THEORETICAL ARTICLE

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

Harnessing the paradoxical phenotypes of APOE ε2 and APOE ε4 to identify genetic modifiers in Alzheimer's disease

Young Won Kim[1] | Ismael Al-Ramahi[2,3] | Amanda Koire[4,5] | Stephen J. Wilson[6] |
Daniel M. Konecki[4] | Samantha Mota[3] | Shirin Soleimani[3] | Juan Botas[2,3,4] |
Olivier Lichtarge[1,2,3,4,5,6]

[1] Program in Integrative Molecular and Biomedical Sciences, Baylor College of Medicine, Houston, Texas, USA

[2] Jan and Dan Duncan Neurological Research Institute, Houston, Texas, USA

[3] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA

[4] Graduate Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, Texas, USA

[5] Medical Scientist Training Program, Baylor College of Medicine, Houston, Texas, USA

[6] Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas, USA

Correspondence
Olivier Lichtarge, Professor, Molecular and Human Genetics, One Baylor Plaza, BCM225, Houston, TX 77030, USA.
E-mail: lichtarge@bcm.edu

Young Won Kim and Ismael Al-Ramahi contributed equally to this study.

Funding information
Oskar Fischer Foundation to OL; Huffington Foundation to JB; National Institute of Aging, Grant/Award Numbers: R01 AG061105 to OL, R01 AG057339 to JB; National Institute of General Medical Sciences, Grant/Award Numbers: R01 GM079656 to OL, R01 GM066099 to OL; Darrell K Royal Research Fund to IA

Abstract

The strongest genetic risk factor for idiopathic late-onset Alzheimer's disease (LOAD) is apolipoprotein E (*APOE*) ε4, while the *APOE* ε2 allele is protective. However, there are paradoxical *APOE* ε4 carriers who remain disease-free and *APOE* ε2 carriers with LOAD. We compared exomes of healthy *APOE* ε4 carriers and *APOE* ε2 Alzheimer's disease (AD) patients, prioritizing coding variants based on their predicted functional impact, and identified 216 genes with differential mutational load between these two populations. These candidate genes were significantly dysregulated in LOAD brains, and many modulated tau- or β42-induced neurodegeneration in *Drosophila*. Variants in these genes were associated with AD risk, even in *APOE* ε3 homozygotes, showing robust predictive power for risk stratification. Network analyses revealed involvement of candidate genes in brain cell type-specific pathways including synaptic biology, dendritic spine pruning and inflammation. These potential modifiers of LOAD may constitute novel biomarkers, provide potential therapeutic intervention avenues, and support applying this approach as larger whole exome sequencing cohorts become available.

KEYWORDS
apolipoprotein E, *Drosophila* models, late-onset Alzheimer's disease, paradoxical phenotypes

1 | NARRATIVE

As pathological changes that lead to late-onset Alzheimer's disease (LOAD) begin well before the presentation of clinical symptoms,[1] predicting those at risk remains an important challenge for developing effective prevention, treatment, and clinical trials. Despite the many loci associated with LOAD so far, we are still short of a robust, unified method of risk stratification based on genome variations. This difficulty mainly stems from the clinical heterogeneity and polygenic nature of the disease. Because LOAD manifests late in life, it often overlaps with

RESEARCH IN CONTEXT

1. **Systematic review**: Despite considerable insight from genome-wide association studies (GWAS) and sequencing studies, much of late-onset Alzheimer's disease (LOAD) heritability remains unexplained leading to uncertain risk stratification of patients and no available disease-modifying therapies. To date, the apolipoprotein E (*APOE*) gene remains the strongest genetic risk factor for LOAD. Carriers of the *APOE* ε4 allele are at a greater risk, while the *APOE* ε2 allele plays a protective role. However, many *APOE* ε4 carriers remain disease free and some *APOE* ε2 carriers develop LOAD. Genetic modifiers that override the effects of *APOE* alleles may explain these paradoxical cases and suggest attractive candidate biomarkers and therapeutic targets, while pointing to novel risk/protective alleles. However, to this day the search for such modifiers has remained beyond the reach of conventional case-control studies.

2. **Interpretation**: Using a novel approach that predicts the functional impact of coding variants and a sequential regression analysis on the largest LOAD whole exome sequencing (WES) dataset (Alzheimer's Disease Sequencing Project), we identified 216 genes that showed differential mutational load between healthy *APOE* ε4 carriers and *APOE* ε2 Alzheimer's disease (AD) patients. These genes were significantly dysregulated in the AD transcriptome and highly connected to known LOAD genes identified by GWAS. Furthermore, many of the genes were relevant in vivo as they ameliorated neurodegeneration caused by tau and secreted $\beta42$ using well-validated Drosophila models. Some identified genes had variants significantly associated with risk beyond the *APOE* ε2 or ε4 individuals extending into the ε3 homozygotes. The 216 identified genes could constitute risk/protective factors, biomarkers, and even therapeutic targets, as many of them are druggable. Importantly, variants in the identified genes showed robust predictive power for patient stratification within the ε2 and ε4 population.

3. **Future directions**: Our study has identified specific coding variants in genes not previously known to be associated with LOAD. Our future work will focus on identifying what the specific effect of each variant is on protein function to better decipher the role they play in LOAD pathogenesis, with particular focus on the druggable targets whose knockdown ameliorated neurodegeneration. As network analyses reveal a potential involvement of the identified genes in a variety of brain cell type-specific pathways ranging from stress granules and synaptic biology to dendritic spine pruning and inflammatory response, our future work will focus on deciphering the specific role of the identified genes in those pathways. As available WES expands, we will be able to replicate this approach in larger cohort studies and in more diverse populations to identify genes that robustly enhance risk stratification and therapeutic target discovery.

other neuropathologies, vascular disorders, and age-associated cognitive impairments.[2,3] Typically, a definitive LOAD diagnosis is performed *post mortem* by neuropathological analysis. Here too, there is variability in the type and number of structures observed (plaques, tangles, and other protein aggregates) and their diagnostic potential,[3] because these deposits are also observed in normal aging and their frequency can overlap between non-LOAD and LOAD patients.[4–6] One unifying characteristic of all LOAD patients is the severe loss of synapses and neurons in the cortex that correlates with the devastating loss of cognitive capacity.[7–12] With respect to the polygenicity of LOAD, estimates of heritability have been as high as 80% in twin studies,[13] with more recent reports using genetic variance analyses being closer to 50%.[14] While genome-wide association studies (GWAS) on increasingly larger cohorts have revealed more than 40 LOAD susceptibility loci,[15–21] much of the LOAD genetic risk remains unaccounted for,[22,23] even as we approach the power limit of traditional GWAS methods. Furthermore, an analysis of the molecular and biological functions of the LOAD susceptibility loci reveals the involvement of multiple pathways and cellular processes. Together, these polygenic phenomena point to multiple pathways contributing to dementia.

Among risk factors for LOAD, the apolipoprotein E (*APOE*) gene remains the strongest genetic risk factor.[24] Clinical and autopsy-based studies have shown that, among Whites, heterozygous *APOE* ε4 carriers (ε4/ε2 or ε4/ε3) are two to three times more likely to develop LOAD

than non-carriers while in homozygotes (ε4/ε4) the risk can be > 14-fold higher,[25] with ≈40% of all LOAD cases carrying the ε4 allele.[26] Conversely, the ε2 allele of *APOE* (either ε2/ε3 or ε2/ε2) plays a protective role, with ≈40% decrease in chance of developing LOAD. However, whether *APOE* ε4 or ε2 result in a loss or a gain of function remains unclear,[27] and the specific biological pathways they affect in the context of AD remain to be fully elucidated. Although *APOE* status is such a robust LOAD risk predictor, the pathogenicity of the ε4 allele and the protection of ε2 are not completely penetrant. Many ε4 carriers remain disease free and some ε2 carriers develop Alzheimer's disease (AD), raising the question of what keeps these paradoxical individuals from manifesting the effects typically associated with their respective alleles? The answer may be a combination of environmental and genetic factors, but to our knowledge, no study has attempted to answer this question. We hypothesize here that a number of paradoxical individuals harbor variants in other genes that overcome the benefit of ε2 or neutralize the pathogenesis of ε4. Identifying these variants would provide a rich set of hypotheses that may (1) reveal novel LOAD susceptibility genes and protective alleles, (2) highlight potential therapeutic targets and pathways, and (3) provide geneticists and neurologists with biomarkers for patient prognosis and stratification.

Here, to search for genetic variants underlying the *APOE* paradoxical population we tapped into the largest whole exome dataset currently available for AD, Alzheimer's Disease Sequencing Project

(ADSP; dbGaP accession: phs000572.v7.p4; $n = 5686$). To maximize power, we included both heterozygotes and homozygotes, assuming ε3 is neutral (see Figure S1 in supporting information for demographics). This resulted in 179 AD patients carrying *APOE* ε2/ε2 or ε2/ε3 (henceforth designated AD-ε2) and 301 healthy individuals with *APOE* ε4/ε4 or ε3/ε4 (henceforth designated HC-ε4). As age is the most important risk factor for LOAD, we confirmed that healthy individuals were on average significantly older than AD patients at the time of diagnosis, in both the *APOE* ε4 population and the *APOE* ε2 population (Figure S1). The small size of these populations lacks the power to perform GWAS. The complex, polygenic nature of LOAD suggests that few additional susceptibility genes may be identified by searching for a sheer excess of mutations across a cohort. In addition, most GWAS LOAD loci fall in non-coding regions, making it hard to decipher the specific genes affected and therefore the functional consequences. For these reasons, we focused on the coding sequences using a method that predicts the functional impact of amino acid (AA) substitutions: evolutionary action (EA).[28]

EA is unlike methods that train on large datasets of human variations and apply supervised machine learning algorithms to weigh mutational impact. Instead, it interprets mutations of interest by tapping a vast but usually unused reservoir of phylogenomic data coupling sequence variations to functional divergences across species. In practice, EA is computed as a continuous score normalized to range from 0 (low impact) to 100 (high impact), and produced by multiplying the evolutionary importance, or *weight*, of an individual protein sequence position with the size of the evolutionary perturbation, or *move*, caused by a mutation at that position.[28] As a product of a force, describing mutational resistance, and of a move, describing mutational displacement, EA becomes interpretable as the *work* of a mutation as it pushes a genome along in the evolutionary landscape. This approach has three main advantages: (1) EA consistently performs well against state-of-the-art predictors of mutational impact in objective, blinded challenges;[29] (2) multiple experimental applications confirmed its utility, including in clinical contexts;[30–33] and (3) finally, because EA is based on basic principles of evolution rather than any specific training datasets, it is general, unbiased, and broadly applicable, even across different species.

In this study, we used EA to compare the mutational burden for each gene between the two paradoxical patient groups through a series of linear regressions and identified genes with differential imputed deviation in EA load (iDEAL; Figure 1A; see Methods). How iDEAL works can be exemplified by *TREM2*, a well-known AD susceptibility gene.[26–29] *TREM2* is identified by iDEAL as one of the genes underlying the AD-ε2/HC-ε4 paradox, mainly due to the potentially pathogenic R47H (EA = 55.66), found significantly more often in AD-ε2 than in HC-ε4 individuals (odds ratio [OR] = 6.99; $P = .015$). However, iDEAL also takes into account other *TREM2* variants as well. For example, the more impactful variant T96K (EA = 96.59) may also overcome the protective effect of ε2 (OR = 5.11), though it is rarer and does not reach significance on its own. EA analysis also includes some less impactful *TREM2* variants (L211P with EA = 13.21, and V27M with EA = 6.43), as well as nonsense variants (Q33X with EA = 100). Therefore, even *de novo* and/or extremely rare variants that would not reach GWAS signif-

icance are used by iDEAL to create an aggregate score for each gene. In total, iDEAL identified 216 genes with significant differences in mutational EA load associated with the paradoxical AD-ε2/HC-ε4 phenotypes (Table S1 in supporting information).

These iDEAL genes can be separated in two groups, those carrying potentially protective variants in healthy ε4 individuals (68 genes henceforth referred to as HCε4-iDEAL) and those enriched in potentially pathogenic variants that overcome the ε2 protection in AD-ε2 patients (148 genes henceforth referred to as ADε2-iDEAL). As a validation, we compared the EA score distributions in the ADε2-iDEAL genes and in the HCε4-iDEAL genes in AD versus HC. We reasoned that if the ADε2-iDEAL genes foster LOAD pathogenesis, they would be enriched for high-impact variants in AD patients. Likewise, if the HCε4-iDEAL genes keep ε4 carriers healthy, then they should be depleted of high-impact variants in AD patients. As predicted, ADε2-iDEAL genes were enriched for high-impact variants in patients compared to HC (Figure S2A in supporting information, red, $P = 1.1E-06$; Kolmogorov-Smirnov [K-S] test), and the HCε4-iDEAL genes were depleted of high-impact variants in patients (Figure S2A, blue, $P = 2.6E-07$; K-S test). Next, we performed the same analysis on either ε2 or ε4 carriers. Consistent with the previous result, we found that the potentially pathogenic ADε2-iDEAL variants were enriched in AD-ε2 individuals compared to HCε2 (Figure S2B, red, $P = 1.0E-14$; K-S test) while the potentially protective HCε4-iDEAL variants were depleted in AD-ε4 patients compared to HC-ε4 (Figure S2B, blue, $P = 4.1E-06$; K-S test). Furthermore, we found that AD-ε3 patients were depleted of high-impact variants in HCε4-iDEAL genes (Figure S2C, blue, $P = .015$; K-S test) compared to HC-ε3 and showed a trend for enrichment of high-impact variants in ADε2-iDEAL genes, though it did not reach statistical significance (Figure S2C, red, $P = .25$; K-S test). Taken together, these results indicate that the variants responsible for the paradoxical AD-ε2/HC-ε4 phenotypes are significantly linked to AD status and may be causative of neurodegeneration or provide protection regardless of *APOE* genotype.

To further investigate the relevance of iDEAL genes in AD biology beyond *APOE*, we asked whether they are connected to known AD-related changes and risk factors. First, we assessed their expression in AD brains of the Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) sequence repository.[34–39] Out of the 174 iDEAL genes present in the AMP-AD data, 75 genes were significantly dysregulated (adjusted P-value < .01) in AD patients compared to controls in at least one brain region (Figure 1B). This statistically significant enrichment ($P = .018$; hypergeometric test) indicates that expression of many iDEAL genes either respond to or are causative of AD-related insults, supporting that iDEAL genes may act as modifiers of *APOE* through a broader role in AD pathology. Next, we measured the degree of connectivity between the iDEAL genes and other AD susceptibility GWAS candidates[15–21] using the STRING v11 database. We found that the set of iDEAL genes were significantly more connected to AD-GWAS candidates than expected by random chance (Figure 1C, Figure S4A in supporting information, z-score = 2.24). Moreover, seven iDEAL genes (*GOLGA5, PTBP1, SYTL2, SMARCD2, GAMT, TREM2, AZU1*) fall within ±500kb of linkage disequilibrium (LD) regions for each locus
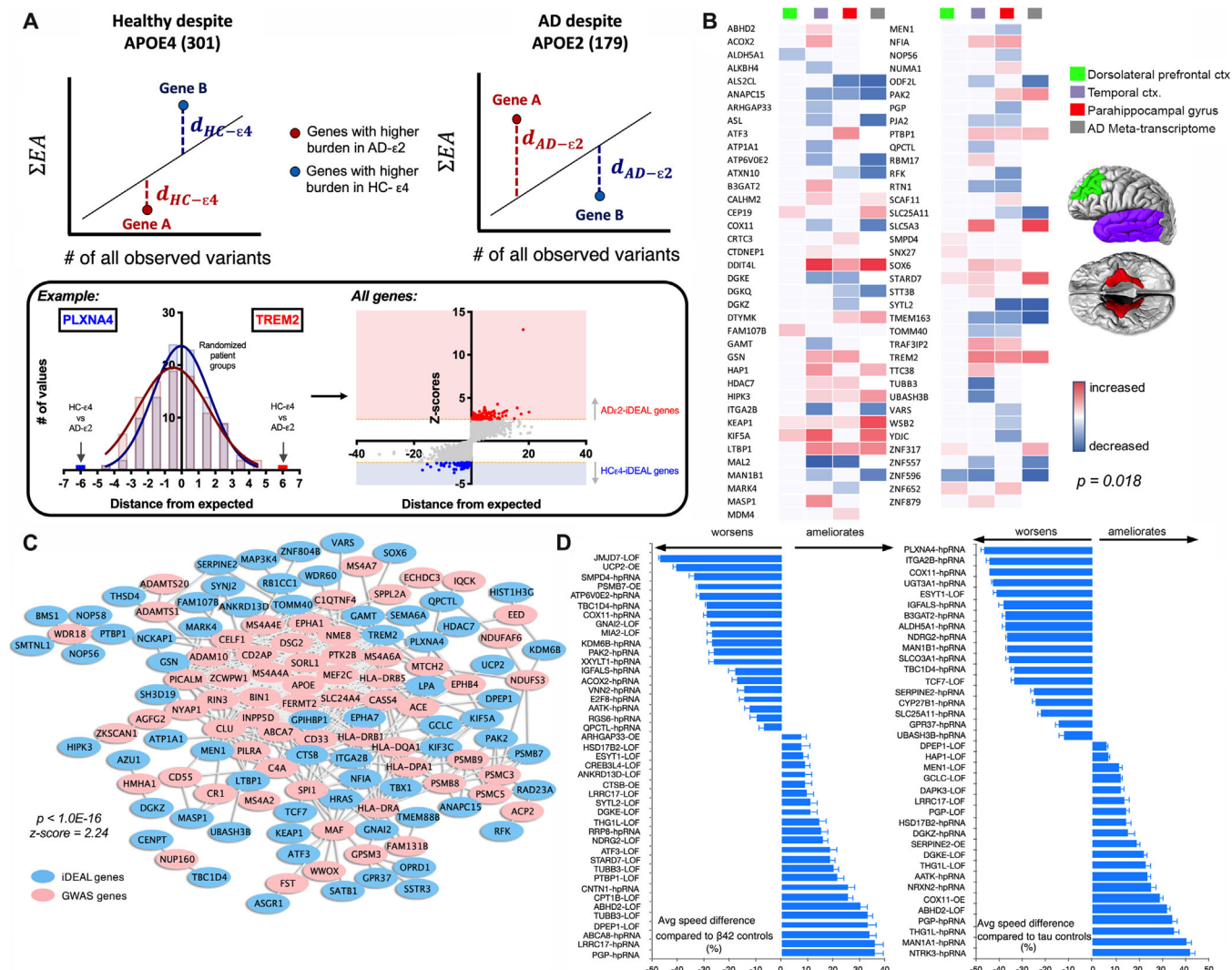
FIGURE 1 Identification and validation of genes with differential functional mutational load in the paradoxical AD-ε2 versus HC-ε4 population. (A) For each gene, the mutational burden, defined as the sum of all evolutionary action (EA) scores, was plotted against the number of all coding variants observed in that gene, and a regression line was fitted to establish the expected mutational burden given background mutation rate across the two paradoxical patient groups. Then, in each patient group (AD-ε2 vs HC-ε4), the observed mutational burden and the background mutation rate for each gene were plotted, and the distance (*d*) from regression line was measured and compared. To control for noise from passive mutations, we assessed the significance of each gene's signal by calculating a *z*-score from randomizing the labels of AD-ε2 and HC-ε4 patients 100 times to build a background distribution of imputed deviation in evolutionary action load (iDEAL) scores; 216 genes had an absolute value of *z*-score above 2.5 (>99th percentile). For more detail, refer to Detailed Methods. (B) Enrichment of iDEAL genes for differentially expressed genes (DEG) in Alzheimer's disease (AD) brains from Accelerating Medicines Partnership-Alzheimer's Disease sequence repository. DEG were defined as genes significantly up- or downregulated using adjusted *P*-value cutoff of .01 in at least one brain region. Hypergeometric test was used to assess enrichment. (C) Enrichment of iDEAL genes for first-degree neighbors of genome-wide association study AD genes. STRING v11 was used to construct protein-protein interaction (PPI) network. Interaction sources are from Textmining, Experiments, and Databases, with interaction scores above 0.400. For calculation of z-score, see Figure S4A in supporting information. (D) The average worsening or amelioration (%) of neurodegeneration measured as the loss in climbing speed of *Drosophila* expressing either secreted *β*42 (left) or human 2N4R tau (right) together with the indicated allele in the *Drosophila* homolog of the gene shown. *β*42/scramble or tau/scramble animals are used as the reference (error bars indicate standard deviation). Seven genes that showed conflicting evidence are not included

found in a recent GWAS by Kunkle et al.[21] These data suggest that the iDEAL candidates act through pathways also affected by known AD susceptibility loci.

The main neuropathological hallmark of AD is the severe loss of synapses and neurons. The above data indicate a link between the 216 genes with higher mutational load in the paradoxical patients and neu-

rodegeneration. To investigate the capacity of each of the identified candidates to modulate neurodegeneration in vivo, we used two well-characterized *Drosophila* models that express either human wild-type 2N4R tau[40] or secreted *β*42 peptide pan-neuronally.[41] In these animals, tau is hyperphosphorylated and forms ALZ50 positive intracellular aggregates while *β*42 is detected both in its soluble form (by IP) and

in Th-S positive extracellular aggregates (data not shown). Importantly, these models present progressive behavioral (ie, locomotor) impairments that can be precisely assessed longitudinally in aging animals and provide a quantitative functional assay of neuronal dysfunction and neurodegeneration. We used an automated data acquisition system that enables high-throughput assessment of movement metrics from video-recorded trajectories of individual animals.[42] Our previous work showed that genes modulating AD neuropathology in *Drosophila* were validated in mouse and human cell AD models,[40,43] supporting the validity of these models. We tested all 134 genes with available overexpression and/or loss of function alleles (classical or shRNA) and identified 69 genes that when modulated worsened or ameliorated tau- or β42-induced neuronal dysfunction ($P = 1.13e-107$; hypergeometric test) (Figure 1D, Figures S5-S8, and Table S2 in supporting information). Among these genes, knockdown of *Drosophila* homologs of *ATP6V0E2*, *COX11*, *E2F8*, *IGFALS*, and *TBC1D4* exacerbated neuronal dysfunction in both models while knockdown of the *Drosophila* homologs of *ABHD2*, *CNTN1*, *DGKE*, *DPEP1*, *HSD17B2*, *LRRC17*, *PGP*, *RRP8*, and *THG1L* ameliorated neurodegeneration in both tau and β42 animals, indicating that these genes may play shared roles in mechanisms underlying the AD pathogenesis. Previous analysis of *APOE* ε4 carriers revealed a positive correlation of *APOE* ε4 and the presence of neurofibrillary tangles and amyloid plaques.[44,45] Furthermore, experimental data in mice and induced pluripotent stem cell (iPSC)-derived neurons indicates that *APOE* ε4 can increase phosphorylation and deposition of tau as well as promote the production and impair the clearance of amyloid.[46-51] Note that even though *Drosophila* does not have a clear *APOE* homolog, the kinases leading to tau phosphorylation are conserved in *Drosophila*, and the same is true for the proteases that can degrade amyloid extracellularly (neprilysin and insulin degrading enzyme) and intracellularly (lysosome and proteasome). Therefore, some of the genes that show modifier effects in *Drosophila* may act as pathogenic effectors or protective factors of tau and amyloid in humans. Other modifiers may act on pathways linking *APOE* ε4 to tau and amyloid accumulation. A third class of modifiers may be acting in parallel to *APOE* ε4 as a comorbidity risk in the case of those potentiating neurodegeneration or as generally neuroprotective alleles. In line with this, it is noteworthy that in addition to LOAD, *APOE* ε4 is also a risk factor for other diseases.[45] The observed enrichment in modifiers in *Drosophila* is even more remarkable, if we consider the physiological differences that exist with mammalians that surely preclude us from mimicking the effect of some iDEAL candidates in the fruit fly system (eg, *Drosophila* lacks blood vessels in their brain).

Generally, higher EA corresponds to higher predicted impact of variants on protein function. While EA alone cannot precisely predict loss-of-function (LOF) versus gain-of-function (GOF) of variants, experimental testing in *Drosophila* increases our confidence in the iDEAL genes and informs how specific genetic variants may be affecting each iDEAL gene—LOF versus GOF—and in turn, indicates potential means of therapeutic intervention (inhibition or activation). For example, we observed that reduced *COX11* function in *Drosophila* worsens both *tau*- and *β42*-induced neurodegeneration while *COX11* overexpression ameliorates tau-induced deficits. In the human data,

its variant (V223G with EA = 52.34) had an OR of 0.60. Taken together, this suggests the variant in this gene is likely to be GOF. Consistent with this, *COX11* encodes a mitochondrial protein involved in the terminal stage of cytochrome c oxidase synthesis and is under-expressed in AD-affected individuals.[52] Klotho (KL-VS)[53] and Christchurch[54] variants are known to be protective against *APOE* ε4. While these would serve as convincing true positives if picked up by the iDEAL method, neither variant was present in the studied dataset.

To assess which iDEAL genes may be good drug targets, we searched the Drug-Gene Interaction Database (DGIdb)[55] for pharmacological compounds that interact with these genes and found that 39 genes interact with 390 compounds (Table S3 in supporting information). Of these, three genes (*ITGA2B*, *ALDH5A*, and *HDAC7*) interact with two medications (enoxaparin and valproic acid) that have been associated with lower incidence of AD in a population study.[56] In addition, we searched PubMed for publications that co-mention the term "Alzheimer" and any of the 390 compounds. Three additional drugs had literature evidence for having potential neuroprotective effects in animal models: the cathepsin inhibitor LHVS,[57] URMC-099,[58,59] and CX-4945.[60,61] LHVS is an inhibitor of *CTSB*, and URMC-099 and CX-4945 are inhibitors of *DAPK3* (Table S3). Interestingly, inhibition of the *Drosophila* homologs of these two iDEAL genes also results in neuroprotection (Figure 1D). Given their robust effect and druggability, we believe these genes are examples of top candidates to characterize further in AD mouse models using these existing inhibitors as well as by genetic knockdown either using viral delivery of shRNAs or targeted CRISPR knock-out.

Next, we asked whether iDEAL genes could also be used for risk prediction and patient stratification. First, we used machine learning to test whether iDEAL gene variants could separate between the two paradoxical patient groups. In a five-fold cross-validation, the AdaBoost-SVM algorithm[62,63] trained on the mutational features represented by EA across the 216 iDEAL genes (see Detailed Methods) could classify AD-ε2 versus HC-ε4 individuals with an average area under the curve (AUC) of 0.92 (Figure 2A). To assess which of the 216 genes have the highest predictive power, we implemented permutation feature importance,[64] which pointed to 94 genes that contributed to risk prediction (Table S4 in supporting information). Surprisingly, 41 iDEAL genes were better classifiers for AD risk in the paradoxical AD-ε2/HC-ε4 population than *TREM2* (Figure 2B). As expected, many of these top genes had variants with significant OR in the AD-ε2/HC-ε4 comparison (Table S4). Interestingly, among the top genes, there were also five (*ZNF804B*, *SYTL2*, *AKNAD1*, *LAMA2*, *LRRC17*-K119E) that showed variants with OR > 1 in the AD-ε3 population, and eight genes with variants with OR < 1 in the AD-ε3 individuals (*LTBP1*, *AKNAD1*, *LRRC17*-G187A, *PRKAG3*, *MIA2*, *WDR60*, *SMTNL1*, and *ARHGAP33*). This further underscores the possibility that iDEAL candidates may play a role in AD in the broader population. Next, we tested a more clinically relevant question, namely, whether the variants in the 216 iDEAL genes could predict which of the *APOE* ε2 carriers would develop AD, and conversely, which *APOE* ε4 carriers would remain healthy. This would provide a useful tool for geneticists and neurologists to stratify individuals with
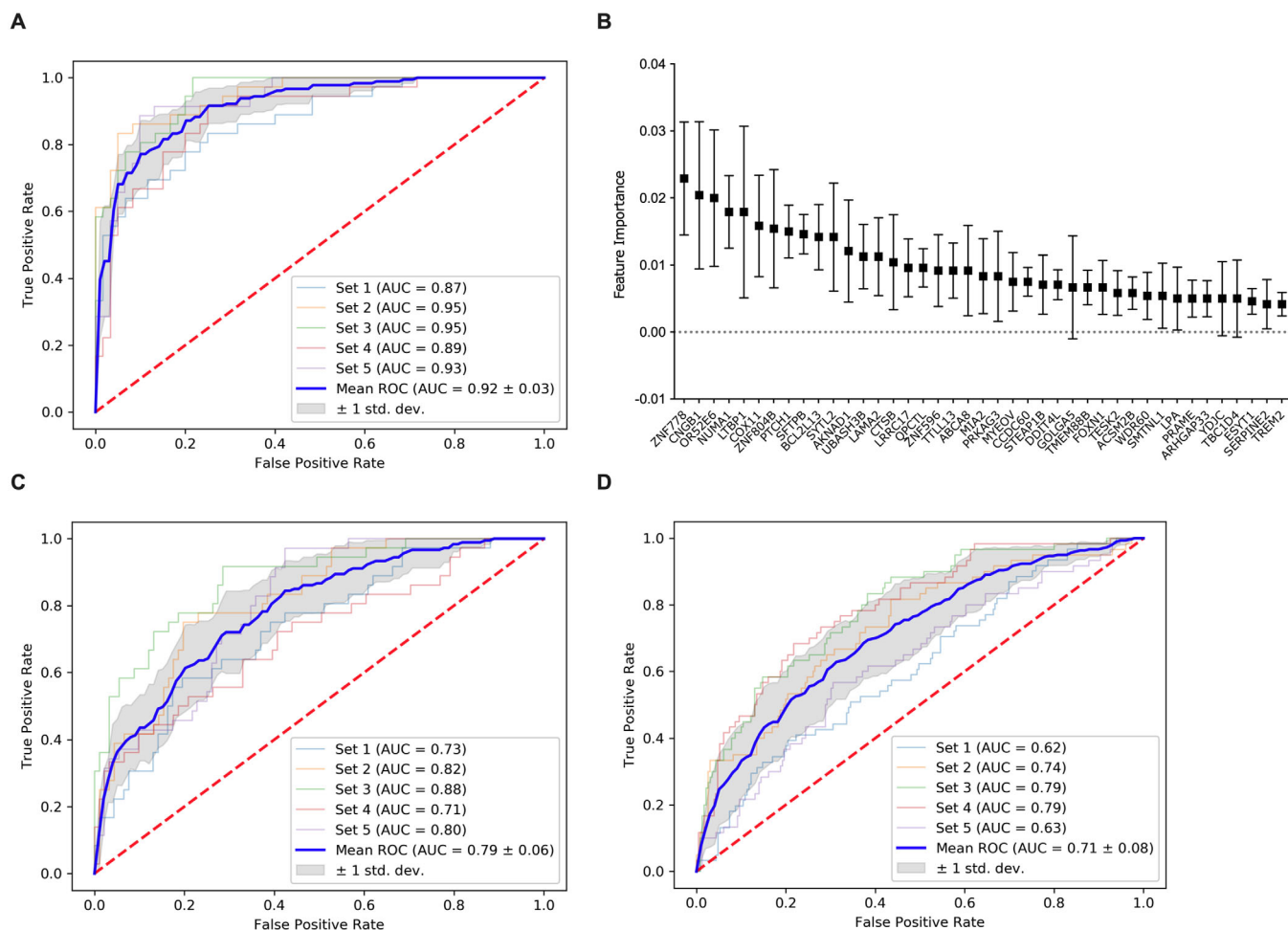
FIGURE 2 Receiver operating characteristic (ROC) curves for imputed deviation in evolutionary action load (iDEAL) genes as diagnostic markers. Adaboost-SVM algorithm was trained to classify individuals with (A) AD-ε2 versus HC-ε4, (C) AD-ε2 versus HC-ε2, and (D) HC-ε4 versus AD-ε4, using aggregate evolutionary action (EA) burden in the 216 iDEAL genes as features in a five-fold cross-validation. Blue line represents the mean ROC curve, and the gray area represents ±1 standard deviation (std. dev.). Red dotted line represents the ROC curve for a random classifier (area under the curve [AUC] = 0.50). (B) Permutation feature importance returned 94 genes that positively contributed to risk prediction. Data shown as mean ± standard error of the mean of five folds

these genotypes and select those at highest risk, for example, ahead of clinical trials. We again used the AdaBoost-SVM algorithm in five-fold cross-validation and could separate the AD-ε2 patients from the HC-ε2 individuals with an average AUC of 0.79 (Figure 2C). Likewise, we could stratify HC-ε4 individuals from the AD-ε4 patients with an average AUC of 0.71 (Figure 2D). While further validation is required when more exomes become available, these data suggest that the 216 iDEAL genes have the potential to be used as stratification biomarkers on top of the usual *APOE*-genotype risk prediction for AD.

Finally, we investigated the biological pathways in which the iDEAL genes are involved. We constructed a STRING-based network with our genes and applied the Markov cluster algorithm (MCA) to detect densely connected regions (inflation 1.9). There are likely multiple pathways toward dementia, but ultimately, synaptic dysfunction and substantial loss of neurons are major contributors of AD pathogenesis[65] and the most proximal event that leads to clinical features observed in AD. Indeed, of the 26 clusters iDEAL genes formed,

15 showed significant enrichment for biological pathways (Methods; Figure 3A; Table S5 in supporting information), many of which are relevant for synaptic integrity. For example, pathways such as vesicular and protein traffic, neuron projection and axon guidance, dendritic spine, microtubule transport and related processes are involved in normal axonal transport of proteins and organelles such as mitochondria, which is essential for healthy synapse function.[66] Specifically, the neuron projection and axon guidance pathway was of particular interest because three of the genes—*TREM2*, *PLXNA4*, and *PAK2*—have already been associated with AD.[67–69] *TREM2* is a well-known risk gene that has been studied extensively.[67,70–72] *PLXNA4*, which ranked top (highest z-score) among the HCε4-iDEAL genes, has previously been shown to have a protective role,[68] and *PAK2* has been implicated in AD synaptic dysfunction.[69] Other pathways consistent with *APOE* functions (ie, plasma lipoprotein remodeling, lipid catabolism, and vesicular traffic) were enriched in genes predictive of AD status in the paradoxical patient groups (*GPIHBP1*, *LPA*, *VNN2*, *SERPINE2*, *CPT1B*,
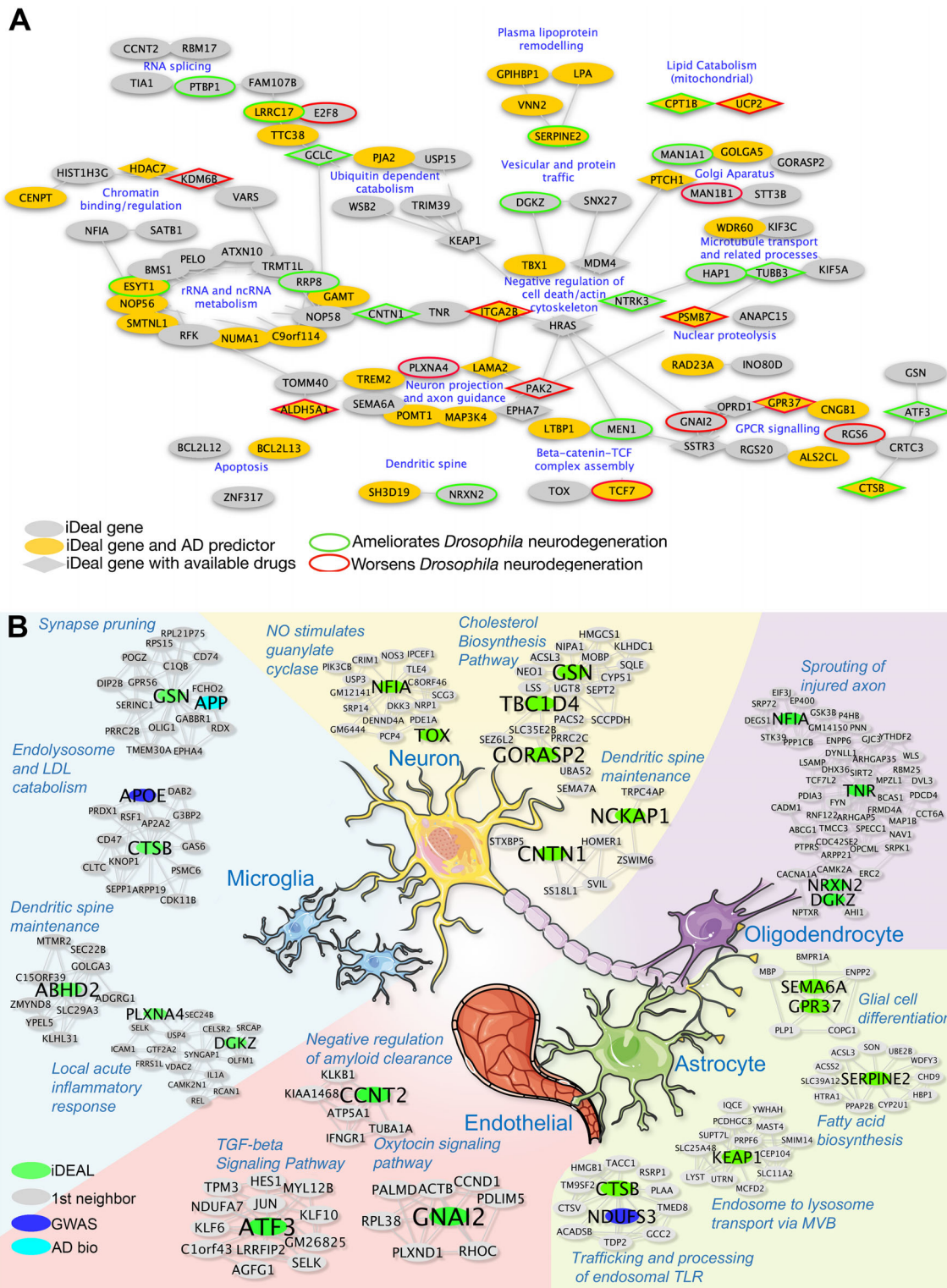
FIGURE 3 Network-based functional enrichment of imputed deviation in evolutionary action load (iDEAL) genes. (A) Networkbuilt using those genes among the 216 candidates which interacted with each other (stringency 0.4) using STRING. Gene modules were established by applying Markov cluster algorithm with an inflation of 1.9. Functional enrichment analysis was then performed for each cluster.15 out of 26 clusters showed functional enrichment (indicated in blue font) with FDR q-value < 0.05. (Table S5 in supporting information). Information on drug availability (rhomboid-shaped nodes), ability to ameliorate (green outline), or worsen (red outline) neurodegeneration in vivo, and whether the genes were among the most significant in patient stratification (orange nodes) are superimposed on the network. (B) Examples of coexpression communities and their functional enrichment in specific brain-cell types. Green nodes indicate iDEAL genes, gray nodes indicate their first degree coexpressed neighbors, light blue indicates first neighbors of ideal genes that are also central players in AD biology, and purple nodes are genome-wide association study candidates that are first coexpression neigbors of iDEAL genes. Edges represent weighed coexpression. Based on networks built by McKenzie et al.[73] (Cell images modified from Servier Medical Art in accordance with the Creative Commons license.)

UCP2, GOLGA5, and PTCH1). These data suggest that iDEAL genes are involved in pathways related to synaptic connection, which may be responsible for the paradoxical phenotypes we observe in AD-ε2 and HC-ε4 patients.

To gain brain-specific pathway information, we took advantage of cell type-specific coexpression networks.[73] Using the HiDef-Louvain algorithm, we identified the coexpression communities of iDEAL genes in neurons, microglia, oligodendrocytes, brain endothelium, and astrocytes, and performed functional enrichment analysis on each community (Figure S9–S13, Table S6 in supporting information and examples in Figure 3B). Pathways such as actin cytoskeleton, RNA metabolism and stress granules, and mitochondrial function featured across cell types, as well as cell type-specific pathways relevant to LOAD pathogenesis. In neurons, iDEAL genes may potentially regulate GABAergic and glutamatergic synapse, postsynaptic densities, synaptic plasticity, dendritic spine maintenance, cholesterol biosynthesis, and nitric oxide signaling, which plays a role in neurodegeneration in AD[74] (Figure S9 and Figure 3B). In microglia, as expected, we find involvement in inflammation, autophagy/lysosome, and regulation of microglial cell migration (Figure S10 and Figure 3B). Interestingly, other enriched pathways relate to the neuron–microglia interaction such as synapse pruning and dendritic spine maintenance (Figure 3B). In oligodendrocytes, a cluster of iDEAL genes (NFIA, TNR, NRXN2, and DGKZ) are potentially involved in focal adhesion kinase-mediated sprouting of injured axons (Figure S11 and Figure 3B). In endothelial cells, iDEAL genes potentially mediate extracellular signaling by hormones (oxytocin, insulin) and growth factors (transforming growth factor [TGF]-beta, epidermal growth factor receptor [EGFR]). Interestingly, this analysis also reveals an involvement of CCNT2 in amyloid clearance in the endothelium (Figure S12 and Figure 3B). In the astrocytes, the enriched pathways include glial cell differentiation and synaptic vesicles, as well as protein degradation pathways (autophagy or ubiquitin/proteasome) and the trafficking and processing of endosomal Toll-like receptors (TLRs), which regulate astrocytic neuroinflammation[75,76] (Figure S13 and Figure 3B). These results reveal that the iDEAL candidates are potentially involved in numerous dementia-related pathways and emphasize the value of using brain-specific and cell type–specific data. Moreover, this analysis provides insights into how variants in different genes may lead to convergent pathogenic or protective effects. For example, five iDEAL genes in two different cell types (CNTN1 and NCKAP in neurons and NRXN2, ABHD2, TIA1 in microglia) are involved in the same process, dendritic spine maintenance. Coexpression functional enrichment also reveals novel potential gene functions. For example, GSN is found in the synapse pruning coexpression community in microglia (Figure 3B and Figure S10). To our knowledge, GSN has not been associated to synapse pruning in mammalians. However, prompted by this result we found that the GSN C. elegans homolog mediates synapse pruning in nematodes,[77] suggesting a similar role in the AD context. Of note, in this analysis, APOE only appeared in three coexpression communities: (1) endolysosome and low-density lipoprotein (LDL) catabolism in microglia together with the iDEAL gene CTSB, (2) protein localization to endoplasmic reticulum in endothelium together with STT3B, and (3) insulin-mediated glucose transport, also in the endothelium as part of

a large community. This raises the possibility that a number of iDEAL genes do not work in close interaction with APOE, but they exert their effect in parallel pathogenic or neuroprotective roles.

In summary, there is an urgent focus in the field to account for the missing heritability in AD. A number of studies have succeeded in uncovering rare variants associated with AD risk.[67,78–81] These studies either rely on GWAS statistics to establish a link to AD, focus on familial inheritance followed by validation, target specific candidate genes, or use basic criteria for calling functional variants (ie, frameshift, stop gain, simple AA substitution). Although this study may be underpowered by classical genetics standards, iDEAL complements these approaches by adding a vast amount of evolutionary information in predicting the impact of mutations. This new information is then aggregated into a mutational burden to identify candidate genes, and rare variants can then be validated using more directed analysis. As the number of WES or whole genome sequencing datasets expand, we plan future AD studies to assess the impact of variants in the genes identified here on larger patient cohorts, as well as the role of non-coding variants and intergenic variations not included in this study. Success in this validation phase would mean that they could be used by clinicians to improve assessment of risk for AD patients. Future work to define the specific effect of each variant on protein function will facilitate deciphering the exact role they play in AD pathogenesis. Particularly promising are genes whose reduced function is protective in Drosophila and for which pharmacological compounds are available. These targets can be pursued in mouse models by CRISPR-based knock-in of the identified variants for functional assessment. Due to concern for potential lack of power, this study did not take sex effect into account. APOE ε4 risk is known to be greater in women,[82] and as sequencing studies become larger future studies will investigate the sex-specific roles of the candidate alleles we identified. Additionally, local ancestry may be a contributing factor on the effects of APOE in AD risk.[83] However, such information was absent in the ADSP dataset, and future studies would benefit from looking at ancestry-specific genetic variations that may modify the roles of APOE. In a similar vein, as this study was limited only to White samples, the candidate modifiers presented here might not all be translatable to other ethnic groups. Future studies that distinguish between White-specific modifiers and pan-ethnic modifiers, and identify novel alleles specific to other ethnic groups would add to our understanding of LOAD.

While the genes identified by iDEAL reinforce pathways known to participate in AD and are related to known APOE biology, they also highlight other unsuspected pathways like rRNA and ncRNA metabolism or chromatin binding and regulation. Furthermore, we identify potential cell type–specific functions of the iDEAL candidates, like synapse pruning in microglia or amyloid clearance in endothelial cells. The potential interplay of these pathways to APOE biology argues for additional characterization in the context of AD and potential therapeutic paths. Moreover, as APOE is implicated in neurodegeneration and inflammation beyond AD,[84] we may hypothesize that the alleles found in this study could play broader roles in neuropathology or neuroprotection. Further studies could assess whether some of the alleles presented here are specific to AD pathology. Genes and variants that

contribute to neuropathology or neuroprotection in general, especially those strong enough to neutralize *APOE* ε4, could point to therapeutic opportunities that cut across neurodegenerative diseases.

2 | DETAILED METHODS

2.1 | Data acquisition and quality assessment

Variant calls files (.vcf) and sample phenotype data produced by ADSP were downloaded from dbGaP (dbGaP accession: phs000572.v7.p4) in February 2018. The whole exome sequencing data encompassed 5686 individuals, of which only 5561 remained after excluding samples whose AD status had changed. AD status was determined based on clinical diagnosis. The number of samples and age at diagnosis for *APOE* ε2 and *APOE* ε4 patient groups are in Figure S1. Family-based data in the ADSP dataset was not used, and only unrelated non-Hispanic White individuals were analyzed to avoid potential confounding genetic background. While variants were jointly called using Atlas (Baylor) or Genome Analysis Toolkit (GATK; Broad), due to a known issue with variant calls from the GATK pipeline, only genotype calls from the Atlas calling pipeline were used. The resulting variants displayed high quality, with an average transition to transversion (TiTv) ratio of 3.52 ± 0.05 and lambda value[85] of 0.039 ± 0.001.

2.2 | Predicting variant impact using EA

All EA scores are available on the public server at: http://eaction.lichtargelab.org/. Synonymous variants were given a score of zero, while stop-gain or start_loss variants were given a max score of one hundred. When a variant affected multiple isoforms of a protein, the score was averaged across all affected isoforms.

2.3 | iDEAL

2.3.1 | Imputed deviation in EA load

First, we established the expected functional mutational burden for a given background mutation rate across the two patient groups by calculating the sum of EA scores (Y) and the number of all protein-coding variants (X) for each gene and then regressing Y on X. How much each gene deviated from expectation in a given patient group was then determined by its distance from the regression line: $d_{AD-\varepsilon2}$ for AD-ε2 group and $d_{HC-\varepsilon4}$ for HC-ε4 group. Next, to identify genes differentially mutated in the two patient groups, we performed a second regression on $d_{AD-\varepsilon2}$ and $d_{HC-\varepsilon4}$. For each gene, we measured the distance from this second regression line, which we refer to as iDEAL. Genes with positive iDEAL scores (above the regression line) have greater than expected EA load, or functional mutational burden, in the AD-ε2 group ($d_{AD-\varepsilon2}$) than in the HC-ε4 group ($d_{HC-\varepsilon4}$). Conversely, genes with negative iDEAL scores (below the regression line) have greater EA load in the HC-ε4 group ($d_{HC-\varepsilon4}$) than in the AD-ε2 group ($d_{AD-\varepsilon2}$). To

control for noise from random mutations, we assessed the significance of each gene's signal using a *z*-score measured against a background distribution of iDEAL scores built by randomizing the labels of AD-ε2 and HC-ε4 patients 100 times. This yielded 216 genes with absolute *z*-score values above 2.5 (>99th percentile).

2.4 | AD differential gene expression enrichment analysis

To integrate the iDEAL hits with transcriptionally dysregulated genes in AD brains, we used the RNAseq data available from the AMP-AD knowledge portal[37,39,86] (https://doi.org/10.7303/syn2580853; see Acknowledgments), which has been re-analyzed to normalize across the different studies as detailed in a previous study.[86] The definitions for AD patients and controls were identical to the ones defined in the aforementioned paper. Specifically, we used the following datasets: syn8484987, syn8466812, syn8456629 for the brain region–specific differential expression, using the data portion specifically comparing AD to control. For the AD meta-transcriptome, we used dataset syn11914606.[86] The brain samples are from ROSMAP (DLPFC 155 AD/86 Control), MayoRNAseq (TCX 80-AD/73-control), and Mount Siani Brain Bank (FP 167-AD/93-control, IFG 151-AD/79-control, PHG 143-AD/82-control, STG 151-AD/89-control). The *P*-values for differential expression were adjusted for multiple hypothesis testing using false discovery rate (FDR) estimation, and we selected the genes with an adjusted *P*-value below .01 (much more detailed methodological description can be found in Logsdon et al.[86]). Enrichment in iDEAL genes was calculated using hypergeometric test applied using as background gene number those genes for which there was data in both the AMP-AD and the ADSP datasets.

2.5 | Network analysis

2.5.1 | Protein-protein interaction (PPI)

PPI network was defined by the Homo sapiens STRING v11[87] using the combined score of all evidence types or of Textmining, Experiments, or Databases, and were considered only if they were above an interaction score threshold of 0.400, 0.450, or 0.700, depending on the analysis. For control, random sets of 216 genes were selected 1000 times to establish a background distribution of expected number of genes that would interact with the same GWAS AD genes the iDEAL genes interact with. From the background distributions, *z*-scores were calculated.

For Markov clustering and gene set enrichment analysis of iDEAL genes, we built a network with the 216 iDEAL genes using STRING v11[75] at a stringency level of 0.400. Next, we applied the Markov cluster algorithm (MCA) provided in the STRING interface, with an inflation value of 1.9, which yielded 26 clusters. Finally, using the "analysis" tool, we looked at the functional enrichment within each module on the various databases covered by STRING (FDR q-value <0.05).

2.5.2 | Coexpression

The single cell RNAseq, cell type–specific WGCNA networks were obtained from McKenzie et al.[73] Using Cytoscape, we identified the primary degree coexpressed nodes for the iDEAL genes and built coexpression communities using the HiDef-Louvain algorithm tool in the Community Detection extension. We obtained ≈100 communities for each cell type, and then ran the functional enrichment tool in the Community Detection extension to explore functional overlap in gProfiler, enrichR, and iQuery databases applying an FDR q < 0.05. We also integrated the results with the AD-specific Alzpathway manually curated database.[88,89] Pathways that appeared in more than one database were highlighted in Figures S9–S13.

2.6 | AdaBoost-SVM

Adaboost is an ensemble method that combines weak estimators into a single strong classifier. Here, the AdaBoost classifier used support vector machine (SVM) as the base estimator. For features, EA scores were averaged to represent a single score for each gene.

2.7 | *Drosophila* strains and motor performance assay

Genetics and strains: the *Drosophila* lines carrying *UAS*-Tau, and $UAS-Aos:\beta_{42}$ have been previously characterized[40,41] and are available from the Bloomington Drosophila Stock Center (BDSC, University of Indiana). For pan-neuronal expression we used the *elav-GAL4*(C155) driver from BDSC. The alleles tested as candidate modifiers targeting the *Drosophila* homologs of iDEAL genes were obtained from the BDSC and from the Vienna Drosophila Resource Center (VDRC). Homologs were identified using Blast and also the DRSC Integrative Ortholog Prediction Tool (Diopt score).[90] Genotypes used are summarized in Table S2.

For the motor performance tests, we used a highly automated behavioral assay based on the *Drosophila* startle-induced negative geotaxis response as previously described.[42] To assess motor performance of fruit flies as a function of age, we used 10 age-matched virgin females per replica per genotype. Flies are collected in a 24-hour period and transferred into a new vial containing $300\mu l$ of media every day. Four replicates were used per genotype. Using an automated platform, the animals are taped to the bottom of a plastic vial and recorded for 7.5 seconds. Videos are analyzed using custom software to assess the speed of each individual animal. Three trials per replicate are performed each day shown, and four replicates per genotype are used. A linear mixed effect model analysis of variance was run using each four replicates to establish statistical significance across genotypes.

3 | DETAILED RESULTS

3.1 | iDEAL modifiers have bias in high-impact variants in AD compared to HC with the same *APOE* allele

We reasoned that if genes with greater than expected mutational burden in the AD-ε2 versus HC-ε4 (ADε2-iDEAL genes) foster LOAD pathogenesis, they should also be enriched for high-impact variants in AD-ε2 compared to healthy controls without the risk allele (HC-ε2, n = 457). Likewise, if genes with greater than expected mutational burden in the HC-ε4 versus AD-ε2 (HCε4-iDEAL genes) keep ε4 carriers healthy, then they should also be depleted of high-impact variants in the AD-ε4 group (n = 1148) relative to HC-ε4. We tested these hypotheses by comparing the EA score distributions of the 148 ADε2-iDEAL genes in AD-ε2 versus HC-ε2 and the 68 HCε4-iDEAL genes in AD-ε4 versus HC-ε4. As predicted, ADε2-iDEAL genes were enriched for high-impact variants in the AD-ε2 group in comparison to the HC-ε2 group (Figure S2B, red, P = 1.0E-14; K-S test); likewise, the HCε4-iDEAL genes were depleted of high-impact variants in AD-ε4 individuals (Figure S2B, blue, P = 4.1E-06; K-S test). No such bias existed when the patients were randomized (Figure S3, one randomization, left panel; 100 randomization trials, right panel). These data support iDEAL genes and variants as plausible modifiers of LOAD phenotypes associated with *APOE* ε2 and ε4 alleles.

To evaluate whether these modifier effects were generalizable beyond *APOE* ε2 or ε4 allele status, we asked whether the same biases also existed in individuals homozygous for the most common *APOE* allele, *APOE* ε3. If so, *APOE* ε3 homozygous AD patients (AD-ε3) would show enrichment for high impact variants in ADε2-iDEAL genes relative to *APOE* ε3 homozygous healthy individuals (HC-ε3) and depletion of high impact variants in HCε4-iDEAL genes. Indeed, compared to 1657 HC-ε3 control individuals, the 1346 AD-ε3 patients were depleted of high impact variants in HCε4-iDEAL genes (Figure S2C, blue, P = .015; K-S test) and showed a trend for enrichment of high impact variants in ADε2-iDEAL genes, though it did not reach statistical significance (Figure S2C, red, P = .25; K-S test). Next, we reasoned that if the ADε2-iDEAL genes foster LOAD pathogenesis, they would be enriched for high-impact variants in AD patients compared to HC, regardless of *APOE* status. Indeed, ADε2-iDEAL genes were enriched for high-impact variants in AD patients compared to HC (Figure S2A, red, P = 1.1E-06; K-S test), and the HCε4-iDEAL genes were depleted of high-impact variants in AD patients (Figure S2A, blue, P = 2.6E-07; K-S test). These results indicate that the iDEAL gene variants are significantly linked to AD status and may be causative of neurodegeneration or provide protection. Taken together, these observations indicate that variants responsible for the paradoxical ADε2/HCε4 phenotypes may also be relevant for the entire AD population regardless of *APOE* genotype.

3.2 | iDEAL genes are significantly dysregulated in AD brains and show association with AD risk genes and pathways

To further investigate the relevance of iDEAL genes in AD biology beyond *APOE*, we assessed their expression in AD brains provided in several datasets of the AMP-AD sequence repository.[34–39] Out of the 174 iDEAL genes analyzed in the AMP-AD data, 75 genes were significantly up- or downregulated (adjusted *P*-value cutoff of .01) in AD patients compared to controls, defined following the criteria provided by Logsdon et al.,[86] in at least one brain region (Figure 1B). This statistically significant enrichment (*P* = .018; hypergeometric test) indicates that expression of many iDEAL genes either responds to or is causative of AD-related insults, further supporting that iDEAL genes may act as modifiers of *APOE* through a broader role in AD pathology.

Among these differentially expressed iDEAL genes is *TREM2*, a well-studied AD risk gene.[67,70–72] Building on this, we measured the degree of connectivity between the iDEAL genes and known AD susceptibility genes identified by GWAS[15–21] using STRING v11.[87] We found that iDEAL genes (n = 65) were significantly more connected to AD-GWAS candidates than expected by random chance (Figure 1C, Figure S4A, z-score = 2.24). This higher connectivity to AD susceptibility genes was retained regardless of the stringency applied in defining STRING network edges (Figures S4B and S4C). We also find that 25 iDEAL genes interact directly with genes involved in *APP* and tau biology (Figure S3D), further supporting a direct link between many iDEAL genes and AD biology. Moreover, seven iDEAL genes fall within ±500kb of LD regions for each locus found in a recent GWAS by Kunkle et al.[21] These data suggest that the iDEAL candidates share mechanisms of action with known AD risk factors and increases confidence in their potential connection to AD biology.

3.3 | iDEAL genes are enriched in modifiers of tau- and β_{42}-induced neuronal dysfunction in vivo

Next, we investigated the functional consequences of modulating the levels of the iDEAL genes in an in vivo model of AD. Since *APOE* exerts its effects at least partially through amyloid production and/or tau accumulation, which are prominent pathologies in AD,[49,50,91,92] we hypothesized that a number of the iDEAL genes would exert their effects by modulating the pathogenesis of the AD-driving proteins tau and amyloid beta (Aβ). To test this, we used two well-characterized *Drosophila* models that express either human wild-type 4R tau[40] or secreted β_{42} peptide pan-neuronally.[41] Expression of either protein in *Drosophila* neurons leads to progressive locomotor performance deficits. We used an automated data acquisition system that enables movement recording for individual animals to assess their speed (mm/s) while climbing a vial.[42] This system provides a quantitative measurement of locomotor performance that can be longitudinally assessed in aging animals and constitutes a precise, high-throughput functional assay of neuronal dysfunction. We determined the *Drosophila* homolog of each iDEAL gene[90] and tested all genes with

available overexpression and/or loss of function alleles (classical or shRNA). Of the 134 genes with homologs in flies, modulating the levels of 69 genes worsened or ameliorated tau- or β_{42}-induced neuronal dysfunction (51.5% hit rate, *P* = 1.13e-107; hypergeometric test) (Figure 1D, Figures S5–S8, and Table S2 in supporting information). Twenty genes had alleles with a protective effect in the tau model (18 when knocked down and 2 when overexpressed) while 22 ameliorated the β_{42}-induced dysfunction (20 when knocked down and 2 when over-expressed). Knockdown of the *Drosophila* homologs of 20 iDEAL genes worsened the tau-induced neuronal dysfunction while decreased levels of 19 iDEAL genes enhanced the β_{42}-induced phenotype. Among these genes, knockdown of *Drosophila* homologs of *ABHD2*, *CNTN1*, *DGKE*, *DPEP1*, *HSD17B2*, *LRRC17*, *PGP*, *RRP8*, and *THG1L* ameliorated the neuronal dysfunction in tau and β_{42} animals, while knockdown of *Drosophila* homologs of *ATP6V0E2*, *COX11*, *E2F8*, *IGFALS*, and *TBC1D4* exacerbated neuronal dysfunction in both models, indicating that these genes may play a role in common mechanisms underlying the pathogenesis of *APOE* in relation to both tau and β_{42} (Figure S14 in supporting information).

To assess which iDEAL genes may be good drug targets, we searched the DGIdb[55] for pharmacological compounds that interact with these genes and found that 39 genes interact with 390 compounds (Table S3). Of these, three genes (*ITGA2B*, *ALDH5A*, and *HDAC7*) interact with two medications (enoxaparin and valproic acid) that have been associated with lower incidence of AD in a population study.[56] In addition, we searched PubMed for publications that co-mention the term "Alzheimer" and any of the 390 compounds. Three additional drugs had literature evidence for having potential neuroprotective effects in animal models: the cathepsin inhibitor LHVS,[57] URMC-099,[58,59] and CX-4945.[60,61] LHVS is an inhibitor of *CTSB*, and URMC-099 and CX-4945 are inhibitors of *DAPK3* (Table S3). Interestingly, inhibition of the *Drosophila* homologs of these two iDEAL genes also results in neuroprotection (Figure 1D). Given their robust effect and druggability, we believe these genes are top candidates to characterize further in AD mouse models using these existing inhibitors as well as by genetic knockdown (either using viral delivered shRNAs or targeted CRISPR knock-out).

3.4 | Variants in iDEAL genes show strong potential to be used for AD risk prediction

Because these data suggest that iDEAL genes are linked to AD and neurodegeneration, we asked next whether they could also be used for AD risk prediction and patient stratification. First, we used machine learning to test whether iDEAL gene variants could separate between the two paradoxical patient groups. In a five-fold cross-validation, the AdaBoost-SVM algorithm[62,63] trained on the mutational features represented by EA across the 216 iDEALgenes (see Methods) could classify AD-ε2 versus HC-ε4 individuals with an average AUC of 0.92 (Figure 2A). To assess which of the 216 genes have the highest predictive power, we implemented permutation feature importance,[64] which pointed to 94 genes that contributed to risk prediction (Table S4).

Surprisingly, 41 iDEAL genes were better classifiers for AD risk in the paradoxical AD-ε2/HC-ε4 population than *TREM2* (Figure 2B). As expected, many of these top genes had variants with significant OR in the AD-ε2/HC-ε4 comparison (Table S4). Interestingly, among the top genes, there were also five (*ZNF804B, SYTL2, AKNAD1, LAMA2, LRRC17*-K119E) that showed variants with OR > 1 in the AD-ε3 population, and eight genes with variants with OR < 1 in the AD-ε3 individuals (*LTBP1, AKNAD1, LRRC17*-G187A, *PRKAG3, MIA2, WDR60, SMTNL1,* and *ARHGAP33*). This further underscores the possibility that iDEAL candidates may play a role in AD in the broader population. Next, we tested a more clinically relevant question, namely, whether the variants in the 216 iDEAL genes could predict which of the *APOE* ε2 carriers would develop AD, and conversely, which *APOE* ε4 carriers would remain healthy. This would provide a useful tool for geneticists and neurologists to stratify individuals with these genotypes and select those at highest risk, for example, ahead of clinical trials. We again used the AdaBoost-SVM algorithm in five-fold cross-validation and could separate the AD-ε2 patients from the HC-ε2 individuals with an average AUC of 0.79 (Figure 2C). Likewise, we could stratify HC-ε4 individuals from the AD-ε4 patients with an average AUC of 0.71 (Figure 2D). While further validation is required when more exomes become available, these data suggest that the 216 iDEAL genes have the potential to be used as stratification biomarkers on top of the usual *APOE*-genotype risk prediction for AD.

3.5 | iDEAL genes form clusters in the PPI and coexpression networks and may be involved in pathways related to synaptic integrity and cell type–specific pathways related to dementia

Finally, we investigated the biological pathways in which the iDEAL genes are involved. We constructed a STRING-based network with our genes and applied the Markov cluster algorithm to detect densely connected regions (inflation 1.9). There are likely multiple pathways toward dementia, but ultimately, synaptic dysfunction and substantial loss of neurons are major contributors of AD pathogenesis[65] and the most proximal event that leads to clinical features observed in AD. Indeed, of the 26 clusters iDEAL genes formed, 15 showed significant enrichment for biological pathways (Methods; Figure 3; Table S5), many of which are relevant for synaptic integrity. For example, pathways such as vesicular and protein traffic, neuron projection and axon guidance, dendritic spine, microtubule transport and related processes are involved in normal axonal transport of proteins and organelles such as mitochondria, which is essential for healthy synapse function.[66] Specifically, the neuron projection and axon guidance pathway was of particular interest because three of the genes—*TREM2, PLXNA4,* and *PAK2*—have already been associated with AD.[67–69] *TREM2* is a well-known risk gene that has been studied extensively.[67,70–72] *PLXNA4*, which ranked top (highest *z*-score) among the HCε4-iDEAL genes, has previously been shown to have a protective role,[68] and *PAK2* has been implicated in AD synaptic dysfunction.[69] Other pathways consistent with *APOE* functions (ie, plasma lipoprotein remodeling, lipid catabolism,

and vesicular traffic) were enriched in genes predictive of AD status in the paradoxical patient groups (*GPIHBP1, LPA, VNN2, SERPINE2, CPT1B, UCP2, GOLGA5,* and *PTCH1*). These data suggest that iDEAL genes are involved in pathways related to synaptic connection, which may be responsible for the paradoxical phenotypes we observe in AD-ε2 and HC-ε4 patients.

To gain more brain-specific pathway information on the iDEAL genes, we turned to coexpression networks followed by functional enrichment analysis. Using human single cell RNAseq weighed coexpression networks,[73] we extracted the first-degree neighbors of the iDEAL genes for each cell type (neurons, microglia, oligodendrocytes, brain endothelium, and astrocytes). Next, we constructed the iDEAL gene coexpression communities for each cell type using the HiDef-Louvain algorithm. Finally, the communities were functionally annotated using gProfiler, enrichR, and iQuery (Figure S9–S13, Table S6 and selected examples in Figure 3B). Pathways such as actin cytoskeleton, microtubule, RNA metabolism and stress granules, and mitochondrial function still featured across cell types. Interestingly, however, this more rigorous analysis also revealed the potential involvement of iDEAL genes in cell-specific pathways relevant to LOAD pathogenesis. In neurons, iDEAL genes may potentially regulate GABA metabolism and the GABAergic synapse, glutamatergic synapse, PSD 95 postsynaptic densities, and synaptic plasticity (Figure S9). We also find genes potentially involved in nitric oxide signaling, which may play a role in neurodegeneration in AD,[74] cholesterol biosynthesis, and dendritic spine maintenance (Figure 3B). In microglia, as would be predicted, iDEAL genes fall in pathways involved in inflammation, autophagy/lysosome and regulation of microglial cell migration (Figure S10 and Figure 3B). Strikingly, we also find an enrichment in genes potentially involved in synapse pruning and dendritic spine maintenance (Figure 3B). In oligodendrocytes, a cluster of iDEAL genes (*NFIA, TNR, NRXN2,* and *DGKZ*) are potentially involved in focal adhesion kinase-mediated sprouting of injured axon (Figure S11 and Figure 3B). In the brain endothelial cells, iDEAL genes may mediate aspects of extracellular signaling by hormones (oxytocin, insulin) and growth factors (TGF-beta, EGFR). Interestingly, this analysis also reveals a potential involvement of *CCNT2* in amyloid clearance in the endothelium (Figure S12 and Figure 3B). In the astrocytes, we find functional enrichment of iDEAL gene modules consistent with astrocyte functions like glial cell differentiation or synaptic vesicles. We also find several communities enriched in protein degradation functions by autophagy or ubiquitin/proteasome and the trafficking and processing of endosomal TLRs which could be involved in the neuroinflammatory response of astrocytes[75,76] (Figure S13 and Figure 3B). These results reveal that the iDEAL candidates are potentially involved in numerous dementia-related pathways and emphasize the importance of using brain-specific and cell type–specific data. Moreover, this analysis provides insight on how variants in different genes can lead to convergent pathogenic or protective effects. For example, five iDEAL genes in two different cell types (*CNTN1* and *NCKAP* in neurons and *NRXN2, ABHD2, TIA1* in microglia) are involved in the same process, dendritic spine maintenance. This approach also provides a means to infer the potential function of genes that have not

been well characterized in a specific cell type. For example, *GSN* falls within the synapse pruning coexpression community in microglial cells. To our knowledge, *GSN* has not been associated to synapse pruning in mammalians. However, prompted by the functional analysis result we found that its *C. elegans* homolog is a mediator of synapse pruning in nematodes,[77] suggesting a potentially similar role in the context of AD. Of note, in this analysis, APOE only appeared in three coexpression communities: (1) endolysosome and LDL catabolism in microglial cells together with the iDEAL gene *CTSB*, (2) protein localization to endoplasmic reticulum in endothelial cells together with *STT3B*, and (3) insulin-mediated glucose transport, also in the brain endothelium as part of a large cluster. This raises the possibility that most iDEAL genes are not directly working with APOE, but that they rather exert their pathogenic or protective effect up/downstream of APOE-mediated processes, or in parallel pathogenic or neuroprotective roles.

3.6 | Identification of specific variants in iDEAL genes associated with increased or decreased AD risk

We next investigated whether specific coding variants of iDEAL genes were associated with AD risk or protection. We calculated ORs for iDEAL gene variants in AD-*ε*2 patients versus HC-*ε*4. We found 62 variants in 54 iDEAL genes significantly associated with increased AD risk (OR > 1) and 31 variants in 21 iDEAL genes significantly associated with decreased AD risk (OR < 1, see Table S7 in supporting information). The EA scores of these variants are in line with SIFT[93] and PolyPhen2[94] scores. Among these 31 potentially protective variants, 28 were present even in *ε*4 homozygotes, suggesting a stronger effect. We also repeated these analyses in *ε*3 homozygotes and found variants in 19 iDEAL genes with OR > 1 and variants in 16 iDEAL genes with OR < 1 for AD (Table S8 in supporting information). In both paradoxical *APOE* groups and in homozygote *ε*3 carriers, variants in the protein products of *TREM2* (R47H), *OPRD1* (C27F), *ZAR1* (Q42H), and *GAMT* (T209M) were associated with increased AD risk while variants *PELO* (L221M), *TRAF3IP2* (D10N), *SMTNL1* (R345G), *LRRC17* (G187A), *UGT3A1* (C67G; C121G), and *NOP56* (M475T) were associated with decreased AD risk. These data suggest that the iDEAL method may be useful in prioritizing novel biomarkers for AD risk or protection, some of which are specific to the *APOE* allelic background while others are general.

In summary, we identify many new genes implicated in AD using diverse tests (Figure S15 in supporting information for the most robust iDEAL genes and in Table S9 in supporting information for all iDEAL genes). Together, these genes may serve to expand the ability of clinicians to assess the risk of patients for developing AD and to target novel candidates for mechanistic and therapeutic studies. More broadly, while our analyses focus on identifying genes with differential mutational EA load in AD, the universality of evolutionary information makes this a generalizable approach complementary to GWAS and applicable to other conditions with a strong risk phenotype.

CONFLICTS OF INTEREST

The authors declare no competing interests.

REFERENCES

1. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* 2010;9(1):119-128.
2. Attems J, Jellinger KA. The overlap between vascular disease and Alzheimer's disease–lessons from pathology. *BMC Med.*. 2014;12:206.
3. Rabinovici GD, Carrillo MC, Forman M, et al. Multiple comorbid neuropathologies in the setting of Alzheimer's disease neuropathology and implications for drug development. *Alzheimers Dement (N Y)*. 2017;3(1):83-91.
4. Negash S, A. Bennett D, S. Wilson R, A. Schneider J, E. Arnold S. Cognition and neuropathology in aging: multidimensional perspectives from the rush religious orders study and rush memory and aging project. *Curr Alzheimer Res*. 2011;8(4):336-340.

5. Negash S, Wilson R, Leurgans S, et al. Resilient brain aging: characterization of discordance between Alzheimer's disease pathology and cognition. *Curr Alzheimer Res.* 2013;10(8):844-851.

6. Negash S, Xie S, Davatzikos C, et al. Cognitive and functional resilience despite molecular evidence of Alzheimer's disease pathology. *Alzheimers Dement.* 2013;9(3):e89-e95.

7. Andrade-Moraes CH, Oliveira-Pinto AV, Castro-Fonseca E, et al. Cell number changes in Alzheimer's disease relate to dementia, not to plaques and tangles. *Brain.* 2013;136(Pt 12):3738-3752.

8. Arendt T. Synaptic degeneration in Alzheimer's disease. *Acta Neuropathol.* 2009;118(1):167-179.

9. Coleman P, Federoff H, Kurlan R. A focus on the synapse for neuroprotection in Alzheimer disease and other dementias. *Neurology.* 2004;63(7):1155-1162.

10. Dekosky ST, Scheff SW. Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Ann. Neurol..* 1990;27(5):457-464.

11. Giannakopoulos P, Gold G, von Gunten A, Hof PR, and Bouras C. Pathological substrates of cognitive decline in Alzheimer's disease. *Front Neurol Neurosci.* 2009;24:20-29.

12. Terry RD, Masliah E, Salmon DP, et al. Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann. Neurol.* 1991;30(4):572-580.

13. Gatz M, Pedersen NL, Berg S, et al. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol Series A Biol Sci Med Sci.* 1997;52A(2):M117-M125.

14. Ridge PG, Hoyt KB, Boehme K, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol. Aging.* 2016;41:200.e13 e13-200 e20.

15. Lambert J-C, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45(12):1452-1458.

16. Lambert J-C, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet.* 2009;41(10):1094-1099.

17. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet.* 2011;43(5):436-441.

18. Seshadri S. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA.* 2010;303(18):1832-40.

19. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet.* 2009;41(10):1088-1093.

20. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet.* 2011;43(5):429-435.

21. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet.* 2019;51(3):414-430.

22. Ridge PG, Mukherjee S, Crane PK, Kauwe JSK. Alzheimer's Disease Genetics, "Alzheimer's disease: analyzing the missing heritability". *PLoS One.* 2013;8(11):e79771.

23. Génin E. Missing heritability of complex diseases: case solved?. *Hum Genet.* 2020;139(1):103-113.

24. Zhao Na, Ren Y, Yamazaki Yu, et al. Alzheimer's risk factors age, APOE genotype, and sex drive distinct molecular pathways. *Neuron.* 2020;106(5):727-742.e6 e6.

25. Liu C-C, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat. Rev. Neurol..* 2013;9(2):106-118.

26. Farrer LA Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer disease meta analysis consortium. *JAMA.* 1997;278(16):1349-56.

27. Belloy ME, Napolioni V, Greicius MD. A Quarter Century of APOE and Alzheimer's disease: progress to date and the path forward. *Neuron.* 2019;101(5):820-838.

28. Katsonis P, Lichtarge O. A formal perturbation equation between genotype and phenotype determines the evolutionary action of protein-coding variations on fitness. *Genome Res.* 2014;24(12):2050-2058.

29. Katsonis P, Lichtarge O. CAGI5: Objective performance assessments of predictions based on the Evolutionary Action equation. *Hum. Mutat.* 2019;40(9):1436-1454.

30. Neskey DM, Osman AA, Ow TJ, et al. Evolutionary action score of TP53 identifies high-risk mutations associated with decreased survival and increased distant metastases in head and neck cancer. *Cancer Res.* 2015;75(7):1527-1536.

31. Osman AA, Neskey DM, Katsonis P, et al. Evolutionary action score of TP53 coding variants is predictive of platinum response in head and neck cancer patients. *Cancer Res.* 2015;75(7):1205-1215.

32. Osman AA, Monroe MM, Ortega Alves MV, et al. Wee-1 kinase inhibition overcomes cisplatin resistance associated with high-risk TP53 mutations in head and neck cancer through mitotic arrest followed by senescence. *Mol. Cancer Ther.* 2015;14(2):608-619.

33. Chun YS, Passot G, Yamashita S, et al. Deleterious effect of RAS and evolutionary high-risk TP53 double mutation in colorectal liver metastases. *Ann Surg.* 2019;269(5):917-923.

34. A. Bennett D, A. Schneider J, S. Buchman A, L. Barnes L, A. Boyle P, S. Wilson R. Overview and findings from the rush Memory and Aging Project. *Curr Alzheimer Res.* 2012;9(6):646-663.

35. A. Bennett D, A. Schneider J, Arvanitakis Z, S. Wilson R. Overview and findings from the religious orders study. *Curr Alzheimer Res.* 2012;9(6):628-645.

36. Wang M, Beckmann ND, Roussos P, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data.* 2018;5:180185.

37. Allen M, Carrasquillo MM, Funk C, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data.* 2016;3:160089.

38. De Jager PL, Ma Y, Mccabe C, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data.* 2018;5:180142.

39. Mostafavi S, Gaiteri C, Sullivan SE, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci.* 2018;21(6):811-819.

40. Lasagna-Reeves CA, De Haro M, Hao S, et al. Reduction of nuak1 decreases tau and reverses phenotypes in a tauopathy mouse model. *Neuron.* 2016;92(2):407-418.

41. Chouhan AK, Guo C, Hsieh Yi-C, et al. Uncoupling neuronal death and dysfunction in Drosophila models of neurodegenerative disease. *Acta Neuropathol Commun.* 2016;4(1):62.

42. Al-Ramahi I, Giridharan SSP, Chen Yu-C, et al. Inhibition of PIP4Kgamma ameliorates the pathological effects of mutant huntingtin protein. *Elife.* 2017;6.

43. Rousseaux MWc, De Haro M, Lasagna-Reeves CA, et al. TRIM28 regulates the nuclear accumulation and toxicity of both alpha-synuclein and tau. *Elife.* 2016;5.

44. Farfel JM, Yu L, De Jager PL, Schneider JA, Bennett DA. Association of APOE with tau-tangle pathology with and without β-amyloid (in eng). *Neurobiol Aging.* 2016;37:19-25.

45. Safieh M, Korczyn AD, Michaelson DM. ApoE4: an emerging therapeutic target for Alzheimer's disease. *BMC Med..* 2019;17(1):64.

46. Zhou M, Huang T, Collins N, et al. APOE4 Induces Site-Specific Tau Phosphorylation Through Calpain-CDK5 Signaling Pathway in EFAD-Tg Mice. *Curr Alzheimer Res.* 2016;13(9):1048-1055.

47. Wadhwani AR, Affaneh A, Van Gulden S, Kessler JA. Neuronal apolipoprotein E4 increases cell death and phosphorylated tau release in alzheimer disease. *Ann Neurol.* 2019;85(5): 726-739.

48. Liraz O, Boehm-Cagan A, Michaelson DM. ApoE4 induces Aβ42, tau, and neuronal pathology in the hippocampus of young targeted replacement apoE4 mice. *Mol Neurodegener*. 2013;8:16.

49. Wang C, Najm R, Xu Q, et al., Gain of toxic apolipoprotein E4 effects in human iPSC-derived neurons is ameliorated by a small-molecule structure corrector. *Nat Med*. 2018;24(5):647-657.

50. Lin Y-Ta, Seo J, Gao F, et al. APOE4 causes widespread molecular and cellular alterations associated with alzheimer's disease phenotypes in human iPSC-derived brain cell types. *Neuron*. 2018;98(6):1141-1154.e7 e7.

51. Huang Yu-WA, Zhou Bo, Wernig M, Südhof TC. ApoE2, ApoE3, and ApoE4 differentially stimulate APP transcription and Aβ secretion *Cell*. 2017;168(3):427-441.e21.e21.

52. Myhre O, Utkilen H, Duale N, Brunborg G, Hofer T. Metal dyshomeostasis and inflammation in Alzheimer's and Parkinson's diseases: possible impact of environmental exposures. *Oxid Med Cell Longev*. 2013;2013:1.

53. Dubal DB, Yokoyama JS. Longevity gene KLOTHO and Alzheimer disease-a better fate for individuals who carry APOE ε4. *JAMA Neurol*. 2020;77:798.

54. Arboleda-Velasquez JF, Lopera F, O'hare M, et al. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. *Nat. Med.*. 2019;25(11)1680-1683.

55. Cotto KC, Wagner AH, Feng Y-Y, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic. Acids. Res.* 2018;46(D1):D1068-D1073.

56. Kern DM, Cepeda MS, Lovestone S, Seabrook GR. Aiding the discovery of new treatments for dementia by uncovering unknown benefits of existing medications. *Alzheimers Dement (N Y)*. 2019;5:862-870.

57. Xu J, Wang H, Ding Ke, et al. Inhibition of cathepsin S produces neuroprotective effects after traumatic brain injury in mice. *Mediators Inflamm*. 2013;2013:1.

58. Kiyota T, Machhi J, Lu Y, et al. URMC-099 facilitates amyloid-beta clearance in a murine model of Alzheimer's disease. *J Neuroinflamm*. 2018;15(1):137.

59. Bos PH, Lowry ER, Costa J, et al. Development of MAP4 kinase inhibitors as motor neuron-protecting agents. *Cell Chem Biol*. 2019;26(12):1703-1715.e37 e37.

60. Rosenberger AFN, Morrema THJ, Gerritsen WH, et al. Increased occurrence of protein kinase CK2 in astrocytes in Alzheimer's disease pathology. *J Neuroinflammation*. 2016;13:4.

61. Kim H, Lee KS, Kim AK, et al. A chemical with proven clinical safety rescues Down-syndrome-related phenotypes in through DYRK1A inhibition. *Dis Model Mech*. 2016;9(8):839-848.

62. Zhu J, Zou H, Rosset S, and Hastie T. "Multi-class AdaBoost," Statistics and Its Interface. 2009;2:349-360.

63. Fan R, Chang K, Hsieh C, Wang X, and Lin C. LIBLINEAR: a library for large linear classification. *J Machine Learning Res*. 2008;9:1871-1874.

64. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikitlearn: Machine Learning in Python. *JMLR*. 2011;12:2825-2830.

65. Kashyap G, Bapat D, Das D, et al. Synapse loss and progress of Alzheimer's disease -a network model. *Sci Rep*. 2019;9(1):6555.

66. Mandelkow E. Clogging of axons by tau, inhibition of axonal traffic and starvation of synapses. *Neurobiol Aging*. 2003;24(8);1079-1085.

67. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med*. 2013;368(2):117-127.

68. Kang SS, Kurti A, Wojtas A, et al. Identification of plexin A4 as a novel clusterin receptor links two Alzheimer's disease risk genes. *Hum Mol Genet*. 2016;25(16):3467-3475.

69. Ma Q-L, Yang F, Calon F, et al. p21-activated kinase-aberrant activation and translocation in Alzheimer disease pathogenesis. *J Biol Chem*. 2008;283(20):14132-14143.

70. Ulland TK, Colonna M. TREM2 - a key player in microglial biology and Alzheimer disease. *Nat Rev Neurol*. 2018;14(11):667-675.

71. Jin SC, Benitez BA, Karch CM, et al. Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum Mol Genet*. 2014;23(21):5838-5846.

72. Sirkis DW, Bonham LW, Aparicio RE, et al. Rare TREM2 variants associated with Alzheimer's disease display reduced cell surface expression. *Acta Neuropathol Commun*. 2016;4(1):98.

73. Mckenzie AT, Wang M, Hauberg ME, et al. Brain cell type specific gene expression and co-expression network architectures. *Sci Rep*. 2018;8(1):8868.

74. Balez R, Ooi L. Getting to NO Alzheimer's disease: neuroprotection versus neurotoxicity mediated by nitric oxide. *Oxid Med Cell Longev*. 2016;2016:1.

75. Madeddu S, Woods TA, Mukherjee P, Sturdevant D, Butchi NB, Peterson KE. Identification of glial activation markers by comparison of transcriptome changes between astrocytes and microglia following innate immune stimulation. *PLoS One*. 2015;10(7): e0127336.

76. Sochocka M, Diniz BS, Leszek J. Inflammatory response in the CNS: friend or foe?. *Mol. Neurobiol.*. 2017;54(10):8071-8089.

77. Meng L, Mulcahy B, Cook SJ, et al. The cell death pathway regulates synapse elimination through cleavage of gelsolin in caenorhabditis elegans neurons. *Cell Rep*. 2015;11(11):1737-1748.

78. Wetzel-Smith MK, Hunkapiller J, Bhangale TR, et al. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. *Nat Med*. 2014;20(12):1452-1457.

79. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl J Med*. 2013;368(2):107-116.

80. Logue MW, Schu M, Vardarajan BN, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. *Alzheimers Dement*. 2014;10(6):609-618.e11 e11.

81. Cruchaga C, Karch CM, Jin SC, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature*. 2014;505(7484):550-554.

82. Ungar L, Altmann A, Greicius MD. Apolipoprotein E, gender, and Alzheimer's disease: an overlooked, but potent and promising interaction. *Brain Imaging Behav*. 2014;8(2):262-273.

83. Blue EE, Horimoto ARVR, Mukherjee S, Wijsman EM, Thornton TA. Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. *Alzheimers Dement*. 2019;15(12):1524-1532.

84. Tzioras M, Davies C, Newman A, Jackson R, Spires Jones T. Invited review: APOE at the interface of inflammation, neurodegeneration and pathological protein spread in Alzheimer's disease. *Neuropathol Appl Neurobiol*. 2019;45(4):327-346.

85. Koire A, Katsonis P, and Lichtarge O, Repurposing germline exomes of the cancer genome atlas demands a cautious approach and sample-specific variant filtering. *Pac Symp Biocomput*. 2016;21:207-18.

86. Logsdon BA, et al. Meta-analysis of the human brain transcriptome identifies heterogeneity across human AD coexpression modules robust to sample collection and methodological approach. *bioRxiv*. 2019:510420.

87. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613.

88. Mizuno S, Iijima R, Ogishima S, et al. AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. *BMC Systems Biol*. 2012;6:52.

89. Mizuno S, Ogishima S, Kitatani K, et al. Network analysis of a comprehensive knowledge repository reveals a dual role for ceramide in Alzheimer's disease. *PLoS One*. 2016;11(2):e0148431.

90. Hu Y, Flockhart I, Vinayagam A, et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinform*. 2011;12:357.

91. Liu L, Mackenzie KR, Putluri N, Maletić-Savatić M, Bellen HJ, The glia-neuron lactate shuttle and elevated ROS promote lipid synthesis in neurons and lipid droplet accumulation in glia via APOE/D. *Cell Metab.* 2017;26(5):719-737.e6 e6.

92. Huang YWA, Zhou Bo, Wernig M, Südhof TC, ApoE2, ApoE3, and ApoE4 differentially stimulate APP transcription and abeta secretion. *Cell.* 2017;168(3):427-441 e21.

93. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins (in eng). *Nucleic Acids Res.* 2012;40(Web Server issue):W452-W457.

94. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.