ARTICLE

Non-coding and Loss-of-Function Coding Variants in *TET2* are Associated with Multiple Neurodegenerative Diseases

J. Nicholas Cochran,[1] Ethan G. Geier,[2] Luke W. Bonham,[2] J. Scott Newberry,[1] Michelle D. Amaral,[1] Michelle L. Thompson,[1] Brittany N. Lasseigne,[1,3] Anna M. Karydas,[2] Erik D. Roberson,[4] Gregory M. Cooper,[1] Gil D. Rabinovici,[2,5] Bruce L. Miller,[2] Richard M. Myers,[1,6] Jennifer S. Yokoyama,[2,5,6,*] and Alzheimer's Disease Neuroimaging Initiative

We conducted genome sequencing to search for rare variation contributing to early-onset Alzheimer's disease (EOAD) and frontotemporal dementia (FTD). Discovery analysis was conducted on 435 cases and 671 controls of European ancestry. Burden testing for rare variation associated with disease was conducted using filters based on variant rarity (less than one in 10,000 or private), computational prediction of deleteriousness (CADD) (10 or 15 thresholds), and molecular function (protein loss-of-function [LoF] only, coding alteration only, or coding plus non-coding variants in experimentally predicted regulatory regions). Replication analysis was conducted on 16,434 independent cases and 15,587 independent controls. Rare variants in *TET2* were enriched in the discovery combined EOAD and FTD cohort ($p = 4.6 \times 10^{-8}$, genome-wide corrected $p = 0.0026$). Most of these variants were canonical LoF or non-coding in predicted regulatory regions. This enrichment replicated across several cohorts of Alzheimer's disease (AD) and FTD (replication only $p = 0.0029$). The combined analysis odds ratio was 2.3 (95% confidence interval [CI] 1.6–3.4) for AD and FTD. The odds ratio for qualifying non-coding variants considered independently from coding variants was 3.7 (95% CI 1.7–9.4). For LoF variants, the combined odds ratio (for AD, FTD, and amyotrophic lateral sclerosis, which shares clinicopathological overlap with FTD) was 3.1 (95% CI 1.9–5.2). TET2 catalyzes DNA demethylation. Given well-defined changes in DNA methylation that occur during aging, rare variation in *TET2* may confer risk for neurodegeneration by altering the homeostasis of key aging-related processes. Additionally, our study emphasizes the relevance of non-coding variation in genetic studies of complex disease.

Introduction

Neurodegeneration with a clinical onset prior to the age of 65 can be devastating for affected individuals, their families, and caregivers, imposing financial burden and hardship during a period of life when individuals are often most productive.[1] Early-onset neurodegenerative diseases such as early-onset Alzheimer's disease (EOAD) and frontotemporal dementia (FTD) are typically thought of as disease forms with highly penetrant genetic contributions, and indeed both can result from Mendelian pathogenic mutations (with Mendelian causes being more common in FTD).[2] However, these diseases exhibit a high degree of heritability that remains unexplained by currently known genetic contributors.[3,4] This suggests that additional genetic factors likely contribute to disease but have not yet been identified. Despite attempts at genome-wide association study (GWAS) of relatively sizeable cohorts, only modest association signals have been identified for FTD[5] and one form of EOAD, posterior cortical atrophy.[6] In contrast, by examining rare variation, sequencing studies have been successful in identifying more moderately to highly penetrant contributions to disease. Successes in Alzheimer's disease (AD) include *ABCA7*, *SORL1*, and *TREM2* (reviewed elsewhere[7]). Similar successes for the amyotrophic lateral sclerosis (ALS)-FTD spectrum include *TBK1*,[8] *MFSD8*,[9] *DPP6*, *UNC13A*, and *HLA-DQA2*.[10] Despite these successes, the rarity of these diseases, along with the high cost of sequencing studies, has resulted in limited cohort sample sizes. Furthermore, prior studies have focused on coding regions of the genome, leaving non-coding regions largely unexplored for their contribution to disease risk.

Here we leveraged a large cohort of 683 individuals with neurodegenerative disease, many of whom have an early age of disease onset (<65), and 856 healthy adult controls (with no known neurological abnormalities) who have undergone genome sequencing to probe both coding and non-coding rare and predicted deleterious variants across the genome for association with disease risk. We assessed variant associations between EOAD and FTD versus controls both separately and together (all cases versus controls), with the hypothesis that genetic pleiotropy—where variation in a single gene associates with multiple, different phenotypes—may play a role, as previously described for neurodegenerative diseases.[11–15]

[1]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, United States; [2]Memory and Aging Center, Department of Neurology, University of California, San Francisco, San Francisco, CA 94158, United States; [3]Department of Cell, Developmental and Integrative Biology, University of Alabama at Birmingham, Birmingham, AL 35294, United States; [4]Center for Neurodegeneration and Experimental Therapeutics, Alzheimer's Disease Center, Departments of Neurology and Neurobiology, University of Alabama at Birmingham, Birmingham, AL 35294, United States; [5]Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA 94158, United States
[6]These authors contributed equally to this work
*Correspondence: jennifer.yokoyama@ucsf.edu
https://doi.org/10.1016/j.ajhg.2020.03.010.

Material and Methods

Sample Selection

The majority of cases were selected from the University of California, San Francisco (UCSF) Memory and Aging Center with an intentional selection of early-onset cases when possible to maximize the likelihood of identifying genetic contributors, along with healthy older adult controls (a total of 664 cases and 102 controls, with 71 of these cases previously described[9]). All UCSF cases and controls were clinically assessed using previously described methods.[9] This cohort was intentionally depleted of cases with known Mendelian variants associated with neurodegenerative diseases, and any cases with known Mendelian variants identified after genome sequencing were excluded (see Results). A small number of samples (19 cases and 21 controls) were obtained from the University of Alabama at Birmingham (UAB) from an expert clinician who employed the same diagnostic procedures (case studies described elsewhere[16]). The resulting cohort was enriched for early-onset cases (Table 1). Additional neurologically healthy controls sequenced at HudsonAlpha were also included from two cohorts: a healthy aging control set from the National Institute of Mental Health (NIMH) (132 controls) and healthy unaffected parents from a childhood disease study where *de novo* mutations are the most common cause of disease,[17] making these parents reasonably representative population controls (601 controls). All participants or their surrogates provided written informed consent to participate in this study, and the institutional review boards at each site approved all aspects of the study.

Genome Sequencing

The majority of genome sequencing was performed at the HudsonAlpha Institute for Biotechnology on the Illumina HiSeq X platform (1,468 samples from UCSF, UAB, NIMH, and HudsonAlpha), while a small subset was sequenced at the New York Genome Center (NYGC), also on the HiSeq X platform (71 samples from UCSF, described previously[9]). Mean depth was 34× with an average of 92% of bases covered at 20×. Sequencing libraries at HudsonAlpha were prepared by Covaris shearing, end repair, adaptor ligation, and PCR using standard protocols. Library concentrations were normalized using KAPA qPCR prior to sequencing. All variants meeting either Mendelian diagnostic criteria or variants in top hits from the discovery cohort were validated through the use of Sanger sequencing.

Data Quality Control

All sequencing reads from both sequencing centers were aligned to the hg19 reference genome with bwa-0.7.12.[18] Binary alignment maps (BAMs) were sorted, and duplicates were marked using Sambamba 0.5.4.[19] Insertion-deletion polymorphisms (indels) were realigned, bases were recalibrated, and genomic variant cell formats (gVCFs) were generated through the use of GATK 3.3.[20] Variants were called across all samples in a single batch through the use of GATK 3.8 using the -newQual flag to minimize false negative singleton calls. The variant cell format (VCF) was quality filtered with a genotype level requirement for 95% of sites to have a minimum genotype quality (GQ) of 20 and DP of 10 (applied using VCFtools 0.1.15[21]), and a variant level filter of variant quality score log-odds (VQSLOD) > -3. The small number of missing genotypes remaining after that quality filtering step were assumed to be reference (filled with bcftools 1.6–19[22]) in order to avoid errors in downstream processing using the package GEMINI 0.20.2,[23] which adds missing genotypes to non-reference counts with its burden function. We note that assuming these alleles are reference is a conservative assumption because it is biased toward the null and will reduce false positive associations. Goleft indexcov 0.1.17[24] was used for sex checks, and samples failing sex checks were excluded. KING 2.2[25] was used to check for familial relationships, and related individuals (up to 4th degree relatives identified using identity by descent segment analysis, as well as a small number of duplicate samples between different cohorts) were excluded. Ancestry was elucidated by using both principal component analysis using plink 1.9[26] compared to 1000 Genomes data[27] (using common variation overlapping with 1000 Genomes calls) and analysis using ADMIXTURE 1.3.0[28] (Figure S1), and only samples from the largest cluster (European ancestry) were retained for discovery analysis in order to minimize potential confounding population effects.

Annotation, Filtering, and Burden Analysis

In order to facilitate annotation and burden analysis, multi-allelic sites were split using Vt.[29] All variants were annotated with CADD v1.3,[30] including all indels. SnpEff 4.3 s[31] was used to annotate with the gene definitions from human genome build Ensembl GRCh37.75. Population database frequency annotations included 1000 Genomes Phase 3, TOPMed Bravo[32] (lifted over from hg38 to hg19 using CrossMap 0.2.7[33]), and several population database sets annotated using WGSA 0.7,[34] including ExAC,[35] gnomAD,[36] ESP, and UK10K. Variants were also annotated using dbSNP release 151.[37] A final important annotation set was the union of regions called by GenoSkyline-Plus[38] as potential regulatory regions. GenoSkyline-Plus incorporates chromatin marks, DNA accessibility, RNA-seq, and DNA methylation to predict function. All tracks derived from direct human tissue sources were included (sources propagated in culture were excluded), and a total of 50 of 66 tissue and cell types described in Table S2 from the publication describing GenoSkyline-Plus[38] were used for annotation (see Table S1 for included epigenome tracks in the union).

Variants were filtered using SnpSift 4.3 s. In addition to the quality filters described, variants were further filtered based on local and population frequency, predicted deleteriousness (CADD v1.3), and segmentation for function. To enrich for rare variation, variants were pre-filtered for a maximum minor allele count of three (approximately 0.1% local allele frequency), and a maximum allele frequency of 1 in 10,000 in any population (i.e., a minor allele frequency [MAF] cutoff of 0.0001) included in the aforementioned population databases. In addition, non-coding variants were more strictly filtered to include only variants present in a GenoSkyline-Plus qualifying region as described above and required to be absent from dbSNP 151.

From the initial pre-filtered file, we conducted further filtering to arrive at nine total filter conditions. First, we evaluated variants meeting either (1) an MAF cutoff of 0.0001 for all populations and a CADD score greater than 10 or 15, or (2) private variation and CADD score greater than 10 or 15, for a total of four conditions that include non-coding variants. We also confined our study to coding variants with the same allele frequency combinations and CADD cutoffs listed for four total coding-only conditions. For canonical loss of function (LoF), we only considered the base MAF cutoff of 0.0001 for all populations and CADD 10 cutoff for a total of one canonical LoF condition (also note that all canonical LoF variants meeting these criteria were included in the other eight filter conditions regardless of allele frequency or

Table 1. Demographic Information for Discovery and Replication Cohorts

Group	n	Sex (% F)	Med. Age (Rng)[a]	>65[b]	0 ε4[c]	1 ε4[c]	2 ε4[c]	Batch Call	Seq. Type	Seq. Center
UCSF Discovery										
AD	227	52%	59 (45–84)	23%	44%	41%	14%	HA	HiSeq X (100%)	HA (100%)
FTD	208	48%	65 (33–89)	49%	74%	24%	2%	HA	HiSeq X (100%)	HA (100%)
Ctrl	671	56%	40 (21–86)	9%	74%	23%	3%	HA	HiSeq X (100%)	HA (100%)
UCSF Replication										
AD	66	55%	61 (48–71)	21%	55%	30%	15%	HA	HiSeq X (100%)	HA (80%), NYGC (20%)
FTD	136	46%	67 (29–88)	57%	78%	21%	1%	HA	HiSeq X (100%)	HA (80%), NYGC (20%)
Ctrl	157	54%	54 (22–85)	15%	77%	20%	3%	HA	HiSeq X (100%)	HA (80%), NYGC (20%)
ADSP Replication										
AD	1,723	60%	76 (50–90+)	90%	47%	42%	11%	ADSP	HiSeq X (83%), HiSeq 2k (17%)	Broad (31%), WashU (30%), Baylor (26%), Illumina (13%)
Ctrl	1,860	66%	73 (48–90+)	87%	74%	25%	1%	ADSP	HiSeq X (83%), HiSeq 2k (17%)	Broad (31%), WashU (30%), Baylor (26%), Illumina (13%)
AMP-AD Replication										
AD	741	68%	89 (60–90+)	98%	60%	35%	4%	AMP-AD	HiSeq X (100%)	Broad (54%), NYGC (46%)
FTD	183	52%	81 (61–90+)	93%	78%	21%	2%	AMP-AD	HiSeq X (100%)	Broad (54%), NYGC (46%)
Ctrl	440	59%	85 (57–90+)	98%	82%	17%	1%	AMP-AD	HiSeq X (100%)	Broad (54%), NYGC (46%)
All Replication										
AD	2,530	62%	79 (48–90+)	90%	51%	40%	9%	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center
FTD	319	49%	76 (29–90+)	78%	78%	21%	2%	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center
Ctrl	2457	64%	74 (22–90+)	84%	76%	23%	1%	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center
Combined Discovery + Replication										
AD	2757	61%	77 (45–90+)	85%	51%	40%	9%	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center
FTD	527	49%	71 (29–90+)	66%	76%	22%	2%	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center
Ctrl	3128	62%	71 (21–90+)	68%	75%	23%	2%	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center	merged w/ adjustment for batch call, seq. type, and seq. center

Demographic information for the discovery cohort and replication cohorts for genome sequencing findings. For the UCSF cohort, demographic information by individual sample is available in Table S5 along with phenotype sub-type and the first four principal component loadings used to impute ancestry and filter to the largest ancestral group (European ancestry) for the discovery cohort. AD—Alzheimer's disease. FTD—fronto-temporal dementia. Ctrl—control. UCSF—University of California, San Francisco. NYGC— New York Genome Center. HA—HudsonAlpha Institute for Biotechnology. ADSP—Alzheimer's Disease Sequencing Project. AMP-AD— Accelerating Medicines Partnership—Alzheimer's Disease.

[a]Med. Age (Rng) is median age at enrollment (range).

[b]>65 indicates age at enrollment was greater than 65.

[c]0, 1, or 2 ε4 indicates number of *APOE* ε4 alleles.

CADD cutoff given the known deleteriousness of canonical LoF variants). Table S2 summarizes all filtering conditions employed. We note that these are extensively overlapping filter conditions (see Figure S2 for correlations), and thus these filter conditions often yield similar results. For example, all conditions constrained to private variation will be a subset of matched conditions with an MAF cutoff of 0.0001 for all populations, and all coding-only conditions are a subset of the conditions that allow rare non-coding variation.

A VCF for each of the nine filtering conditions was loaded into a GEMINI 0.20.2 database,[23] which was used to aggregate counts of variants for each individual by gene. By default, GEMINI is constrained to coding variation, so GEMINI python scripts were edited to allow for counting of variants in non-coding regions as well. Variants upstream or downstream (within 5kb, the SnpEff default) were also assigned to their adjacent genes. The number of qualifying individuals was the final count unit, where one or more qualifying variants in a gene for a given individual resulted in that individual having a qualifying count for that gene (i.e., if an individual had two qualifying variants, they would still only be counted once in order to account for the possibility of a recessive model of inheritance or negligibility of the second variant if on the same allele). Individuals with more than three qualifying variants in a gene were not counted as qualifying because an excess of rare and predicted damaging variants in a single gene may be indicative of a sequencing or variant-calling error.

Burden Analysis Statistics

In order to assess the effect of covariates for the discovery set, as well as for any replication sets where the necessary covariate data were available, we tested via SKAT 1.3.2.1[39] using the adaptive efficient re-sampling method[40] adjusted for sex, number of *APOE* ε4 alleles, the first four principal components from common variant PCA, batch call, sequencer type, and sequencing center. Statistical significance was set at a corrected p value < 0.05 based on a correction factor of 57,354. We arrived at this correction factor by conducting a cross-correlation analysis of p values from all filter conditions. Three main correlation clusters of filter conditions corresponding to case-control test sets (EOAD versus control, FTD versus control, or all cases versus control) were apparent (Figure S2), so we applied a correction factor of three to all protein coding genes in hg19 put forth by the Human Genome Organization (HUGO) Gene Nomenclature Committee (19,118 genes, Table S3); this resulted in a correction factor of 57,354. In order to allow for use of replication cohorts where covariate data was not available, we also utilized a two-sided Fisher's exact test. SKAT and Fisher's tests were highly correlated (Pearson's $r = 0.80$ of log transformed p values).

GWAS

In addition to rare variant burden, we also assessed single common variant contributions by conducting GWAS using plink 2.0.[26] 95% of sites were required to have a minimum GQ of 20 and DP of 10, and only PASS filter variants not in a RepeatMasker region with greater than 1% MAF and a Hardy-Weinberg cutoff of $p > 10^{-6}$ were included. Logistic regression was employed using a genotypic model adjusted for sex and the first four principal components. GWAS results were plotted through the use of qqman.[41] Methods for the cohort used for GWAS replication, the International Genomics of Alzheimer's Project (IGAP), are provided in the Supplemental Information.

Clinical Rate of Progression Analysis

Our study utilized genetic and longitudinal clinical data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to study the clinical profiles and progression of *TET2* rare variant carriers. Data used in the preparation of this article were obtained from the ADNI database (see Web Resources). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. ADNI is a multi-center prospective longitudinal cohort study created to study the genetic, clinical, and imaging correlates of AD,[42–44] and ADNI cases are present in the Alzheimer's Disease Sequencing Project (ADSP) replication cohort. Every study participant undergoes a thorough assessment that includes clinical characteristics, cognitive testing, and genetic sequencing. Participants were diagnosed as either normal controls, MCI, or AD (note that some participants progressed from MCI to AD while being followed, with the last assessment used for case designation in the replication analysis, although they may be designated as beginning at the MCI stage in the following analysis). For clinical rate of progression analysis, we used the Clinical Dementia Scale Sum of Boxes (CDRSB) score [45] a broad measure of clinical progression and impairment well-validated in multiple studies.[46,47]

To test whether variation in *TET2* predicts longitudinal clinical progression, we used linear mixed-effects modeling using R version 3.5.2. We covaried for baseline age, sex, education, and CDRSB score as well as for *APOE* ε4 dose. The model was implemented as follows:$\Delta\ \mathrm{CDRSB} = \beta_0 + \beta_1 \mathrm{Age}_{\mathrm{baseline}} \times \Delta t + \beta_2 \mathrm{Sex}_{\mathrm{female}} \times \Delta t + \beta_3 \mathrm{Education}_{\mathrm{baseline}} \times \Delta t + \beta_4 \mathrm{CDRSB}_{\mathrm{baseline}} \times \Delta t + \beta_5 \mathrm{APOE}\ 4_{\mathrm{dose}} \times \Delta t + \beta_6 TET2_{\mathrm{carrier\ status}} \times \Delta t + (1|\mathrm{subject}) + \varepsilon$.

Data Availability

All data in both discovery and replication sets are available either through an application for access by qualified researchers, or through public availability (access details in Supplemental Acknowledgments). We have also included supplemental text files with summary statistics for all conditions assessed in the discovery cohort burden analysis (Table S8) and GWAS (Table S15).

Results

Of the 1,539 samples in the original set, after quality control, a total of 74 samples were removed from analysis for the following reasons: two failed sex checks; 28 were pruned for relatedness; 12 were pruned due to an identifiable Mendelian variant (all of which were Sanger validated) meeting American College of Medical Genetics pathogenic or likely pathogenic criteria, including one *C9orf72* expansion carrier from the UAB set (for list of Mendelian variants identified, see Table S4); one control was pruned for conversion to MCI after enrollment; and 31 cases were pruned because of phenotypic uncertainty or diagnosis of MCI or Parkinson's disease (PD) rather than EOAD or FTD on re-evaluation after enrollment. The remaining dataset consisted of 1,465 individuals: 637 cases (293 EOAD and 344 FTD) and 828 controls. Of

Table 2. Discovery and Replication for Private, CADD > 10 Coding and Non-coding Variants in *TET2* (Combined Analysis of All Cases, AD and FTD, versus Controls)

Cohort Type	Cohort	Cases w/	Cases w/o	Ctrls w/	Ctrls w/o	SKAT p	Corr. p	FET p	OR (95% CI)
Discovery	UCSF Eur. (EOAD & FTD)	18	417	1	670	4.6×10^{-8}	2.6×10^{-3}[a]	4.9×10^{-7}	28.9 (4.5–1200)
Replication	UCSF Rep. (EOAD & FTD)	8	194	1	156	0.529	NA[a]	0.083	6.4 (0.8–287)
Replication	ADSP Genomes (LOAD)	54	1,669	31	1,829	2.1×10^{-3}	NA[a]	4.2×10^{-3}	1.9 (1.2–3.1)
Replication	AMP-AD (LOAD & FTD)	15	909	7	433	0.716	NA[a]	1.000	1.0 (0.4–3.0)
Replication	all replication cohorts	77	2,772	39	2,418	2.9×10^{-3}	2.9×10^{-3}[a]	6.1×10^{-3}	1.7 (1.2–2.6)
Combined	discovery + all replication	95	3,189	40	3,088	2.2×10^{-7}	NA[b]	7.2×10^{-6}	2.3 (1.6–3.4)

Variants in *TET2* absent from population databases and with a computational prediction of deleteriousness (CADD) score >10 (including non-coding variants in GenoSkyline-Plus regions) in the combined analysis considering both early-onset Alzheimer's disease (EOAD) and fronto-temporal dementia (FTD) cases versus controls was the only qualifying gene and filter set in the discovery analysis to reach statistical significance. Although we applied a correction factor of 57,354 based on genome-wide (19,118 HGNC protein-coding genes) testing of three clusters of correlated filter conditions (Figure S2), *TET2* remains significant if we conservatively do not consider the correlated nature of the different filter sets and instead apply a strict Bonferroni correction ($p = 0.024$). The primary test was SKAT adjusted for number of *APOE* ε4 alleles, sex, principal components 1–4, batch call, sequencer type, and sequencing center. Fisher's exact test yielded similar raw p values and was highly correlated with SKAT (Pearson's $r = 0.80$ of log-transformed p values) and is presented here for consistency with Table 3, in which some cohorts did not have covariate data available for SKAT and therefore relied on Fisher's exact. The main analyses based on pre-determined criteria have a value in the Corr. p column. Replication cohorts are listed individually for reference as well as combined discovery plus replication statistics. Subsets of Alzheimer's disease (AD) only versus control and FTD only versus control are provided in Table S6 and Table S7, respectively. Ctrls—controls. CI—confidence interval. LOAD—late-onset Alzheimer's disease. UCSF—University of California, San Francisco. ADSP—Alzheimer's Disease Sequencing Project. AMP-AD— Accelerating Medicines Partnership—Alzheimer's Disease.
[a]Significance ($p < 0.05$ after correction).
[b]NA = not applicable (not a pre-determined analysis).

these cases and controls, 302 were of non-European ancestry (determined by principal component and admixture analysis, Figure S1). Individuals of non-European ancestry were excluded from the discovery set in order to reduce heterogeneity, and those individuals were retained for replication. In addition, 57 of the remaining qualifying samples were sequenced at the NYGC without matched controls, so these were also excluded from the discovery set in order to avoid batch effects, and these samples were retained for replication because other replication cohorts contained controls sequenced at NYGC. The resultant discovery set consisted of 1,106 individuals of European ancestry: 435 cases (227 EOAD and 208 FTD) and 671 controls. All available demographic information for individual samples (case category, primary clinical diagnosis, sex, age at enrollment, *APOE* ε4 status, self-reported race and/or ethnicity, principal components 1–4 and 5 ADMIXTURE coefficients) is available in Table S5. Note that for the discovery set, all samples were called in one batch, sequenced on Illumina HiSeq X sequencers, and sequenced at HudsonAlpha, precluding the need to adjust for these covariates, although we did adjust for all of these covariates in replication and combined analyses. We provide an overall summary of demographic information for discovery and replication cohorts in Table 1. The majority of cases were clinically diagnosed and did not have autopsy material available for neuropathological sub-grouping at the time of analysis. At the time of this writing, 52% of enrolled subjects from UCSF were living (specified individually in Table S5), and all subjects were enrolled while living. Primary clinical diagnoses included as AD were logopenic variant primary progressive aphasia (29), posterior cortical atrophy (26), frontal AD (17), language AD (17), vascular AD (8), AD + dementia with Lewy bodies (DLB) (5), and AD not otherwise specified (125). Primary clinical diagnoses included as FTD spectrum disorders were behavioral variant FTD (83), corticobasal syndrome (65), nonfluent variant primary progressive aphasia (43), FTD plus ALS (20), primary supranuclear palsy (17), semantic variant primary progressive aphasia (17), argyrophilic grain disease (5), and FTD not otherwise specified (15).

We compared EOAD versus control, FTD versus control, and a combined analysis of EOAD and FTD versus control across all variant filtering conditions (see Material and Methods). In the discovery analysis of combined burden across EOAD and FTD versus control, with variants absent from population databases and with a CADD score >10 (including non-coding variants in GenoSkyline-Plus regions), one gene-disease association passed the multiple-comparison significance threshold: *TET2* (SKAT uncorrected $p = 4.6 \times 10^{-8}$, corrected $p = 0.0026$; Table 2, model corrects for number of *APOE* ε4 alleles, sex, and principal components 1–4). Note that, although we applied a multiple-correction cutoff of 57,354 based on three main clusters of correlated filter conditions (Figure S2), the p value for *TET2* would also pass a strict Bonferroni correction for 516,186 implicit tests (19,118 genes, 27 filter conditions) if we conservatively did not consider the correlated nature of the different filter sets (Bonferroni $p = 0.024$). Statistical tests separately comparing EOAD versus control and FTD versus control did not pass the same degree of multiple testing correction, but results for those comparisons are provided Table S6 (EOAD) and Table S7 (FTD), and these results demonstrate that the nominal enrichment level in *TET2* is similar in both EOAD and FTD. No other gene reached even nominal significance ($p < 1 \times 10^{-5}$) in any filter condition, so *TET2* was the only gene considered for replication analysis. However, in the interest of making data from this study readily available, counts and p values for all genes assessed are provided in Table S8.

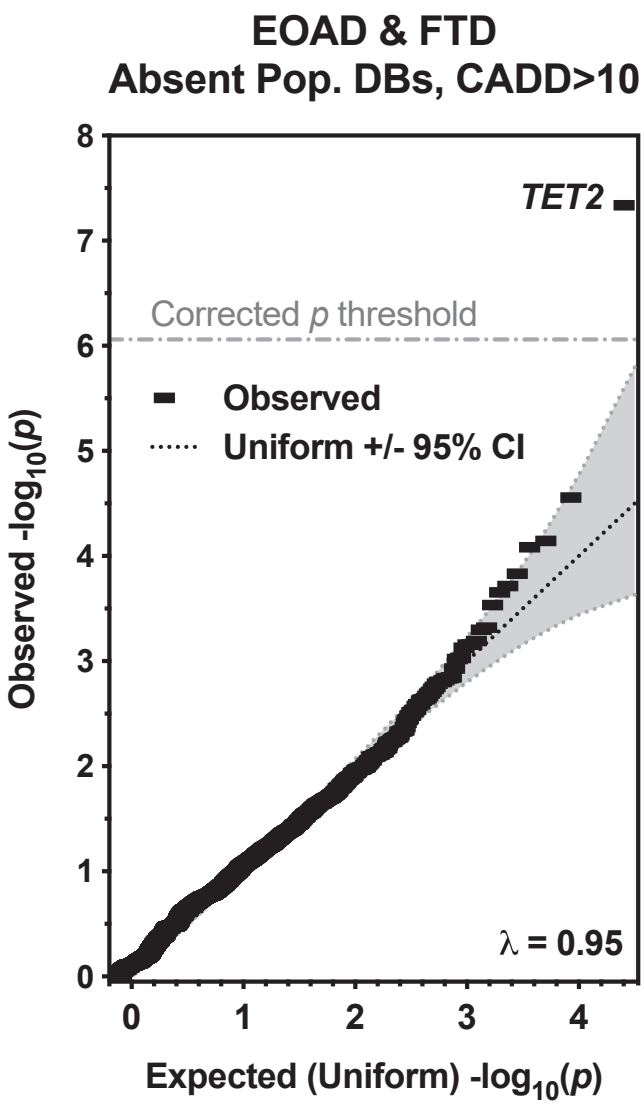

Figure 1. QQ plot of p Values from the Discovery Burden Analysis of Early-Onset Alzheimer's Disease (EOAD) and Fronto-temporal Dementia (FTD) Cases versus Controls and Private, CADD > 10 Variants

TET2 is the top and only hit reaching statistical significance (corrected $p < 0.05$). No genomic inflation was observed ($\lambda = 0.95$). The uniform distribution and theoretical 95% confidence interval based on a beta distribution is shown. Note that, in addition to passing the correction threshold, *TET2* also falls well outside of theoretical random p value distributions.

All qualifying variants in cases for *TET2* were both Sanger validated and visually evaluated using the Integrative Genomics Viewer (IGV). Two variants failed Sanger validation (due to adjacent erroneous indel calls in a single sample) and were excluded from the variant counts in Table 2, all statistics, and in Table S9 where all qualifying variants in *TET2* are listed. In addition, two cases had adjacent variant calls that were found to make up one variant. These were also corrected in all statistical analyses and tables. The single control with a qualifying *TET2* variant did not have material available for Sanger sequencing but appeared valid in IGV (a single-nucleotide variant with eight alternate allele reads among 18 total reads).

Next, we assessed potential confounding due to stratification by a QQ plot of the p value distribution for the filter set that produced the top result, and we observed no genomic inflation; this is consistent with a well-matched case-control dataset ($\lambda = 0.95$, Figure 1).

To help determine the types of sequencing datasets to target for replication, we assessed the variant type (coding or non-coding) and associated disease for all qualifying *TET2* variants in the discovery set. Qualifying variants were observed in 11 EOAD cases, eight FTD cases, and one control. Of the 11 EOAD cases, one had depressive symptoms, one had language symptoms and possible corticobasal syndrome, one had logopenic variant primary progressive aphasia, and one had a previous diagnosis of behavioral variant FTD revised to frontal AD (seven had no additional noted phenotypes). Of the eight FTD cases, three had corticobasal syndrome (one of whom had AD symptoms and possible posterior cortical atrophy), one had FTD plus ALS, and four had behavioral variant FTD (one with AD symptoms and one with seizures). Eight cases in total harbored coding variants, six of which were canonical LoF variants (four EOAD and two FTD). Because non-coding variants make up a large portion of the signal, we assessed coding and non-coding variants separately. We observed a similar level of enrichment for both coding and non-coding variants in EOAD and FTD cases when these types of variants were considered independently of one another (Figure 2A). Furthermore, the non-coding variants were prevalent in regions predicted to have regulatory function (Figure 2B). Combined with the high number of canonical LoF variants, these data support a model whereby *TET2* haploinsufficiency, resulting from either canonical LoF variation or expression-altering non-coding variation, may contribute risk to neurodegenerative disease.

To replicate this finding, we used six additional cohorts (five independent, one internal) with available sequencing data from individuals diagnosed with a neurodegenerative disorder and from healthy controls. Based on the variants discovered in *TET2*, we attempted to replicate the association between disease risk and aggregate rare variant burden in *TET2* by using two arms: the same conditions used in discovery applied to other genome sequencing datasets as a primary measure, and canonical LoF only analysis as a secondary measure to allow for incorporation of exome sequencing datasets. We assessed three cohorts with genome sequencing data for replication using the same conditions applied in the discovery set: ADSP (2,208 late-onset AD [LOAD] cases and 2,208 controls; 1,723 cases and 1,860 controls after relatedness filtering), the Accelerating Medicines Partnership—Alzheimer's Disease (AMP-AD) cohort (749 LOAD, 184 FTD, and 446 controls; 741 LOAD, 183 FTD, and 440 controls after relatedness filtering), and the individuals of non-European ancestry and NYGC–sequenced FTD cases from our cohort not

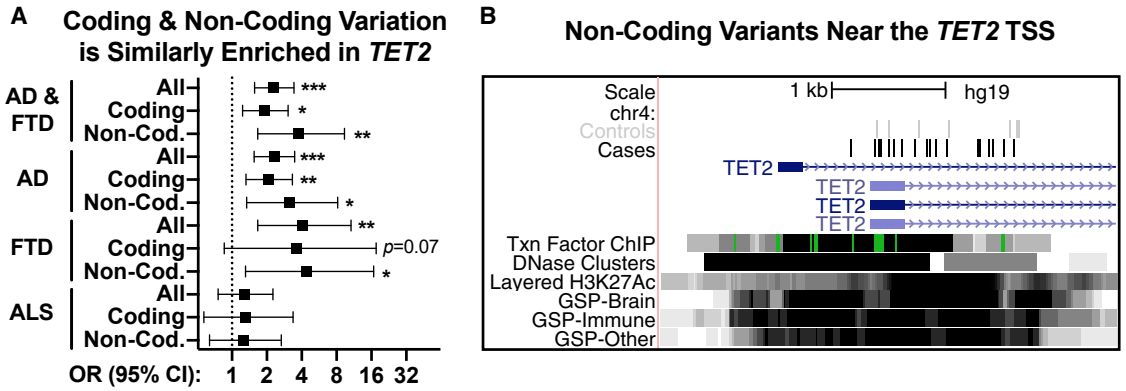

Figure 2. Qualifying Non-coding Variation in *TET2* Is Enriched at a Similar Level as Coding Variation and Occurs in Key Predicted Functional Regulatory Regions

(A) Odds ratios are shown for combined analyses (cohorts described in Table 2). Breaking out coding and non-coding variation reveals similar effect sizes and p values. * indicates $p < 0.01$, ** indicates $p < 0.001$, and *** indicates $p < 0.0001$ by Fisher's exact test.

(B) Non-coding variants near the *TET2* transcription start site (hg19 chr4:106,066,000–106,070,000) serve as an example of variant enrichment in key regions predicted to have regulatory function. GSP indicates GenoSkyline-Plus regions.

assessed in the discovery set (66 EOAD, 136 FTD, and 157 controls). Assessment of these three cohorts revealed replication of the signal for *TET2* overall for EOAD, LOAD, and FTD combined versus control ($p = 0.0029$; Table 2). Although the statistics for separate analyses of EOAD versus control and FTD versus control did not meet significance criteria, secondary analysis of those subgroups revealed similar levels of enrichment within each distinct condition (Table S6 [EOAD] and Table S7 [FTD]). Because of the established genetic overlap between FTD and ALS,[48] we also assessed variants in Project MinE[49] (4,366 ALS cases and 1,832 controls) and observed a non-significant trend toward a slight enrichment in ALS cases (OR 1.3, 95% confidence interval [CI] 1.1–2.7; Table S7). Although this is not a formal replication because no ALS cases were included in the discovery set, we present these findings in Table S7 along with FTD statistics.

Finally, we assessed predicted LoF variants alone in all aforementioned cohorts (UCSF European discovery set, UCSF replication set, ADSP, AMP-AD, and Project MinE) along with exome sequencing data from a second ALS dataset[8] and additional exome samples from ADSP[50] for a total of seven sample sets. We observed a robust signal for association between predicted canonical LoF variants and disease across multiple disease cohorts (Table 3). Specifically, three of the four largest independent replication cohorts (ADSP genomes [LOAD], ADSP exomes [LOAD], and HudsonAlpha-Duke-Stanford ALS exomes) all exhibit independent nominal replication ($p < 0.05$). Meta-analysis of all canonical LoF variants from all available cohorts (across EOAD, LOAD, FTD, and ALS) yielded a p value below commonly used exome-wide significance cutoffs ($p = 9.8 \times 10^{-7}$; Table 3), and subgroup analyses of both AD and FTD–ALS versus controls were each nominally significant ($p < 0.05$) and suggested similar degrees of enrichment.

To assess potential clinical implications of rare variation in *TET2*, we queried the ADNI dataset,[42–44] which includes clinical rate of progression data. We used linear mixed-effects modeling to test whether qualifying rare variation (based on the discovery condition that passed multiple corrections testing) in *TET2* predicts longitudinal clinical progression. We covaried for baseline age, sex, education, and CDRSB score as well as for *APOE* ε4 dose. A total of 786 ADNI participants had *TET2* genotypes available for analysis. There was no significant difference in the distribution of *TET2* rare variant carriers in regards to sex, *APOE* ε4 dose, education, or baseline CDRSB score (Table S10). There was a significant difference between *TET2* rare variant carriers and non-carriers in regards to baseline age (Table S10), but one must keep in mind that baseline age is adjusted for along with sex, education, *APOE* ε4 dose, and baseline CDRSB score. Using linear mixed effects regression, we found a significant relationship between carrying any *TET2* rare variant and clinical progression as measured by change in CDRSB score in ADNI participants with MCI ($\beta \pm SE = 0.64 \pm 0.12$; $^*p = 6.2 \times 10^{-8}$, Bonferroni corrected p value for total number of longitudinal assessments: $^*p = 3.1 \times 10^{-6}$) (Figure 3; Table 4). A similar finding was observed when our analyses were limited to *TET2* LoF variant carriers ($\beta \pm SE = 0.58 \pm 0.13$; $^*p = 7.1 \times 10^{-6}$, Bonferroni corrected p value for total number of longitudinal assessments: $^*p = 3.6 \times 10^{-4}$) (Figure S3; Table 4). (Although we adjusted for covariates for rigor, no covariates were significantly associated with *TET2* LoF carrier status; Table S11.) We also explored whether rare variation in *TET2* predicted changes in CDRSB score and cognition (measured by Mini Mental State Exam [MMSE] score changes[51]) in MCI and control when analyzed separately. When we constrained the analysis to MCI, *TET2* rare variant carriers ($n = 8$) demonstrated a greater CDRSB score change over time compared to noncarriers, and this change was of a higher magnitude and significance compared to that found in the pooled analysis of control, MCI, and AD ($\beta \pm SE = 0.64 \pm 0.12$; $^*p = 6.17 \times 10^{-8}$;

Table 3. Canonical Loss-of-Function Variation in *TET2* is Nominally Enriched in Both AD and FTD-ALS

Cohort Type	Cohort	Cases w/	Cases w/o	Ctrls w/	Ctrls w/o	SKAT p	FET p	OR (95% CI)	Cases Frq.	Ctrls Frq.
Discovery	UCSF Eur. (EOAD & FTD)	6	429	0	671	5.4×10^{-3}	3.6×10^{-3}	∞ (1.8–∞)	1.38%	0.00%
Replication	UCSF Rep. (EOAD & FTD)	2	200	1	156	0.194	1.000	1.6 (0.1–92.6)	0.99%	0.64%
Replication	ADSP genomes (LOAD)	25	1,698	11	1,849	0.021	0.011	2.5 (1.2–5.6)	1.45%	0.59%
Replication	AMP-AD (LOAD & FTD)	0	924	1	439	0.186	0.323	0.0 (0.0–18.6)	0.00%	0.23%
Replication	Project MinE (ALS)	21	4,345	5	1,827	NA[b]	0.289	1.8 (0.6–6)	0.48%	0.27%
Replication	HA-Duke-Stanford (ALS)	11	2,863	5	6,400	NA[b]	2.0×10^{-3}	4.9 (1.6–18.1)	0.38%	0.08%
Replication	All rep. (AD, FTD, ALS)	59	10,030	23	10,671	NA[b]	2.0×10^{-5}	2.7 (1.7–4.6)	0.58%	0.22%
Combined	discovery + replication	65	10,459	23	11,342	NA[b]	9.8×10^{-7}[a]	3.1 (1.9–5.2)	0.62%	0.20%
Combined subset	AD except ADSP exomes	29	2,728	13	3,115	3.1×10^{-3}	4.7×10^{-3}	2.5 (1.3–5.3)	1.05%	0.42%
Sum. stat. set	ADSP exomes (sum. stats)	6,345 total cases		4,893 total controls		0.019 (ADSP model p value)			CMAF 0.49%	
Combined subset	all FTD	4	523	2	1,266	0.440	0.065	4.8 (0.7–53.6)	0.76%	0.16%
Combined subset	all ALS	32	7,208	10	8,227	NA[b]	1.4×10^{-4}	3.7 (1.8–8.3)	0.44%	0.12%
Combined subset	all FTD & ALS	36	7,731	12	9,493	NA[b]	3.0×10^{-5}	3.7 (1.9–7.8)	0.46%	0.13%
Population databases	gnomAD+TOPMed	–	–	284	196,035	–	–	–	–	0.14%

Because of the high number of canonical loss-of-function (LoF) variants in *TET2* observed in the discovery analysis, we performed a separate assessment of LoF variants alone. Although, because of the low number of qualifying counts, the LoF model did not pass multiple testing correction in the discovery analysis, *TET2* was the highest ranked LoF gene (lowest p value for enrichment in cases). Note the additional inclusion of amyotrophic lateral sclerosis (ALS) exomes (HudsonAlpha-Duke-Stanford). SKAT values could not be calculated for ALS sets (and thus not for summed replication and discovery+replication sets) because necessary covariate data were not available for these cohorts, although both ALS cohorts were independently filtered to include only individuals of European ancestry. Below this combined analysis, we also present summaries by each disease which achieved nominal significance ($p < 0.05$) for both combined analysis of all Alzheimer's disease (AD) cases and of all fronto-temporal dementia (FTD) and ALS cases. Note the addition of summary statistics from Alzheimer's Disease Sequencing Project (ADSP) exomes in this section as well. For ADSP exomes, the p value from the VEP HIGH meta-analysis model is shown (publicly available from the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site). For comparison, we have also listed the frequency of *TET2* LoF carriers in population databases (gnomAD minus TOPMed set added to counts from TOPMed), which is similar to the frequencies observed in control groups we analyzed. All frequencies are the percentage of individuals harboring a LoF variant (not minor allele frequency) except ADSP exomes, for which cumulative minor allele frequency (CMAF) for both cases and controls (ctrls) is listed. EOAD—early-onset Alzheimer's disease. OR—odds ratio. CI—confidence interval. UCSF—University of California, San Francisco. Fisher's Exact Test—Fisher's Exact Test.

[a]The combined analysis across all cases and controls was below an arbitrary exome-wide cutoff of 2.5×10^{-6} (a commonly used threshold based on correction of $p < 0.05$ for ~20,000 genes).

[b]NA = not applicable (covariate information was not available for some or all data in the set, precluding SKAT).

Table 4) when correcting for the covariates outlined above. Of note, *TET2* rare variant carriers diagnosed with MCI also demonstrated greater decreases in changes to MMSE results when compared to non-carriers ($\beta \pm SE = -0.47 \pm 0.17$; *p = 0.01; Table 4). Within controls (n = 6), there were no significant associations between *TET2* variant carrier status and either CDRSB or MMSE results.

We chose to focus initial longitudinal analysis on ADNI because of excellent longitudinal data availability in individuals with MCI, where longitudinal changes are most likely to be detected. However, to determine the robustness of the findings from the ADNI dataset, we queried data from samples with longitudinal data available from the UCSF dataset described here (demographics in Table S12) as well as from the Rush Religious Orders Study and Memory and Aging Project (ROSMAP) cohort (demographics in Table S13). The main finding of a difference in CDRSB score did not replicate in the UCSF set, and CDRSB score was not available in the Rush dataset (Table 4). MMSE results were not significantly different over time for the combined analysis across cases and controls for *TET2* rare variant carriers for any of the cohorts queried (Table 4), although constraining to only ADNI individuals with MCI did reveal a signal as mentioned previously. Other measures specific to particular variables did exhibit nominal significance. Three measures in the Rush dataset exhibited negative longitudinal betas (consistent with worse cognitive function) adjusted for sex, education, *APOE* ε4 dose, and baseline score (Table 4). Four measures in the Rush dataset actually exhibited positive longitudinal betas (consistent with *better* cognitive function); however, each of these measures was either trending or significantly lower at baseline, and *TET2* carriers in the Rush dataset also had a significantly higher education level (Table S13). In addition to the 28 measures shown in Table 4, in Table S14, we present 22 additional outcome measures from the Rush dataset where neither baseline nor longitudinal changes of nominal significance were detected. Only a

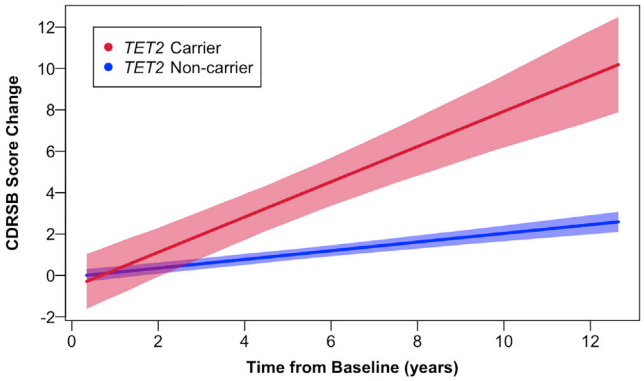

Figure 3. Longitudinal CDRSB Changes in ADNI Participants with MCI and Qualifying *TET2* Rare Variants

TET2 rare variant carriers with mild cognitive impairment (MCI) from the Disease Neuroimaging Initiative (ADNI) show greater Clinical Dementia Scale Sum of Boxes Score (CDRSB) changes over time compared to non-carriers after controlling for age, sex, and education, *APOE* ε4, and baseline CDRSB score ($\beta \pm SE = 0.64 \pm 0.12$; $^{*}p = 6.2 \times 10^{-8}$, Bonferroni corrected p value for total number of longitudinal assessments: $^{*}p = 3.1 \times 10^{-6}$). The lines depicted illustrate CDRSB change with 95% confidence intervals in shading.

handful of measures were significant after Bonferroni correction for the 50 variables tested, including faster progression in CDRSB score in ADNI individuals with MCI under both all-variant and LoF models.

In addition to the rare variant analysis, we also used GWAS to assess the contribution of single common variants for the discovery cohort for EOAD versus control (Figure S4A); including assessments with *APOE* ε4 adjustment (Figure S4B), for FTD versus control (Figure S4C) including with *APOE* ε4 adjustment (Figure S4D), and for combined analysis of EOAD and FTD versus control (Figure S4E) including with *APOE* ε4 adjustment (Figure S4F). Besides the expected effect of *APOE* ε4 in the EOAD versus control comparison (Figure S4A), only two singleton variants (i.e., lacking a pattern of nearby nominally significant variants in LD) on chromosomes 3 (Figure S4G) and 12 (Figure S4H) were identified as passing genome-wide significance ($p < 5 \times 10^{-8}$) in the combined EOAD and FTD versus control comparison. Both lie in regions with high adenine/thymine content, and neither replicate in a large AD GWAS meta-analysis[52] (chr3 rs9869684 replication $p = 0.49$, chr12 rs709216 replication $p = 0.58$); this suggests spurious associations due to sequencing artifact. GWAS summary statistics are provided in Table S15.

Discussion

In this study, we identified a significant excess of rare, likely deleterious variation in *TET2* as a risk factor for multiple neurodegenerative disorders, including EOAD, LOAD, FTD, and ALS. This finding is important for two main reasons. First, *TET2* plays an important role in the conversion of methylation to 5-hydroxymethylation, implicating dysfunction in a pathway known to be critical during aging[53] and learning and memory[54] in age-associated neurodegenerative diseases. Second, it is striking that the effect sizes in both coding and non-coding variant enrichments were comparable. This point suggests that further investigation of non-coding variation in complex disease genome sequencing studies holds potential for the identification of new contributors to disease.

TET2 promotes de-methylation of DNA by catalyzing conversion of methylation to 5-hydroxymethylation, and TET2 is highly expressed in brain (reviewed elsewhere[55]). Defined methylation changes occur with age in humans ("Horvath's clock"[53]) and there is some evidence for an association between increased "methylation age" and disease (systematically reviewed elsewhere[56]). Taken together, this raises speculation that reduced function or loss of TET2, a critical regulator in methylation processes, may have adverse age-associated effects. Evidence from mouse models further supports this idea: using either exercise-induced upregulation of *Tet2*[57] or artificial overexpression of *Tet2*[58] to promote the conversion of methylation to 5-hydroxymethylation improves memory in mice by increasing neurogenesis in the dentate gyrus. Conversely, reducing *Tet2* in mouse hippocampus leads to reduced neurogenesis and impaired memory,[58] consistent with its role in promoting adult neurogenesis in mice.[59] Finally, reduction of *Tet2* in mouse primary neurons also reduced cell survival.[60] All of these observations are consistent with detrimental consequences of loss of *Tet2* function and suggest that neurons may be particularly vulnerable to these effects. Further support for a generally important role of TET enzymes comes from a recent study that implicates mono- and bi-allelic LoF of *TET3* in childhood diseases.[61] (Based on population database estimates, *TET3* is more constrained against LoF,[35] and this [along with bi-allelic contributions] could explain the earlier ages observed.) In addition to general evidence for the importance of TET2 and other TET enzymes, an intriguing and more specific role for *Tet2* has also been proposed, one that implicates *Tet2* in microglial response, particularly around amyloid plaques,[62] and thus suggests that *Tet2* LoF may prevent its recruitment into a protective role (similar to recent findings on *TREM2* which suggest that higher secreted TREM2 levels are protective,[63] and thus support a model where risk-conferring *TREM2* variants result in LoF). Recent additional support for a role of *Tet2* LoF in facilitating the neurodegenerative disease process comes from an AD mouse model: 9-month-old 2xTg-AD (APPswe/PSEN1) mice showed reduced expression of *Tet2*, and *in vivo* reduction of *Tet2* in five5-month-old 2xTg-AD mice accelerated cognitive deficits, increased amyloid plaque load, increased inflammatory markers, and reduced synaptic markers.[64] Finally, the data we analyzed from ADNI is consistent with deleterious consequences of *TET2* rare variants, and our observations support a faster rate of both general clinical decline (measured by CDRSB, from both the combined and MCI-only analysis) and

Table 4. Longitudinal Changes in Dementia and Cognitive Measures in *TET2* Rare Variant Carriers

[a]Set	[b]Model	Cohort	Cognitive Test	*TET2* β ± SE	*TET2* p	*TET2*[c]year β ± SE	*TET2*[c]year p
MCI	All	ADNI	Clin. Dem. Rating Sum of Boxes	−0.52 ± 0.72	0.47	0.64 ± 0.12	6.2×10^{-8}[c,d]
MCI	All	ADNI	Mini Mental State Exam	0.51 ± 1.04	0.63	−0.43 ± 0.17	0.01[c]
MCI	All	ADNI	AD Assessment Scale 11	−0.25 ± 1.90	0.90	0.27 ± 0.33	0.42
MCI	All	ADNI	AD Assessment Scale 13	−0.41 ± 2.30	0.86	0.44 ± 0.38	0.25
MCI	LoF	ADNI	Clin. Dem. Rating Sum of Boxes	−0.26 ± 0.83	0.75	0.58 ± 0.13	7.1×10^{-6}[c,d]
MCI	LoF	ADNI	Mini Mental State Exam	−0.23 ± 1.20	0.85	0.00 ± 0.18	0.99
MCI	LoF	ADNI	AD Assessment Scale 11	0.26 ± 2.19	0.91	−0.30 ± 0.36	0.40
MCI	LoF	ADNI	AD Assessment Scale 13	−0.04 ± 2.65	0.99	−0.20 ± 0.42	0.63
All	All	ADNI	Clin. Dem. Rating Sum of Boxes	0.44 ± 0.46	0.34	0.14 ± 0.06	0.03[c]
All	All	UCSF	Clin. Dem. Rating Sum of Boxes	0.03 ± 0.72	0.97	0.07 ± 0.15	0.65
All	All	ADNI	Mini Mental State Exam	−0.19 ± 0.70	0.79	−0.03 ± 0.09	0.74
All	All	UCSF	Mini Mental State Exam	−0.71 ± 1.24	0.57	−0.20 ± 0.28	0.46
All	All	Rush	Mini Mental State Exam	0.49 ± 1.25	0.70	−0.15 ± 0.10	0.13
All	All	ADNI	AD Assessment Scale 11	0.45 ± 1.27	0.72	−0.23 ± 0.18	0.19
All	All	ADNI	AD Assessment Scale 13	0.30 ± 1.55	0.85	−0.30 ± 0.21	0.16
All	All	Rush	East Boston Immediate Recall	−0.02 ± 0.58	0.97	−0.12 ± 0.05	0.024[c]
All	All	Rush	10-Item Reading Test	−0.42 ± 0.36	0.24	−0.06 ± 0.03	0.037[c]
All	All	Rush	Word List Recall	0.61 ± 0.53	0.25	−0.09 ± 0.04	0.04997[c]
All	All	Rush	Stroop Color Naming	−9.77 ± 4.42	0.03[c]	1.40 ± 0.56	0.013[c]
All	All	Rush	Category Fluency Combined	−3.16 ± 2.12	0.14	0.75 ± 0.17	7.0×10^{-6}[c,d]
All	All	Rush	Category Fluency—Animals	−1.82 ± 1.18	0.12	0.54 ± 0.10	4.3×10^{-8}[c,d]
All	All	Rush	Category Fluency—Fruits	−1.08 ± 1.18	0.36	0.19 ± 0.10	0.048[c]
All	LoF	ADNI	Clin. Dem. Rating Sum of Boxes	0.30 ± 0.62	0.63	0.17 ± 0.09	0.04[c]
All	LoF	UCSF	Clin. Dem. Rating Sum of Boxes	−0.97 ± 0.93	0.30	0.15 ± 0.16	0.36
All	LoF	ADNI	Mini Mental State Exam	−0.63 ± 0.95	0.51	0.15 ± 0.12	0.22
All	LoF	UCSF	Mini Mental State Exam	−1.00 ± 1.69	0.55	−0.20 ± 0.28	0.49
All	LoF	ADNI	AD Assessment Scale 11	0.76 ± 1.72	0.66	−0.62 ± 0.24	0.01[c]
All	LoF	ADNI	AD Assessment Scale 13	0.25 ± 2.09	0.90	−0.63 ± 0.29	0.03[c]

Dementia and cognitive measures both at baseline and with time interaction (adjusted for baseline score) after adjusting for age, sex, education, *APOE* ε4. Note that because the Rush cohort contained only one loss-of-function (LoF) sample, it was not possible to analyze LoF separately for Rush. Note that Alzheimer's Disease Neuroimaging Initiative (ADNI) Clinical Dementia Rating Sum of Boxes data correspond to Figure 3 (All variants) and Figure S3 (LoF variants). MCI—mild cognitive impairment. AD—Alzheimer's disease. UCSF—University of California, San Francisco.

[a]Set—Set of samples analyzed. MCI indicates MCI only. All indicates analysis across dementia, MCI (when present), and control samples.

[b]Model—Type of *TET2* variant, either all meeting the primary rare variant filter, or LoF for loss-of-function only.

[c]Indicates $p < 0.05$. Note that p values are nominal and are not corrected for multiple comparisons.

[d]Indicates $p < 0.05$ after a Bonferroni correction for the 50 presented tests between this table and Table S14. Additional measures from the Rush cohort not reaching a nominal significance cutoff of $p < 0.05$ for either baseline or with time interaction are presented in Table S14.

cognitive decline (measured by MMSE, from the MCI-only analysis). While the findings from ADNI, a cohort with excellent longitudinal data availability in individuals with MCI (where longitudinal changes are most likely to be detected), were supportive of an association between *TET2* rare variants and faster decline, data from other cohorts not designed to detect longitudinal effects either were not significant or exhibited mixed results, emphasizing that this particular aspect of the study requires follow-up in studies of larger, longitudinal cohorts.

The strongest association signal in the discovery cohort was a combined analysis across all EOAD and FTD cases together. There are a number of caveats to a combined burden analysis across phenotypes. First, pleiotropy between diseases may be an invalid assumption if there is an outsized effect from one of the diseases considered. Second, if a strong effect is present with a specific phenotype, this may be diluted by including cases with a different phenotype. We addressed each of these issues through extensive replication and considering each disease cohort

separately, respectively. Despite these caveats, we argue that the benefits outweigh the drawbacks for two critical reasons beyond the increase in sample size: (1) known effects of genetic pleiotropy, and (2) the possibility of identifying shared pathways between diseases.

The first reason supporting comparison across EOAD and FTD is that genetic pleiotropy—in which a single locus contributes variance to multiple different phenotypes—may play a role in neurodegenerative disease risk. Our group and others have provided support for this idea through several studies investigating multiple neurodegenerative diseases through the use of GWAS approaches.[11–15] In addition to common risk variants, there is clear evidence of moderately to highly penetrant rare variation in single genes conferring risk or causality for multiple neurodegenerative diseases. First, rare variants in *TREM2* confer risk for both AD[65] and FTD.[66] Second, rare variation in multiple established genes such as *TBK1* and *C9orf72* confer risk or are causative across the ALS-FTD spectrum.[67] Third, moderately penetrant common risk alleles like *APOE* ε4 are primarily associated with AD, but also associated with risk for DLB,[68] FTD,[12] and age of onset in *C9orf72* carriers.[69] Fourth, *GBA* and *SNCA* were first identified as risk factors for PD, but also confer risk for DLB.[68] Finally, rare pathogenic variants in *MAPT* typically cause FTD,[70,71] but the c.1216C>T (p.Arg406Trp) (Genbank: NM_005910.5) pathogenic variant has also been associated with an EOAD presentation.[72,73] Furthermore, common variants near *MAPT* (tagging the H1 haplotype, which is associated with higher tau expression[74]) are associated with AD, PD, FTD, and ALS.[11,12,15]

A second important reason to analyze across different populations of individuals with neurodegenerative disorders is that performing analyses across cohorts of individuals diagnosed with different neurodegenerative disorders but with partially overlapping underlying neuropathology (i.e., tau-containing protein aggregates in both AD and approximately half of FTD cases; and TDP-43-containing protein aggregates in ALS, approximately half of FTD cases, and some AD cases) may identify shared dysregulated pathways, and has the potential to identify therapeutic targets with relevance across multiple neurodegenerative diseases. Indeed, our discovery that rare variation in *TET2* is associated with multiple neurodegenerative diseases suggests that age-related changes in methylation may be relevant across a broad spectrum of neurodegeneration.

In conclusion, we provide evidence that loss of *TET2* function confers risk for EOAD, LOAD, FTD, and ALS. Specifically, we found that, in aggregate, both coding and noncoding qualifying rare variation in *TET2* is associated with approximately a 2-fold risk increase across diverse populations of individuals with AD, FTD, and ALS, and that canonical LoF variation in *TET2* is associated with approximately a 3-fold risk increase for these diseases. We note that, similarly to any burden test, it is impossible from aggregate enrichment values to deconvolute variable penetrance levels among disease-relevant alleles and the degree of enrichment for truly associated variation. Future work to assess the functional effects of particular alleles and their concomitant levels of risk to individual variant carriers would be helpful in this regard. Additionally, further work is required to understand the local and global mechanisms by which alterations to TET2 levels and/or function contribute to disease risk, whether this risk is anchored to TET2's effects on aging biology, and, if so, whether rare variation in *TET2* also confers risk for other age-associated neurodegenerative diseases.

Accession Numbers

The accession number for UAB genome data is NIAGADS: NG00082. The accession numbers for control genome data from HudsonAlpha are dbGaP: phs001625.v1.p1 and phs001089.v3.p1. The accession number for ADSP genome data is NIAGADS: NG00067.

The accession number for ADSP exome data is dbGaP: phs000572.v7.p4. The accession number for Kunkle et al., 2019[52] GWAS summary statistics is NIAGADS: NG00075.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.03.010.

Acknowledgments

Funding for genomes from UCSF-enrolled participants sequenced at HudsonAlpha was generously provided by donors to the HudsonAlpha Foundation Memory and Mobility Fund (R.M.M). Funding for genomes sequenced at the New York Genome Center was provided from grant support from the Rainwater Charitable Foundation (J.S.Y.). Funding for sequencing of genomes from the University of Alabama at Birmingham was provided by the Daniel Foundation of Alabama (R.M.M.). Additional support was provided by National Institutes of Health (NIH) National Institute on Aging (NIA) grants NIH-NIA K01 AG049152 (J.S.Y.), NIH-NIA R01 AG062588 (J.S.Y.), Larry L. Hillblom Foundation grant 2016-A-005-SUP (J.S.Y.), the Rainwater Charitable Foundation (J.S.Y.), NIA P01 AG1972403 (B.L.M.), NIA P50 AG023501 (B.L.M.), NIA P30 AG062422 (B.L.M.), and R01 AG045611 (G.D.R.).

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (Web Resources). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A link to the complete listing of ADNI investigators is provided in the Web Resources. We thank the Genomic Services Lab at HudsonAlpha for DNA isolations, library generation, quality control, and sequencing. We also acknowledge a key contribution from the availability of datasets through controlled access, without which replication for this study would not have been possible. In particular, we thank ADSP, ADNI, the AMP-AD Knowledge Portal (Web Resources; with contributions from ROSMAP,[75] Mayo,[76] and MSBB), the Project MinE Sequencing Consortium,[49] and HA-Duke-Stanford ALS exomes.[8] More detail on contributing studies is provided in the Supplemental Acknowledgments.

Declaration of Interests

The authors declare competing interests. G.D.R. industry relationships: Research support from Avid Radiopharmaceuticals, Eli Lilly, GE Healthcare, and Life Molecular Imaging. Consultant/scientific advisory board member for Axon Neurosciences, Eisai, and Merck. Speaking honorarium from GE Healthcare. Associate Editor for JAMA Neurology.

Received: September 5, 2019
Accepted: March 20, 2020
Published: April 23, 2020

Web Resources

Alzheimer's Disease Neuroimaging Initiative (ADNI) database, http://adni.loni.usc.edu
AMP-AD Knowledge Portal, https://doi.org/10.7303/syn2580853
Complete listing of ADNI investigators, http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf
dbGAP, https://www.ncbi.nlm.nih.gov/gap/
National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS), https://www.niagads.org/

References

1. Lambert, M.A., Bickel, H., Prince, M., Fratiglioni, L., Von Strauss, E., Frydecka, D., Kiejna, A., Georges, J., and Reynish, E.L. (2014). Estimating the burden of early onset dementia; systematic review of disease prevalence. Eur. J. Neurol. *21*, 563–569.
2. Loy, C.T., Schofield, P.R., Turner, A.M., and Kwok, J.B. (2014). Genetics of dementia. Lancet *383*, 828–840.
3. Goldman, J.S., Farmer, J.M., Wood, E.M., Johnson, J.K., Boxer, A., Neuhaus, J., Lomen-Hoerth, C., Wilhelmsen, K.C., Lee, V.M., Grossman, M., and Miller, B.L. (2005). Comparison of family histories in FTLD subtypes and related tauopathies. Neurology *65*, 1817–1819.
4. Wingo, T.S., Lah, J.J., Levey, A.I., and Cutler, D.J. (2012). Autosomal recessive causes likely in early-onset Alzheimer's disease. Arch. Neurol. *69*, 59–64.
5. Ferrari, R., Hernandez, D.G., Nalls, M.A., Rohrer, J.D., Ramasamy, A., Kwok, J.B., Dobson-Stone, C., Brooks, W.S., Schofield, P.R., Halliday, G.M., et al. (2014). Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neurol. *13*, 686–699.
6. Schott, J.M., Crutch, S.J., Carrasquillo, M.M., Uphill, J., Shakespeare, T.J., Ryan, N.S., Yong, K.X., Lehmann, M., Ertekin-Taner, N., Graff-Radford, N.R., et al. (2016). Genetic risk factors for the posterior cortical atrophy variant of Alzheimer's disease. Alzheimers Dement. *12*, 862–871.
7. Nicolas, G., Charbonnier, C., and Campion, D. (2016). From Common to Rare Variants: The Genetic Component of Alzheimer's disease. Hum. Hered. *81*, 129–141.
8. Cirulli, E.T., Lasseigne, B.N., Petrovski, S., Sapp, P.C., Dion, P.A., Leblond, C.S., Couthouis, J., Lu, Y.F., Wang, Q., Krueger, B.J., et al.; FALS Sequencing Consortium (2015). Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science *347*, 1436–1441.
9. Geier, E.G., Bourdenx, M., Storm, N.J., Cochran, J.N., Sirkis, D.W., Hwang, J.H., Bonham, L.W., Ramos, E.M., Diaz, A., Van Berlo, V., et al. (2019). Rare variants in the neuronal ceroid lipofuscinosis gene MFSD8 are candidate risk factors for frontotemporal dementia. Acta Neuropathol. *137*, 71–88.
10. Pottier, C., Ren, Y., Perkerson, R.B., 3rd, Baker, M., Jenkins, G.D., van Blitterswijk, M., DeJesus-Hernandez, M., van Rooij, J.G.J., Murray, M.E., Christopher, E., et al. (2019). Genome-wide analyses as part of the international FTLD-TDP whole-genome sequencing consortium reveals novel disease risk factors and increases support for immune dysfunction in FTLD. Acta Neuropathol. *137*, 879–899.
11. Karch, C.M., Wen, N., Fan, C.C., Yokoyama, J.S., Kouri, N., Ross, O.A., Höglinger, G., Müller, U., Ferrari, R., Hardy, J., et al.; International Frontotemporal Dementia (FTD)–Genomics Consortium, International Collaboration for Frontotemporal Dementia, Progressive Supranuclear Palsy (PSP) Genetics Consortium, and International Parkinson's Disease Genomics Consortium (2018). Selective Genetic Overlap Between Amyotrophic Lateral Sclerosis and Diseases of the Frontotemporal Dementia Spectrum. JAMA Neurol. *75*, 860–875.
12. Ferrari, R., Wang, Y., Vandrovcova, J., Guelfi, S., Witeolar, A., Karch, C.M., Schork, A.J., Fan, C.C., Brewer, J.B., Momeni, P., et al.; International FTD-Genomics Consortium (IFGC); International Parkinson's Disease Genomics Consortium (IPDGC); and International Genomics of Alzheimer's Project (IGAP) (2017). Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. J. Neurol. Neurosurg. Psychiatry *88*, 152–164.
13. Broce, I., Karch, C.M., Wen, N., Fan, C.C., Wang, Y., Tan, C.H., Kouri, N., Ross, O.A., Höglinger, G.U., Muller, U., et al.; International FTD-Genomics Consortium (2018). Immune-related genetic enrichment in frontotemporal dementia: An analysis of genome-wide association studies. PLoS Med. *15*, e1002487.
14. Yokoyama, J.S., Wang, Y., Schork, A.J., Thompson, W.K., Karch, C.M., Cruchaga, C., McEvoy, L.K., Witoelar, A., Chen, C.H., Holland, D., et al.; Alzheimer's Disease Neuroimaging Initiative (2016). Association Between Genetic Traits for Immune-Mediated Diseases and Alzheimer's disease. JAMA Neurol. *73*, 691–697.
15. Desikan, R.S., Schork, A.J., Wang, Y., Witoelar, A., Sharma, M., McEvoy, L.K., Holland, D., Brewer, J.B., Chen, C.H., Thompson, W.K., et al.; ADNI, ADGC, GERAD, CHARGE and IPDGC Investigators (2015). Genetic overlap between Alzheimer's disease and Parkinson's disease at the MAPT locus. Mol. Psychiatry *20*, 1588–1595.
16. Cochran, J.N., McKinley, E.C., Cochran, M., Amaral, M.D., Moyers, B.A., Lasseigne, B.N., Gray, D.E., Lawlor, J.M.J., Prokop, J.W., Geier, E.G., et al. (2019). Genome sequencing for early-onset or atypical dementia: high diagnostic yield and frequent observation of multiple contributory alleles. Cold Spring Harb. Mol. Case Stud. *5*, a003491. https://doi.org/10.1101/mcs.a003491.
17. Bowling, K.M., Thompson, M.L., Amaral, M.D., Finnila, C.R., Hiatt, S.M., Engel, K.L., Cochran, J.N., Brothers, K.B., East, K.M., Gray, D.E., et al. (2017). Genomic diagnosis for children with intellectual disability and/or developmental delay. Genome Med. *9*, 43.
18. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
19. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034.

20. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
21. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
22. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.
23. Paila, U., Chapman, B.A., Kirchner, R., and Quinlan, A.R. (2013). GEMINI: integrative exploration of genetic variation and genome annotations. PLoS Comput. Biol. *9*, e1003153.
24. Pedersen, B.S., Collins, R.L., Talkowski, M.E., and Quinlan, A.R. (2017). Indexcov: fast coverage quality control for whole-genome sequencing. Gigascience *6*, 1–6.
25. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.
26. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
27. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
28. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
29. Tan, A., Abecasis, G.R., and Kang, H.M. (2015). Unified representation of genetic variants. Bioinformatics *31*, 2202–2204.
30. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
31. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.
32. NHLBI; and University of Michigan (2018). The NHLBI TransOmics for Precision Medicine (TOPMed) Whole Genome Sequencing Program. BRAVO variant browser (University of Michigan). https://bravo.sph.umich.edu/freeze5/hg38/.
33. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics *30*, 1006–1007.
34. Liu, X., White, S., Peng, B., Johnson, A.D., Brody, J.A., Li, A.H., Huang, Z., Carroll, A., Wei, P., Gibbs, R., et al. (2016). WGSA: an annotation pipeline for human genome sequencing studies. J. Med. Genet. *53*, 111–112.
35. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
36. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. https://doi.org/10.1101/531210.
37. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.
38. Lu, Q., Powles, R.L., Abdallah, S., Ou, D., Wang, Q., Hu, Y., Lu, Y., Liu, W., Li, B., Mukherjee, S., et al. (2017). Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. PLoS Genet. *13*, e1006933.
39. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93.
40. Lee, S., Fuchsberger, C., Kim, S., and Scott, L. (2016). An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. Biostatistics *17*, 1–15.
41. Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. https://doi.org/10.1101/005165.
42. Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Cedarbaum, J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., et al.; Alzheimer's Disease Neuroimaging Initiative (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. Alzheimers Dement. *11*, e1–e120.
43. Saykin, A.J., Shen, L., Foroud, T.M., Potkin, S.G., Swaminathan, S., Kim, S., Risacher, S.L., Nho, K., Huentelman, M.J., Craig, D.W., et al.; Alzheimer's Disease Neuroimaging Initiative (2010). Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimers Dement. *6*, 265–273.
44. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jr., Jagust, W.J., Shaw, L.M., Toga, A.W., et al. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology *74*, 201–209.
45. Morris, J.C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology *43*, 2412–2414.
46. Williams, M.M., Storandt, M., Roe, C.M., and Morris, J.C. (2013). Progression of Alzheimer's disease as measured by Clinical Dementia Rating Sum of Boxes scores. Alzheimers Dement. *9* (1, Suppl), S39–S44.
47. O'Bryant, S.E., Lacritz, L.H., Hall, J., Waring, S.C., Chan, W., Khodr, Z.G., Massman, P.J., Hobson, V., and Cullum, C.M. (2010). Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. Arch. Neurol. *67*, 746–749.
48. Lattante, S., Ciura, S., Rouleau, G.A., and Kabashi, E. (2015). Defining the genetic connection linking amyotrophic lateral sclerosis (ALS) with frontotemporal dementia (FTD). Trends Genet. *31*, 263–273.
49. Project MinE ALS Sequencing Consortium (2018). Project MinE: study design and pilot analyses of a large-scale

whole-genome sequencing study in amyotrophic lateral sclerosis. Eur. J. Hum. Genet. *26*, 1537–1546.

50. Bis, J.C., Jian, X., Kunkle, B.W., Chen, Y., Hamilton-Nelson, K.L., Bush, W.S., Salerno, W.J., Lancour, D., Ma, Y., Renton, A.E., et al.; Alzheimer's Disease Sequencing Project (2018). Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. Mol. Psychiatry. https://doi.org/10.1038/s41380-018-0112-7.
51. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. *12*, 189–198.
52. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A., et al.; Alzheimer's disease Genetics Consortium (ADGC); European Alzheimer's Disease Initiative (EADI); Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE); and Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES) (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat. Genet. *51*, 414–430.
53. Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. *19*, 371–384.
54. Heyward, F.D., and Sweatt, J.D. (2015). DNA Methylation in Memory Formation: Emerging Insights. Neuroscientist *21*, 475–489.
55. Antunes, C., Sousa, N., Pinto, L., and Marques, C.J. (2019). TET enzymes in neurophysiology and brain function. Neurosci. Biobehav. Rev. *102*, 337–344.
56. Fransquet, P.D., Wrigglesworth, J., Woods, R.L., Ernst, M.E., and Ryan, J. (2019). The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis. Clin. Epigenetics *11*, 62.
57. Jessop, P., and Toledo-Rodriguez, M. (2018). Hippocampal *TET1* and *TET2* Expression and DNA Hydroxymethylation Are Affected by Physical Exercise in Aged Mice. Front. Cell Dev. Biol. *6*, 45.
58. Gontier, G., Iyer, M., Shea, J.M., Bieri, G., Wheatley, E.G., Ramalho-Santos, M., and Villeda, S.A. (2018). Tet2 Rescues Age-Related Regenerative Decline and Enhances Cognitive Function in the Adult Mouse Brain. Cell Rep. *22*, 1974–1981.
59. Li, X., Yao, B., Chen, L., Kang, Y., Li, Y., Cheng, Y., Li, L., Lin, L., Wang, Z., Wang, M., et al. (2017). Ten-eleven translocation 2 interacts with forkhead box O3 and regulates adult neurogenesis. Nat. Commun. *8*, 15903.
60. Mi, Y., Gao, X., Dai, J., Ma, Y., Xu, L., and Jin, W. (2015). A Novel Function of TET2 in CNS: Sustaining Neuronal Survival. Int. J. Mol. Sci. *16*, 21846–21857.
61. Beck, D.B., Petracovici, A., He, C., Moore, H.W., Louie, R.J., Ansar, M., Douzgou, S., Sithambaram, S., Cottrell, T., Santos-Cortez, R.L.P., et al. (2020). Delineation of a Human Mendelian Disorder of the DNA Demethylation Machinery: TET3 Deficiency. Am. J. Hum. Genet. *106*, 234–245.
62. Carrillo-Jimenez, A., Deniz, Ö., Niklison-Chirou, M.V., Ruiz, R., Bezerra-Salomão, K., Stratoulias, V., Amouroux, R., Yip, P.K., Vilalta, A., Cheray, M., et al. (2019). TET2 Regulates the Neuroinflammatory Response in Microglia. Cell Rep. *29*, 697–713.e8.
63. Ewers, M., Franzmeier, N., Suárez-Calvet, M., Morenas-Rodriguez, E., Caballero, M.A.A., Kleinberger, G., Piccio, L., Cruchaga, C., Deming, Y., Dichgans, M., et al.; Alzheimer's Disease Neuroimaging Initiative (2019). Increased soluble TREM2 in cerebrospinal fluid is associated with reduced cognitive and clinical decline in Alzheimer's disease. Sci. Transl. Med. *11*, 11.
64. Li, L., Qiu, Y., Miao, M., Liu, Z., Li, W., Zhu, Y., and Wang, Q. (2020). Reduction of Tet2 exacerbates early stage Alzheimer's pathology and cognitive impairments in 2 x Tg-ad mice. Hum. Mol. Genet. https://doi.org/10.1093/hmg/ddz282.
65. Pottier, C., Wallon, D., Rousseau, S., Rovelet-Lecrux, A., Richard, A.C., Rollin-Sillaire, A., Frebourg, T., Campion, D., and Hannequin, D. (2013). TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease. J. Alzheimers Dis. *35*, 45–49.
66. Cuyvers, E., Bettens, K., Philtjens, S., Van Langenhove, T., Gijselinck, I., van der Zee, J., Engelborghs, S., Vandenbulcke, M., Van Dongen, J., Geerts, N., et al.; BELNEU consortium (2014). Investigating the role of rare heterozygous TREM2 variants in Alzheimer's disease and frontotemporal dementia. Neurobiol. Aging *35*, 726.e11–726.e19.
67. Nguyen, H.P., Van Broeckhoven, C., and van der Zee, J. (2018). ALS Genes in the Genomic Era and their Implications for FTD. Trends Genet. *34*, 404–423.
68. Orme, T., Guerreiro, R., and Bras, J. (2018). The Genetics of Dementia with Lewy Bodies: Current Understanding and Future Directions. Curr. Neurol. Neurosci. Rep. *18*, 67.
69. van Blitterswijk, M., Mullen, B., Wojtas, A., Heckman, M.G., Diehl, N.N., Baker, M.C., DeJesus-Hernandez, M., Brown, P.H., Murray, M.E., Hsiung, G.Y., et al. (2014). Genetic modifiers in carriers of repeat expansions in the C9ORF72 gene. Mol. Neurodegener. *9*, 38.
70. Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., et al. (1998). Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. Nature *393*, 702–705.
71. Cruts, M., Theuns, J., and Van Broeckhoven, C. (2012). Locus-specific mutation databases for neurodegenerative brain diseases. Hum. Mutat. *33*, 1340–1344.
72. Rademakers, R., Dermaut, B., Peeters, K., Cruts, M., Heutink, P., Goate, A., and Van Broeckhoven, C. (2003). Tau (MAPT) mutation Arg406Trp presenting clinically with Alzheimer's disease does not share a common founder in Western Europe. Hum. Mutat. *22*, 409–411.
73. Reed, L.A., Grabowski, T.J., Schmidt, M.L., Morris, J.C., Goate, A., Solodkin, A., Van Hoesen, G.W., Schelper, R.L., Talbot, C.J., Wragg, M.A., and Trojanowski, J.Q. (1997). Autosomal dominant dementia with widespread neurofibrillary tangles. Ann. Neurol. *42*, 564–572.
74. Kwok, J.B., Teber, E.T., Loy, C., Hallupp, M., Nicholson, G., Mellick, G.D., Buchanan, D.D., Silburn, P.A., and Schofield, P.R. (2004). Tau haplotypes regulate transcription and are associated with Parkinson's disease. Ann. Neurol. *55*, 329–334.
75. Bennett, D.A., Buchman, A.S., Boyle, P.A., Barnes, L.L., Wilson, R.S., and Schneider, J.A. (2018). Religious Orders Study and Rush Memory and Aging Project. J. Alzheimers Dis. *64* (s1), S161–S189.
76. Allen, M., Carrasquillo, M.M., Funk, C., Heavner, B.D., Zou, F., Younkin, C.S., Burgess, J.D., Chai, H.S., Crook, J., Eddy, J.A., et al. (2016). Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci. Data *3*, 160089.