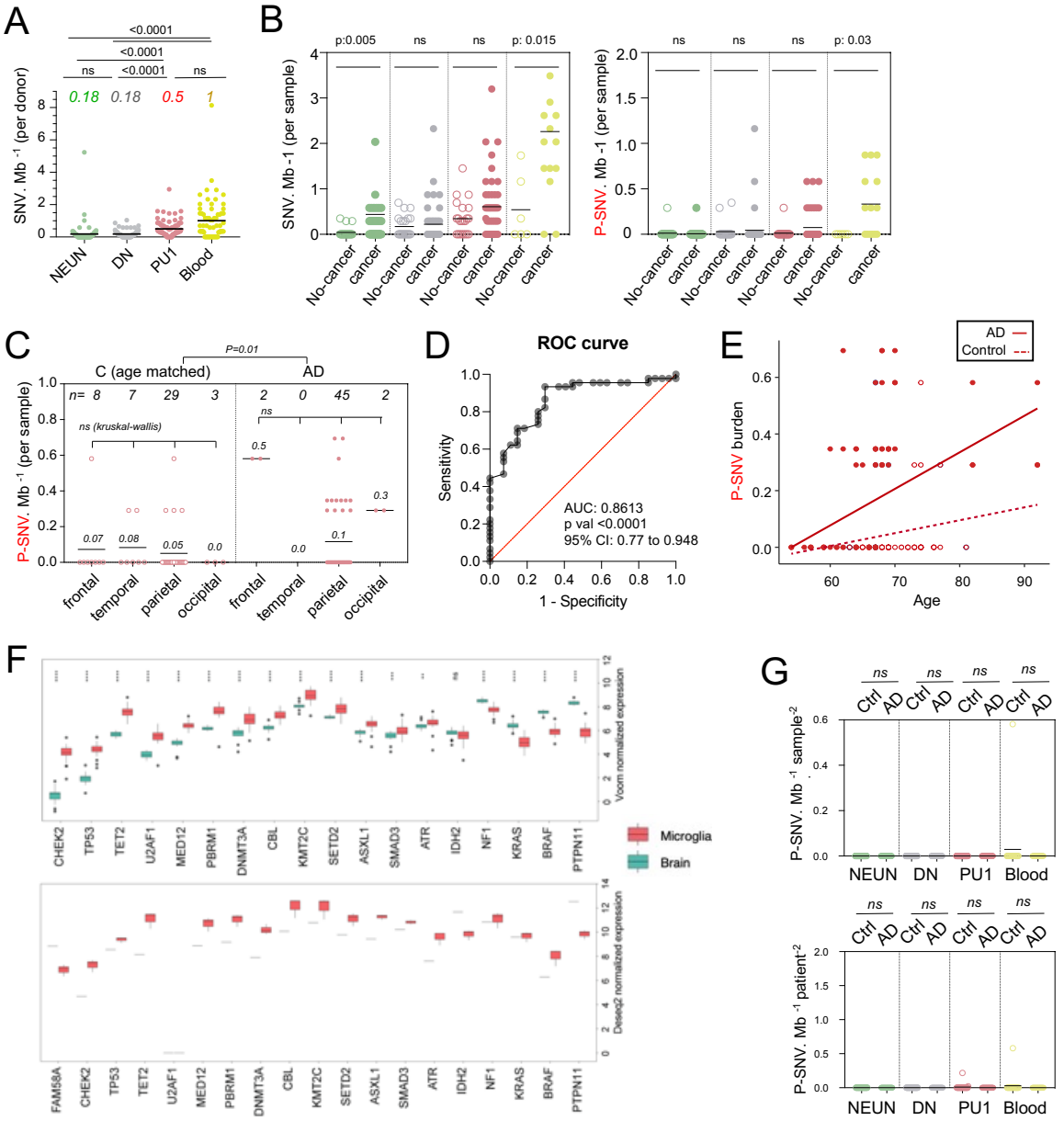


Donor ID and status	NBB_C11_AG NBB_C11, Healthy donor		NBB_AD34_HIP NBB_AD34, Alzheimer's disease		NBB_AD52_HIP NBB_AD52, Alzheimer's disease				NBB_AD53_HIP NBB_AD53, Alzheimer's disease	
Brain region	Cortex		Hippocampus (HIP)		Hippocampus (HIP)		Cortex		Hippocampus (HIP)	
MAPK activating Pu.1 mutations	None		None		KRAS p.A59G; %VAF 5.29		KRAS p.A59G; %VAF 2.67		CBL p.C416Y %VAF 0.6 BRAF L505H %VAF 0.4	
	ncell	% of total	ncell	% of total	ncell	% of total	ncell	% of total	ncell	% of total
Total	9949		7562		6869		9285		9595	
Microglia	9170	92.17	6781	89.67	6437	93.71	8948	96.37	9241	96.31
Astrocyte	156	1.57	184	2.43	44	0.64	43	0.46	44	0.46
Endothelial	2	0.02	201	2.66	10	0.15	28	0.30	7	0.07
Oligodendrocyte	601	6.04	120	1.59	356	5.18	241	2.60	286	2.98
OPC	20	0.20	262	3.46	19	0.28	22	0.24	15	0.16
Pericyte	0	0.00	8	0.11	1	0.01	1	0.01	0	0.00
VLMC	0	0.00	6	0.08	2	0.03	2	0.02	2	0.02

Supplementary Figure 1. Quality control for DNA analysis and snRNAseq. (A) patients characterisitic: age and sex distribution of young controls, age-matched controls and AD patients, Statistics: 2 way Anova. **(B)** Distribution of APOE genotype in a historical cohort of controls and AD patients [49] (Left) and the present series (Right) of Control, AD and AD without and with pathogenic (P-SNV) microglia variants. Numbers on top of the bars show patient number in each group. **(C)** Sorting strategy to separate PU.1^{+}, NEUN^{+} and DN nuclei from post-mortem brain samples. Boxplot represents relative frequencies, median, mean, 25-75th quartiles (boxes) and minimum/maximum (whiskers) of nuclei for each cell type in controls (n=63 brain samples) and AD patients (n=99 brain samples). **(D)** SnRNA-seq analysis of Facs-sorted PU.1+ nuclei from 4 donors. Table indicate donor characteristics, number of nuclei

analyzed after quality control (see methods) and cell types as determined by unsupervised clustering of normalized and integrated gene expression of nuclei from 5 PU.1^{+} samples. **(E)** UMAP representation of cell types from (C). **(E)** Cell proportion plot of the 5 PU.1^{+} samples from (C). **(F)** Boxplot showing the coverage of targeted DNA deep sequencing per cell type in AD and control samples. Box plots show median (+ mean) and 25th and 75th percentiles; whiskers extend to the largest and smallest values. Dots show outliers. **(G)** Expresion of microglia markers by sn-RNAseq across samples and clusters. **(H)** Number (TOP) and proportiton (BOTTOM) of cells from each sample, per-cluster. **(I)** Boxplot showing the coverage of targeted DNA deep sequencing per cell type in AD and control samples. Box plots show median (+ mean) and 25th and 75th percentiles; whiskers extend to the largest and smallest values. Dots show outliers.

Supplementary Fig. S2, related to figure 2



Supplementary Figure 2. Analysis of pathogenic variants. (A) Number of SNV per Mb, per donor, and cell types. Each dot represents the mean of a donor. NeuN n=226, DN n=229, PU.1 n=225, Blood n=66). Values (color, *italics)* indicate the mean number of variants /Mb per cell type. Statistics: *p-value* are calculated by Kruskal–Wallis test and Dunn's test for multiple comparisons. **(B)** Number of SNV (Left) and P-SNV (right) per Mb in controls (age-matched with the patients) with or without cancer per sample and cell types. Each dot represents a sample. Statistics: *p-values* within each group are calculated with Kruskal–Wallis, multiple comparisons. **(C)** Number of SNV per Mb in PU.1 samples across cortical samples, of age-matched controls (n=27) and AD patients (n=45). Each dot represents a sample. Statistics: *p-values* within each group are calculated with Kruskal–Wallis, multiple comparisons. **(D)** Receiver operating characteristic (ROC) curve showing the accuracy of the multivariate logistic regression model in predicting the association of AD and the presence or not of pathogenic variants in PU.1^{+} nuclei. Note: non-parametric tests were used as data did not follow a normal distribution (D'Agostino-Pearson normality test). **(E)** Observed SNV burden for Fig. 2H. **(F)** Expression of pathogenic genes in microglia and whole brain tissue,

reported in [33] (TOP, sorted microglia n= 39 and whole brain n=16) and [34] (BOTTOM, sorted microglia n= 3 and whole brain n=1. **(G)** Graph depicts mean number of pathogenic variants in a group of control genes not expressed by the brain or by microglia (see Supplementary Table 3), per Mb, and samples (LEFT) and donor (RIGHT), in NEUN, DN, PU.1 nuclei and matching blood from all controls and AD patients. Each dot represents the mean for each donor. Statistics: *p-values* are calculated with unpaired two-tailed Mann-Whitney U test comparing AD to controls.

Supplementary Fig. S3, related to Figure 2 and 3

Driver mutations (BRAINPACT)

Patient #	44		47	62		53			60	61		30	38			43	39	52	35	49	36	23		57	24		26	29	50		19	37	59
Gene_variant	CBL p.I383M	IDH2 p.R140Q	CBL p.C404Y	CBL p.C404Y	PBRM1 c.996-7T>A	CBL p.C416S	BRAF p.L505H	ATR c.6318A>G	CBL p.C384Y	CBL p.R420Q	TET2 p.Q1627*	RIT1 p.F82L	RIT1 p.M90I	NF1 p.L2442*	SMC1A p.X285_splice	PTPN11 p.T73I	TEK p.R1099*	KRAS p.A59G	U2AF1 p.S34F	SMAD3 p.R373C	TP53 c.783-1G>A	CHEK2 p.X107_splice	ASXL1 p.Q588*	CHEK2 p.R346H	KMT2C p.Q1764*	TET2 p.S585*	TET2 p.Y1421*	DNMT3A p.R635L	DNMT3A p.Q842*	DNMT3A p.R736C	MED12 c.5400+6C>T	SETD2 p.Y1666C	DNMT3A p.V636M
TKR/MAPK	x		x	x		x	x		x	x		x	x	x		x	x	x	x														
CML	x		x	x		x	x		x	x		x	x	x		x	x	x		x	x												
DNA/Histone methylation		x			x						x						x						x		x	x	x	x	x	x	x	x	x
DNA repair								x						x							x	x		x									
expressed in microglia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
age	67		64	82		68			68	92		60	62			68	65	67	67	69	64	68		69	70		67	68	70		69	69	64
sex	M		F	M		M			F	F		F	F			F	M	M	M	F	M	F		M	F		F	M	F		F	M	M
notes												(1)						(2)															
ApoE	33		33	33		33			43	33		33	43			43	43	44	44	32	44	32		43	43		32	43	33		44	43	33
braak	6		6	2		6			6	2		5	6			6	6	5	6	6	6	6		5	6		6	6	6		5	4	6
CERAD	C		C	C		C			C	C		C	C			C	C	C	C	C	C	C		C	C		C	C	C		C	C	C
Lewis bodies	LBV		+	-		LBV			-	-		+	LBV			LBV	-	LBV	LBV	+	+	-			LBV		-	-	LBV		LBV	LBV	LBV
amyloid angiopathy	-		-	+		-			-	-		-	-			-	-	+	-	-	-	-		+	-		-	-	-		-	-	-
tau	-		-	-		-			-	-		-	-			-	-	ARTAG	-	-	-	-		T	-		-	-	-		-	-	-
alpha synucleinopathy	-		-	-		-			-	-		-	-			-	-	-	-	-	-	-		-	-		-	-	-		-	-	-
Brain regions with mut	H	H	H	FC/OC/CN	FC	H	H	SPG	H	FC	FC/OC/CN	H	H	SPG	H	H	SPG	H/SPG	H/SPG	H/SPG	H	H/SPG	H/SPG	H	H/SPG	SPG	H/SPG	H/SPG	H/SPG	H/SPG	SPG	H	SPG
GOF (G) / LOF (L)	L	G	L	L	L	L	G	?	L	L	L	G	G	L	G	G	G	G	G	L	L	L	L	L	L	L	L	?	L	L	?	L	?
gnomeAD	na	0	0	0	0	na	na	na	na	0	na	na	na	na	na	na	na	na	0	na	na	0	na	na	na	na	na	0	na	0	0	0	0
VAF (%,brain)	5.7	0.9	1.1	1.1	2.4	1.2	0.8	1.1	1.7	1.0	1.2	0.8	1.1	1.9	1	1.2	1.2	4.8	1.0	1.8	1.1	5.6	1.2	1.4	4.8	0.7	0.6	2.1	0.7	1.1	1.8	1.8	1.2
VAF (%,blood)	0	0	0	na	0	0	0	0	0	na	na	0	0	0	0	0	0	2.34	0	1.21	0	0	0	0	6.5	0	0	na	1.4	0.6	0	0	0
VAF (%,DN)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAF (%,NEUN)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CADD score	22	28	25	25	0.8	25	30	12	25	28	38	28	25	46	23	32	54	24.2	32	35	24	27	37	35	38	29	41	35	43	34	1.4	30	16

Canditates (WES)

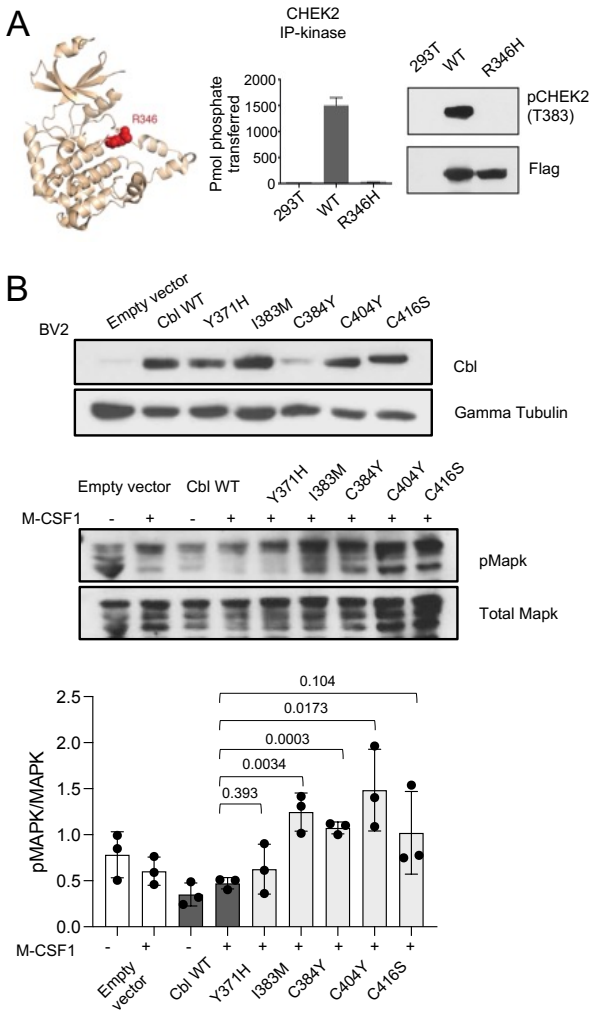
Patient #	58	45	21	18	48	56	34	28
Gene_variant	ARHGAP9 p.E595G	AFF1 p.D58E	CHD8 p.R303W	DIXDC1 p.I60L	RMDN2 p.E104V	ZUFSP p.T317K	FAM214A p.I1023L	ARHGEF26 p.R863S
TKR/MAPK	x		x	x				x
CML								
DNA/Histone methylation		x						
DNA repair						x		
expressed in microglia	+	+	+	-	+	na	+	-
age	67	61	62	67	65	56	59	57
sex	F	F	F	M	F	M	M	F
notes	(3)		(4)	(5)	(6)	(7)		
ApoE	43	44	43	44	43	43	43	43
braak	6	6	6	5	5	5	5	6
CERAD	C	C	C	B	C	C	C	C
Lewis bodies	LBV	-	-	-	+	LBV	LBV	-
amyloid angiopathy	-	-	-	+	-	-	-	-
tau	-	-	-	-	-	-	-	-
alpha synucleinopathy	-	-	-	-	+	-	-	-
Brain regions with mut	SPG	H	H	H	H	H	H	H
GOF (G) / LOF (L)	?	?	?	?	?	?	?	?
gnomeAD	na	0	na	na	na	na	na	na
VAF (%,brain)	6.7	6	5.7	5.3	6.3	5.3	7.2	5
VAF (%,blood)	na	na	na	na	na	na	na	na
VAF (%,DN)	na	na	na	na	na	na	na	na
VAF (%,NEUN)	na	na	na	na	na	na	na	na
CADD score	32	24	32	24	28	27	24	27

No candidate variant detected

Patient #	20	25	27	31	32	33	41	42	51	54	55	22	40	46
Gene_variant														
TKR/MAPK														
CML														
DNA/Histone methylation														
DNA repair														
expressed in microglia														
age	54	65	68	54	59	68	64	66	63	66	66	66	57	63
sex	M	M	M	F	F	M	F	M	M	F	F	F	F	F
notes				8)										
ApoE	43	33	33	43	44	44	44	32	44	33	32	42	43	43
braak	6	6	6	6	5	6	4	6	4	6	6	6	6	6
CERAD	C	B	C	C	C	C	C	C	C	C	C	C	C	C
Lewis bodies	LBV	LBV	LBV	-	-	LBV	-	-	LBV	-	-	LBV	-	-
amyloid angiopathy	-	-	-	-	-	+	+	+	-	+	-	-	-	-
tau	-	-	-	-	-	-	-	-	-	-	-	-	-	-
alpha synucleinopathy	-	-	-	-	-	+	-	-	-	-	-	-	-	-
Brain regions with mut	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GOF (G) / LOF (L)														
gnomeAD														
VAF (%,brain)														
VAF (%,blood)														
VAF (%,DN)														
VAF (%,NEUN)														
CADD score														

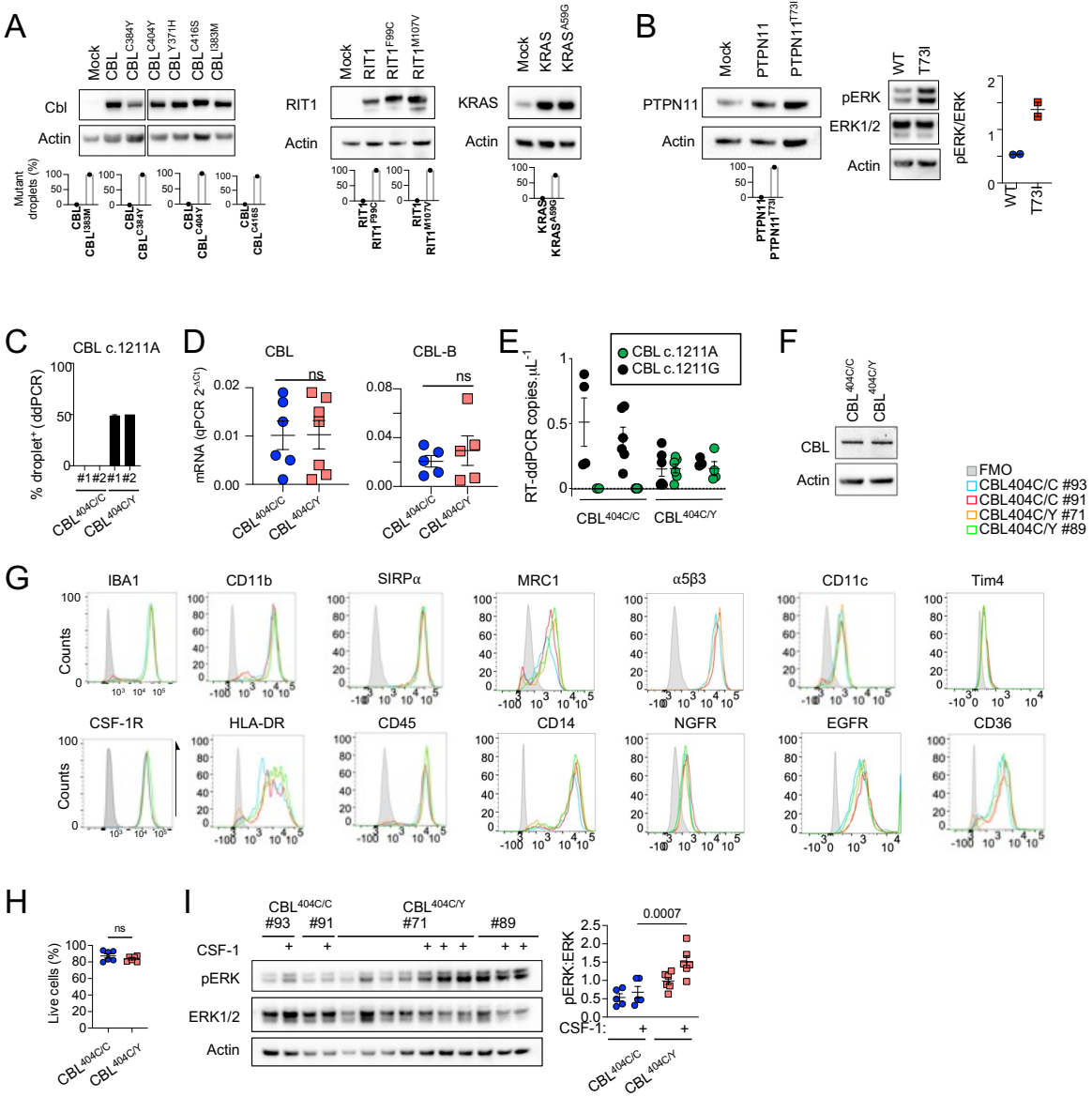
Supplementary Figure 3. Summary of AD patients characteristics and pathogenic variants. Table shows for all AD patients studied, the detection of pathogenic variants by TDS, candidates identified by WES, categories of gene functions (MAPK pathway, DNA repair, DNA/Histone methylation), expression in microglia, and patient information (age/sex/Apoe genotype/braak status/CERAD score/presence of lewis bodies/presence of amyloid angiopathy). #Brain regions: number of brain regions where variant was detected. GOF (G, Gain of Function) / LOF (L, Loss of Function) as reported in bibliography (see manuscript for references). gnomeAD shows the minor allele frequency of each variant in the population. VAF: variant allelic frequency (%) by BRAIN-PACT in brain cell types and matching-blood when available. CADD score (Combined Annotation Dependent Depletion) of each variant. Notes: (1) Trisomy 21, Down syndrome. (2) familial history of AD, no variant in AD associated genes. (3) MAPK docking protein. (4) cooperative interaction with ELK1 on chromatin. (5) inhibits JNK activation, murine KO has a neurological phenotype[50] (6) microtubule binding, involved in b-amyloid aggregation. (7) DNA repair gene. (8) Mosaic trisomy 21.

Supplementary Fig. S4, related to Figure 3



Supplementary Figure 4. Functional analysis of variants in HEK293 and BV2 cell lines (A) Quantification of Western blot from cell lysates from HEK293T cells expressing WT of mutant CBL alleles and stimulated with EGF or control were probed with antibodies against Phospho-p44/42 MAPK (Erk 1/2, Thr202/Tyr204, (pMAPK)), total MAPK (p44/42 MAPK, Erk1/2, (MAPK)), and HA-tag (BOTTOM). n= 4 independent experiments. Statistic: Student t-test. **(B)** HEK293T cells expressing Flag-RIT1 (WT and mutants) were treated -/+ 20% FBS before harvesting and Lysates were probed with antibodies against Phospho-p44/42 MAPK (Erk 1/2, Thr202/Tyr204, (pMAPK)), total MAPK (p44/42 MAPK, Erk1/2, (MAPK)), and Flag. n= 5 independent experiments. Statistic: Student t-test. **(C)** Flag-tagged RIT1 constructs were expressed in HEK293T cells. Lysates were used in pulldown reactions with immobilized GST-PAK1-CRIB domain and in immunoprecipitation reactions with Cdc42 antibody. Bound RIT1 was measured by anti-Flag Western blotting. Lysates were also analyzed by anti-Flag and anti-MAPK Western blotting. **(D)** CHEK2 R346H is a loss-of-function mutant. The R346H variant is located within the catalytic loop of the protein kinase domain and shown in red on the 3D structure of CHEK2 kinase domain (pdb code: 2cn5) (LEFT). CHEK2 R346 Lysates from HEK293T cells expressing Flag-WT or CHEK2 R346 were probed with antibodies that recognizes the auto phosphorylated and activated form of CHEK2 and Flag (MIDDLE). Flag-tagged WT and R346H CHK2 were expressed in HEK293T cells, proteins were isolated by immunoaffinity capture using anti-Flag resin. CHK2 activity was measured with [^{32}P]-labeled ATP and a synthetic CHEK2 substrate peptide. Wild-type CHEK2 showed robust activity, while the R346H mutant was inactive (RIGHT). **(E)** Western-blot analysis of CBL expression (TOP), pMAPK and total MAPK (MIDDLE) and respective quantification (BOTTOM) in BV2 cell lines transduced with empty vector, CBL^{WT}, CBL^{Y371H}, CBL^{I383M}, CBL^{C384Y}, CBL^{C404Y} and CBL^{C416S}. For MIDDLE panel, cells were treated with M-CSF1 100 ng/ml for 5 min. Statistics: *p-values* are calculated with t-test. n=3.

Supplementary Fig. S5, related to figure 4

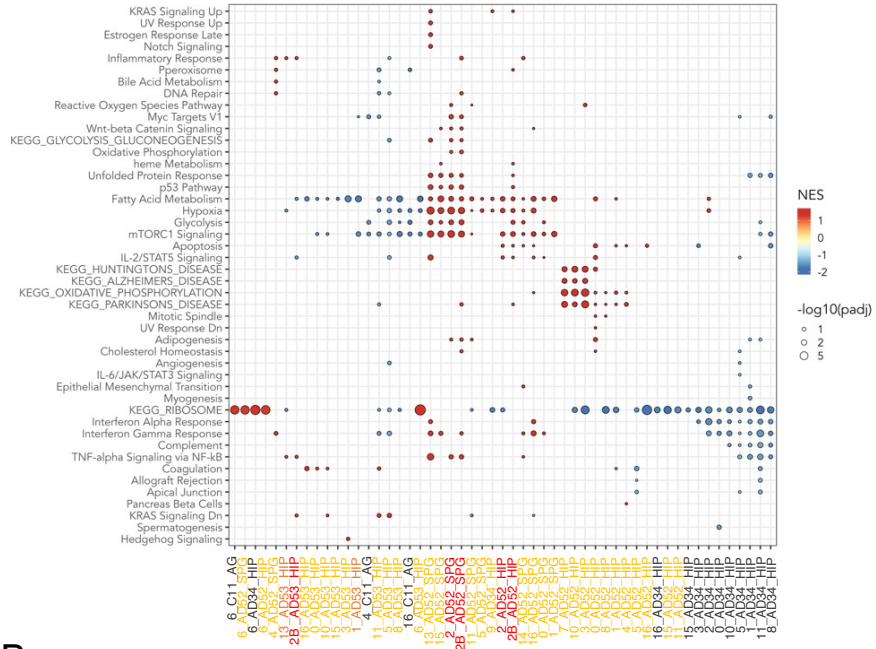


Supplementary Figure 5. Analysis of mouse and human microglia-like cells. (A) Western-blot analysis of CBL, RIT1, and KRAS expression in lysates from a growth factor-dependent macrophage cell line expressing CBL^{WT}, CBL^{I383M}, CBL^{C384Y}, CBL^{C404Y}, CBL^{C416S}, CBL^{R420Q}, RIT1^{WT}, RIT1^{F99C}, RIT1^{M107V}, KRAS^{WT} and KRAS^{A59G} alleles (TOP), and ddPCR analysis of wt and mutant alleles in DNA from the same cell lines (BOTTOM). (**B**) Western-blot analysis of PTPN11 expression and phospho- and total-ERK in lysates from growth factor-dependent macrophage cell line expressing PTPN11^{WT} or PTPN11^{T73I} alleles, and ddPCR analysis of wt and variant alleles in DNA from the same lines. **(C)** Genomic DNA ddPCR of 2 independent hiPSC clones (#1 and #2) of CBL$^{404C/Y}$ heterozygous mutant carrying the c.1211G/A transition on one allele and 2 independent isogenic control CBL$^{404C/C}$ clones all obtained by prime editing. **(D)** CBL and CBL-B mRNA expression assessed by Taqman assay in CBL$^{404C/C}$ and CBL$^{404C/Y}$ iPSC-derived microglia-like cells. Unpaired t-test. **(E)** RT-ddPCR of CBL reference allele (CBL c.1211A) and CBL variant CBL c.1211G transcripts in CBL$^{404C/C}$ and CBL$^{404C/Y}$ iPSC-derived macrophages. n=4-6 independent experiments. **(F)** Western-blot analysis of CBL expression in lysates from CBL$^{404C/C}$ and CBL$^{404C/Y}$ iPSC-derived microglia-like cells. **(G)** Representative flow cytometry analysis of the expression of surface receptors

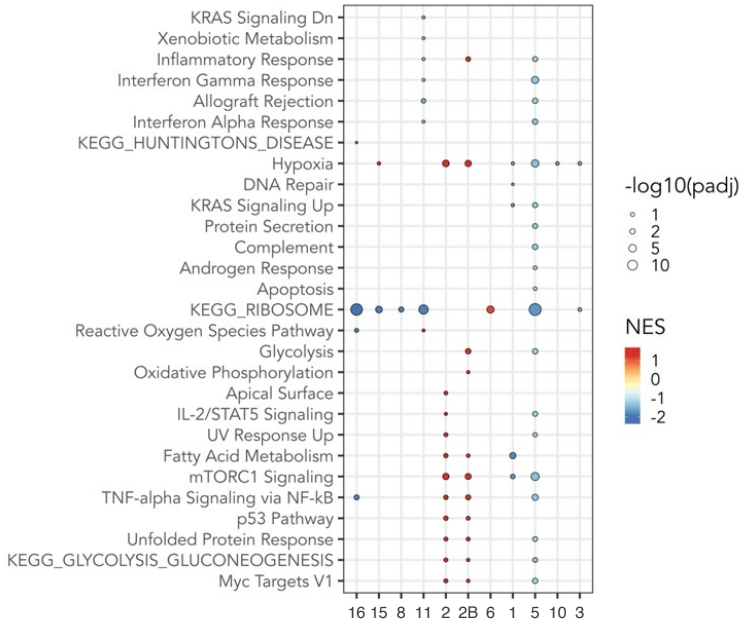
and Iba1 in $CBL^{404C/C}$ and $CBL^{404C/Y}$ cells (n=3) **(H)** Viability of $CBL^{404C/C}$ and $CBL^{404C/Y}$ iPSC-derived microglia-like cells estimated by flow cytometry analysis after DAPI staining. Unpaired t-test. n=6. **(I)** Western-blot analysis and quantification of phospho- and total-ERK proteins in lysates from $CBL^{404C/C}$ and $CBL^{404C/Y}$ iPSC-derived microglia-like cells untreated or re-stimulated with CSF-1 cells (5 min, 100 ng/mL). (Two-way ANOVA, n=6-7).

Supplementary Fig. S6, related to figure 5

A



B



Supplementary Figure 6. snRNAseq analysis of microglia. (A) Dot plot represents the significant pathways by GSEA analysis of HALLMARK and KEGG pathways of snRNAseq analysis of microglia, by samples and clusters. Genes from all samples are pre-ranked per cluster using differential expression analysis with SCANPY [37] and the Wilcoxon rank-sum method. Statistical analysis were performed using the fgseaMultilevel function in fgsea R package [38]f or HALLMARK and KEGG pathways. Only HALLMARK and KEGG gene sets with p-value < 0.05 and adjusted p-value < 0.25 are visualized, using ggpubr and ggplot2 [39] R package. **(B)** Dot plot represents the same GSEA analysis of HALLMARK and KEGG pathways enriched in snRNAseq microglia clusters as in A, but samples from all donors are grouped by microglia clusters.

Supplemental Tables 1-9:

Supplementary Table 1: Characteristics of AD and control donors and samples

Supplementary Table 2: Targeted-Sequencing gene panel

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

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

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

Supplementary Table 6: Predicted deleterious variants by WES

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

Supplementary Table 8: RNAseq analysis of hIPSC derived microglial-like cells: Differential expressed genes and GSEA analysis.

Supplementary Table 9: Single nuclei RNAseq analysis of control and AD microglia: Differential expressed genes per clusters and GSEA analysis.

Supplemental References

1 Dubois, B. *et al.* Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746 (2007). https://doi.org:10.1016/S1474-4422(07)70178-3

2 Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* **82**, 239-259 (1991). https://doi.org:10.1007/BF00308809

3 Braak, H. & Braak, E. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol Aging* **16**, 271-278; discussion 278-284 (1995). https://doi.org:10.1016/0197-4580(95)00021-6

4 McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. & Stadlan, E. M. 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 (1984). https://doi.org:10.1212/wnl.34.7.939

5 McKhann, G. M. *et al.* 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. *Alzheimers Dement* **7**, 263-269 (2011). https://doi.org:10.1016/j.jalz.2011.03.005

6 Cheng, D. T. *et al.* 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. *J Mol Diagn* **17**, 251-264 (2015). https://doi.org:10.1016/j.jmoldx.2014.12.006

7 Karch, C. M., Cruchaga, C. & Goate, A. M. Alzheimer's disease genetics: from the bench to the clinic. *Neuron* **83**, 11-26 (2014). https://doi.org:10.1016/j.neuron.2014.05.041

8 Karch, C. M. & Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry* **77**, 43-51 (2015). https://doi.org:10.1016/j.biopsych.2014.05.006

9 Turner, M. R. *et al.* Controversies and priorities in amyotrophic lateral sclerosis. *Lancet Neurol* **12**, 310-322 (2013). https://doi.org:10.1016/s1474-4422(13)70036-x

10 Bras, J., Guerreiro, R. & Hardy, J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat Rev Neurosci* **13**, 453-464 (2012). https://doi.org:10.1038/nrn3271

11 Renton, A. E., Chiò, A. & Traynor, B. J. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* **17**, 17-23 (2014). https://doi.org:10.1038/nn.3584

12 Ferrari, R. *et al.* A genome-wide screening and SNPs-to-genes approach to identify novel genetic risk factors associated with frontotemporal dementia. *Neurobiol Aging* **36**, 2904.e2913-2926 (2015). https://doi.org:10.1016/j.neurobiolaging.2015.06.005

13 Kouri, N. *et al.* Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nat Commun* **6**, 7247 (2015). https://doi.org:10.1038/ncomms8247

14 Scholz, S. W. & Bras, J. Genetics Underlying Atypical Parkinsonism and Related Neurodegenerative Disorders. *Int J Mol Sci* **16**, 24629-24655 (2015). https://doi.org:10.3390/ijms161024629

15 Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**, 989-993 (2014). https://doi.org:10.1038/ng.3043

16 Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880-886 (2015). https://doi.org:10.1126/science.aaa6806

17 Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911-917 (2018). https://doi.org:10.1126/science.aau3879

18 Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483-1489 (2015). https://doi.org:10.1126/science.aab4082

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

20 Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985 (2014). https://doi.org:10.1093/nar/gkt1113

21 Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017** (2017). https://doi.org:10.1200/PO.17.00011

22 Tidyman, W. E. & Rauen, K. A. Expansion of the RASopathies. *Curr Genet Med Rep* **4**, 57-64 (2016). https://doi.org:10.1007/s40142-016-0100-7

23 Rauen, K. A. The RASopathies. *Annu Rev Genomics Hum Genet* **14**, 355-369 (2013). https://doi.org:10.1146/annurev-genom-091212-153523

24 Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **23**, 703-713 (2017). https://doi.org:10.1038/nm.4333

25 Bardou, P., Mariette, J., Escudie, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**, 293 (2014). https://doi.org:10.1186/1471-2105-15-293

26 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019). https://doi.org:10.1038/s41467-019-09234-6

27 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009). https://doi.org:10.1038/nprot.2008.211

28 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009). https://doi.org:10.1093/nar/gkn923

29 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000). https://doi.org:10.1038/75556

30 The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338 (2019). https://doi.org:10.1093/nar/gky1055

31 Schaefer, C. F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res* **37**, D674-679 (2009). https://doi.org:10.1093/nar/gkn653

32 Zar, J. Biostatistical Analysis: New Jersey: Prentice-Hall. 523 (2010).

33 Galatro, T. F. *et al.* Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat Neurosci* **20**, 1162-1171 (2017). https://doi.org:10.1038/nn.4597

34 Gosselin, D. *et al.* An environment-dependent transcriptional network specifies human microglia identity. *Science* **356** (2017). https://doi.org:10.1126/science.aal3222

35 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8

36 Azizi, E. *et al.* Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* **174**, 1293-1308 e1236 (2018). https://doi.org:10.1016/j.cell.2018.05.060

37 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* **19**, 15 (2018). https://doi.org:10.1186/s13059-017-1382-0

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

39 Hadley, W. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

40 Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R. & Campbell, C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **34**, 511-513 (2018). https://doi.org:10.1093/bioinformatics/btx536

41 Xiong, Y., Song, D., Cai, Y., Yu, W., Yeung, Y. G. & Stanley, E. R. A CSF-1 receptor phosphotyrosine 559 signaling pathway regulates receptor ubiquitination and tyrosine phosphorylation. *J Biol Chem* **286**, 952-960 (2011). https://doi.org:10.1074/jbc.M110.166702

42 Lachmann, N. *et al.* Large-scale hematopoietic differentiation of human induced pluripotent stem cells provides granulocytes or macrophages for cell replacement

therapies. *Stem Cell Reports* **4**, 282-296 (2015). https://doi.org:10.1016/j.stemcr.2015.01.005

43 Lu, B. *et al.* A Transcription Factor Addiction in Leukemia Imposed by the MLL Promoter Sequence. *Cancer Cell* **34**, 970-981 e978 (2018). https://doi.org:10.1016/j.ccell.2018.10.015

44 Ng, P. K. *et al.* Systematic Functional Annotation of Somatic Mutations in Cancer. *Cancer Cell* **33**, 450-462 e410 (2018). https://doi.org:10.1016/j.ccell.2018.01.021

45 Kim, E. *et al.* Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles. *Cancer Discov* **6**, 714-726 (2016). https://doi.org:10.1158/2159-8290.CD-16-0160

46 Zafra, M. P. *et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol* **36**, 888-893 (2018). https://doi.org:10.1038/nbt.4194

47 Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* **10**, 973-976 (2013). https://doi.org:10.1038/nmeth.2600

48 Zhong, A., Li, M. & Zhou, T. Protocol for the Generation of Human Pluripotent Reporter Cell Lines Using CRISPR/Cas9. *STAR Protoc* **1** (2020). https://doi.org:10.1016/j.xpro.2020.100052

49 Sando, S. B. *et al.* 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 Neurol* **8**, 9 (2008). https://doi.org:10.1186/1471-2377-8-9

50 Martin, P. M. *et al.* 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 (2018). https://doi.org:10.1038/mp.2016.184