Polygenic resilience scores capture protective genetic effects for Alzheimer’s disease

Jiahui Hou1,2,3, Jonathan L. Hess1,2,3, Nicola Armstrong4, Joshua C. Bis5, Benjamin Grenier-Boley6, Ida K. Karlsson7,8, Ganna Leonenko9, Katya Numbers10, Eleanor K. O’Brien11,12, Alexey Shadrin13, Anbupalam Thalamuthu10, Qiong Yang14, Ole A. Andreassen13, Henry Brodaty15,10, Margaret Gatz7,15, Nicole A. Kochan10, Jean-Charles Lambert6,9, Simon M. Laws16, Colin L. Masters16, Karen A. Mather16,17, Nancy L. Pedersen7, Daniëlle Posthuma18, Permindar S. Sachdev10, Julie Williams19, the Alzheimer’s Disease Neuroimaging Initiative, Chun Chieh Fan20, Stephen V. Farace21, Christine Fennema-Notestine21, Shu-Ju Lin22, Valentina Escott-Price19,19, Peter Holmans19,19, Sudha Seshadri23, Ming T. Tsuang16,22, William S. Kremen22 and Stephen J. Glatt1,2,3,24,25

Polygenic risk scores (PRSs) can boost risk prediction in late-onset Alzheimer’s disease (LOAD) beyond apolipoprotein E (APOE) but have not been leveraged to identify genetic resilience factors. Here, we sought to identify resilience-conferring common genetic variants in (1) unaffected individuals having high PRS for LOAD, and (2) unaffected APOE-e4 carriers also having high PRSs for LOAD. We used genome-wide association study (GWAS) to contrast “resilient” unaffected individuals at the highest genetic risk for LOAD with LOAD cases at comparable risk. From GWAS results, we constructed polygenic resilience scores to aggregate the additive contributions of risk-orthogonal common variants that promote resilience to LOAD. Replication of resilience scores was undertaken in eight independent studies. We successfully replicated two polygenic resilience scores that reduce genetic risk penetrance for LOAD. We also showed that polygenic resilience scores positively correlate with polygenic risk scores in unaffected individuals, perhaps aiding in staving off disease. Our findings align with the hypothesis that a combination of risk-independent common variants mediates resilience to LOAD by moderating genetic disease risk.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2022

INTRODUCTION

Alzheimer’s disease (AD) is the leading cause of dementia [1]. AD exists as two genetically distinct forms: early-onset AD, which is caused by autosomal dominant mutations in one of several genes (PSEN1, PSEN2, APP, SORL1) and typically has an onset of symptoms between the ages of 40 and 60 years [2], and the more common late-onset AD (LOAD), which is sporadic, polygenic, and typically has an onset of symptoms in the mid-60s [3]. Elevated risk of LOAD is associated with a host of lifestyle factors and medical conditions, such as a high-fat diet, heavy drinking and smoking, cardiovascular disease, type-2 diabetes, and traumatic brain injury [4]. More importantly, the heritability of LOAD from twin studies was estimated at 58–79% [5], and its estimates from single-nucleotide polymorphisms (SNPs) range...
from 13 to 33% [6–9]. The goal of this study is to determine whether genes also play a role in resilience to LOAD. We used an innovative approach first introduced and applied in schizophrenia as a general framework for resilience research [10], focusing on individuals at the highest levels of genetic risk.

To date, genome-wide association studies (GWAS) have discovered close to 50 genome-wide significant loci (P < 5e-08) associated with LOAD risk [9, 11–20]. The e4 allele of apolipoprotein E (APOE) is the polymorphism with the strongest effect on LOAD susceptibility [21]. Beyond APOE-e4, there may be thousands of additional genetic polymorphisms that make small individual contributions to the overall risk for LOAD [22–25]. A polygenic risk score (PRS) [26] can be derived by summing the weighted effect of SNPs to identify a single genetic risk variable that reflects one’s relative susceptibility to LOAD. Recent LOAD PRSs capture most of the SNP heritability for LOAD [9, 24, 27]. Extensive research shows that PRSs boost the accuracy of LOAD diagnosis beyond the performance of APOE [22–25], and capture LOAD phenotypic variability not explained by APOE status [28, 29].

Revealing the genetic architecture of LOAD is vital for understanding its etiology and identifying molecular targets for innovative therapeutic interventions. Yet, knowledge of risk factors might be fruitfully complemented by an understanding of resilience-associated and -promoting mechanisms as well. As such, some AD research has shifted focus from symptomatic cases to healthy aging individuals or asymptomatic individuals at elevated risk [30]. This was motivated by the premise that high-risk asymptomatic individuals, yet unaffected, may provide clues that protect them against AD. Here, we employ the term “resilience” to indicate individuals who show better than expected outcomes in the face of high genetic risk for disease [30–35].

Increasing evidence suggests that several factors—including education, literacy, physical activity, and mental activity—can moderate the risk for LOAD [31, 32, 36], and it is estimated that one-third to 40% of dementia cases might be preventable [36, 37]. These moderation effects may be explained by reverse causation [38], but genetic influences—which are not subject to reverse causation—also underlie these factors. Educational attainment [39, 40] and, particularly, general cognitive ability [40, 41] are heritable. Thus, some of these factors may also confer resilience-enhancing genetic effects. Notably, some genetic variants, such as APOE-ε2 [42] and the APP A673T variant [43], have been identified as protective for LOAD. However, the biological mechanisms that drive these protective effects remain largely unknown. Importantly, we consider such protective effects to be fundamentally different from the “resilience” effects we sought in our study, in that protective factors are generally operative across the full range of risk, whereas resilience factors are only operative in those at the highest risk for disease. Very little work has been aimed at identifying additional genetic resilience factors that potentially moderate the genetic risk established by the cumulative effects of risk-associated alleles and their corresponding protective alleles. Genetic resilience against risk for LOAD has been investigated for identifying additional genetic resilience factors that potentially moderate the genetic risk established by the cumulative effects of risk-associated alleles [or their alternative protective alleles]. We hypothesized that there exist resilience-associated variants that lower LOAD susceptibility in a manner that is statistically independent of the effects of risk-associated alleles (or their alternative protective alleles). We tested this hypothesis by capitalizing on the most comprehensive known PRS for LOAD [18] and APOE allelic status to develop two designs identifying unaffected individuals with the highest genetic likelihood of developing LOAD. Design 1 defined “resilient” individuals as normal controls with the highest PRSs for LOAD. Design 2 defined “resilient” individuals as normal controls with at least one APOE-ε4 allele and the highest LOAD PRSs (excluding the APOE region). We aimed to discover residual common genetic variants that confer resilience to unaffected individuals in the highest genetic risk tiers for LOAD. We then leveraged this profile of resilience-promoting genetic variants to build a polygenic resilience score for LOAD. We hypothesized that polygenic resilience scores would account for significant variation in affection status for LOAD among individuals with high genetic risk, and would show a significant positive correlation with PRSs in unaffected controls.

**METHODS**

**Research design**

Our workflow is shown in Fig. 1. In stage 1, a recent GWAS meta-analysis for LOAD [18] was leveraged for identifying risk variants and polygenic risk scoring. In stage 2, we compared two analytic designs to identify high-risk “resilient” normal controls and “risk-matched” LOAD cases. In stage 3, a resilience GWAS was conducted for each design using the identified high-risk individuals. Then the polygenic resilience score weights were derived from resilience GWAS meta-analysis summary statistics. Finally, polygenic resilience scores were replicated in independent external studies for evaluating the performance in distinguishing high-risk “resilient” normal controls from “risk-matched” LOAD cases. The parameters of each analysis step are summarized in Supplemental Table 3.

**Samples and genotypes**

We acquired the largest available collection of genome-wide SNP data for clinically diagnosed or autopsy-confirmed LOAD to ensure adequate power. Table 1 shows the number of normal controls, LOAD cases, high-risk “resilient” normal controls, and “risk-matched” LOAD cases in each study. Summary statistics of age-at-onset (AAO) for LOAD cases and age-at-first-exam (AAE) for normal controls are presented in Supplemental Table 1. In design 1 and design 2, the mean AAE of high-risk “resilient” normal controls and the mean AAO of “risk-matched” LOAD cases ranged from 70.3 to 80.9, and there were no significant age differences between groups. A common lower bound for AAO of LOAD is 65; however, the age cutoff has no specific biological significance [3], and many genetic studies of LOAD have included cases with AAO as low as 60 (and the same AAE for unaffected comparison subjects). Therefore, we included all participants in this analysis having an AAO of 60 years old. The full name and accessibility of each study can be found in Supplemental Table 2. All 26 studies in the discovery stage came from the stage-1 AD GWAS meta-analysis of Kunkle et al. [18]. The eight studies in the replication stage are fully independent of the discovery studies. Full descriptions of the discovery and replication samples were published previously [9, 17, 18, 51, 52]. Genotypes for all studies were imputed using the Haplotype Reference Consortium (HRC) r1.1 2016 reference panel [53]. Detailed quality control (QC) steps for samples and genotypes are described in Supplemental Methods.

**Identifying individuals at high genetic risk**

In design 1, a PRS was used to select individuals with high genetic risk. At the time of deploying our analyses, the Kunkle et al., 2019 study [18] was
the largest publicly available GWAS using clinically diagnosed or autopsy-confirmed AD cases and CN controls, as opposed to proxy AD cases and controls that might lead to inaccurate risk estimation [27]. Therefore, we consider that this study will give the most accurate measure of AD risk and derived the PRS weights from its stage-1 AD GWAS meta-analysis summary statistics [18]. See Supplementary Methods for further details. The variance in AD explained by PRS maximizes at a $P$-value threshold of 0.5 in participants from GERAD (Genetic and Environmental Risk for Alzheimer’s disease) [23] and 22 locally available ADGC (Alzheimer’s Disease Genetics Consortium) studies (Supplementary Fig. 4). We, therefore, adopted this threshold to ensure that our risk measure captures as much of the genetic risk for AD as possible. This very conservative threshold will thus ensure that potential risk SNPs with even very small effect sizes will not be advanced for consideration as resilience SNPs. However, if studies other than Kunkle et al. 2019 were used to estimate genetic risk for AD, it may be the case that smaller $P$-value thresholds may be optimal (e.g., $5 \times 10^{-8}$, $1 \times 10^{-7}$, 0.1) [9, 22, 54, 55]. Within each study, LOAD cases and normal controls were ranked based on their PRSs. Note that as the true prevalence of resilience to AD in the population is unknown, we adopted the same high-risk percentile cutoff that proved effective in our original work [10], and classified the 10% of controls with the highest PRSs as “resilient.” The LOAD cases whose risk scores were between the 90th percentile and the maximum PRS in controls were retained as risk-matched LOAD cases for comparison.

In design 2, within the subset of normal controls who had at least one apolipoprotein E ($\text{APOE}$-ε4) allele, a threshold of $\geq$80th percentile of PRSs (excluding SNPs in the $\text{APOE}$ region) was used to define high-risk controls as “resilient.” Stage 3: Resilience GWAS and replication of polygenic resilience scores. GWAS was performed using “resilient” individuals and risk-matched affected cases from each of the two designs. For each design, polygenic resilience scores were derived and evaluated in external replication datasets. LD linkage disequilibrium, OR odds ratio, SNPs single-nucleotide polymorphisms.
| Study       | Sub-study     | Design 1 | LOAD cases | Normal controls | Design 2 | LOAD cases | Normal controls |
|-------------|---------------|----------|------------|----------------|----------|------------|----------------|
|             |               |          |            |                |          |            |                |
|             |               | N        | Total      | % Retained     | N        | Total      | % Retained     |
|             |               | High-risk |            |                | Risk-matched |            |                |
| Discovery   | ADGC          | 52       | 512        | 10.2           | 842      | 1524       | 55.2           |
|             |               | 16       | 155        | 10.3           | 386      | 620        | 62.3           |
|             |               | 57       | 566        | 10.1           | 379      | 666        | 56.9           |
|             |               | 38       | 376        | 10.1           | 222      | 304        | 73.0           |
|             |               | 51       | 503        | 10.1           | 187      | 285        | 65.6           |
|             |               | 34       | 337        | 10.1           | 68       | 213        | 31.9           |
|             | MAYO          | 112      | 1117       | 10.0           | 626      | 754        | 83.0           |
|             | UVMU5SM       | 113      | 1108       | 10.2           | 753      | 1123       | 67.1           |
|             | ACT + GenDiff | 157      | 1556       | 10.1           | 425      | 524        | 81.1           |
|             | MIRAGE        | 51       | 509        | 10.0           | 101      | 137        | 73.7           |
|             | WASHU1        | 18       | 176        | 10.2           | 137      | 253        | 54.2           |
|             | ROSMAP        | 72       | 717        | 10.0           | 192      | 237        | 81.0           |
|             | UPIT/T        | 82       | 811        | 10.1           | 794      | 1152       | 68.9           |
|             | OHUS          | 18       | 180        | 10.0           | 119      | 174        | 68.4           |
|             | Tgen II       | 36       | 360        | 10.0           | 364      | 613        | 59.4           |
|             | NIA-LOAD      | 102      | 1007       | 10.1           | 582      | 760        | 76.6           |
|             | ROSSMAP2      | 22       | 214        | 10.3           | 50       | 59         | 84.7           |
|             | WASHU2        | 8        | 71         | 11.3           | 29       | 36         | 80.6           |
|             | MTC           | 19       | 188        | 10.1           | 220      | 252        | 87.3           |
|             | TARRC         | 18       | 176        | 10.2           | 241      | 306        | 78.8           |
|             | WHICAP        | 56       | 553        | 10.1           | 17       | 72         | 23.6           |
|             | ADNI          | 13       | 127        | 10.2           | 174      | 350        | 49.7           |
|             | CHARGE        | 168      | 1661       | 10.1           | 274      | 451        | 60.8           |
|             | FHS           | 191      | 1901       | 10.0           | 137      | 288        | 47.6           |
|             | EADI          | 620      | 6172       | 10.0           | 1636     | 2167       | 75.5           |
|             | GERAD         | 139      | 1388       | 10.0           | 2354     | 2992       | 78.7           |
|             |               | 2263     | 22441      | 10.1           | 11309    | 16312      | 69.3           |
| Replication | AddNeuroMed   | 19       | 183        | 10.4           | 38       | 223        | 17.0           |
|             | ADGC          | 79       | 784        | 10.1           | 45       | 513        | 8.8            |
|             | ADNI          | 38       | 374        | 10.2           | 37       | 211        | 17.5           |
|             | PGC-ALZ       | 75       | 741        | 10.1           | 163      | 1085       | 15.0           |
|             | GENDER/ SATSA/ HARMONY | 77 | 764 | 10.1 | 39 | 306 | 12.7 |
|             | TwinGene      | 608      | 6080       | 10.0           | 30       | 287        | 10.5           |

Table 1. The number of LOAD cases and normal controls, high-risk normal controls (“resilient” individuals), and risk-matched LOAD cases identified in each of the discovery and replication studies.
Derivation, replication, and statistical analysis of polygenic resilience scores

GWASs of resilience were performed using logistic regression with Plink (version 1.9) [56]. Selected principal components, AAO/AAE, and sex were used as covariates. A GWAS meta-analysis was conducted in METAL [57] software using an inverse-variance random-effects model with genomic control. In accord with the pipeline described by Hess et al. [10], SNPs known to be associated with LOAD risk were excluded from the resilience-scoring algorithm; these were defined as those SNPs that showed an association with AD risk ($P < 0.5$) from the GWAS meta-analysis summary statistics [18], and variants that were in linkage disequilibrium (LD) ($r^2 \geq 0.2$ in a 1-Mb window) with those risk variants with associations of $P < 0.5$. This pruning step of excluding risk variants from consideration as resilience loci serves as a conservative measure to avoid re-discovering risk variants for resilience scoring. For both resilience designs, the polygenic resilience score weights were generated from the marginal SNPs of resilience GWAS meta-analysis summary statistics following the same series of QC steps (see Supplementary Methods).

Polygenic resilience scores were derived for 10 $P$-value thresholds, in a manner similar to the PRS algorithm, by summing up the weighted effective allele counts of SNPs [26]. Logistic regression was used to assess the likelihood of “resilient” group inclusion based on harboring a higher polygenic resilience score. Selected principal components, AAO/AAE and sex were used as covariates. For each polygenic resilience score, we meta-analyzed the natural logarithm of the odds ratio (OR) of being a high-risk resilient versus a risk-matched LOAD case using a random-effects inverse-variance model using the $R$ package metafor, and pooled variance explained in resilience across independent replication studies. All tests were two-tailed unless specified otherwise. See Supplementary Methods for further details.

RESULTS
Resilience GWAS
Design 1 produced 2263 high-risk “resilient” normal controls and 11,309 risk-matched LOAD cases for the resilience GWAS meta-analysis. As expected, the sample size retained in design 2 was smaller, totalling 988 high-risk “resilient” normal controls and 6541 risk-matched LOAD cases (Table 1). Because our analytic approaches used only subsets of all available LOAD case-control GWAS data, we neither had nor anticipated having sufficient power to detect individual SNPs with genome-wide significant association with resilience (Supplementary Fig. 3). Instead, our focus was on deriving and evaluating polygenic resilience scores. As a necessary step to generate SNP-weights for summation in those scores, we performed individual-SNP association tests and briefly reported the results in Supplementary Results.

Replication and evaluation of polygenic resilience scores
After removing risk-associated SNPs ($P < 0.5$) and SNPs in LD with those risk-associated SNPs ($r^2 \geq 0.2$), clumping the remaining marginal SNPs, and applying QC steps, a profile of 18,723 SNPs was included in the resilience score for design 1, and 18,122 SNPs in design 2. Resilience scores for all 10 $P$-value thresholds were significantly associated with “resilient” group inclusion (“resilient” normal controls versus risk-matched LOAD cases) when tested in locally downloaded discovery datasets. Results of the association between “resilient” group inclusion and polygenic resilience scores from the replication datasets were meta-analyzed, yielding 1056 high-risk “resilient” normal controls and 381 risk-matched LOAD cases in design 1, and 583 high-risk “resilient” normal controls and 331 risk-matched LOAD cases in design 2 (Table 1).

In design 1, the meta-analysis found significant replication of the association between “resilient” group inclusion and polygenic resilience scores at two $P$-value thresholds ($P < 0.1$, $P < 0.2$) (Fig. 2A). The most significant association was found for the polygenic resilience score containing all independent marginal SNPs with resilience GWAS $P < 0.1$ (OR $= 1.24$, 95% confidence interval $[CI] = 1.05–1.47$, $P = 0.010$). Resilience scores for the 0.1 $P$-value threshold explained an average of 1.3% (standard deviation $[SD] = 5.3\%$) of the variance in “resilient” group inclusion or 1.2%
(SD = 4.3%) (Fig. 2B) of the variance on the liability scale, i.e., SNP heritability of resilience. No significant (P < 0.05) replication of the association between “resilient” group inclusion and polygenic resilience scores was observed for any of the 10 polygenic resilience scores in design 2.

Note that the association between “resilient” group inclusion and polygenic resilience scores (P < 0.1, P < 0.2 of design 1) was not significant after multiple-testing correction for 10 P-value thresholds using the false discovery rate (FDR) or Bonferroni method. However, considering polygenic resilience scores were derived by aggregating SNPs within a series of escalating P-value thresholds, they were nested models and not independent. Therefore, a typical FDR or Bonferroni correction under the assumption of independence would be overly conservative.

**Interaction of risk and resilience effects**

In the full samples from three locally downloaded replication studies (Alzheimer Disease Centers Wave 7 [ADC7], AddNeuroMed, and Alzheimer’s Disease Neuroimaging Initiative stage GO/2/3 [ADNI-GO/2/3]; normal controls, n = 1,056; risk-matched LOAD cases, n = 381), we tested for correlations between PRSs and polygenic resilience scores. As hypothesized,
Risk-countering effects of polygenic resilience scores

Individuals with higher polygenic resilience scores \((P < 0.1\) and \(P < 0.2\) thresholds of design 1) had higher odds of being a “resilient” high-risk normal control than a risk-matched LOAD case. Polygenic resilience scores (design 1) significantly increased with higher PRSs in normal controls, but not in LOAD cases. Taken together, these results support the hypothesis that polygenic resilience scores capture risk-countering polygenic effects against the penetrance of high polygenic risk for LOAD, and that normal controls with higher PRSs are protected from LOAD by harboring correspondingly higher polygenic resilience scores. Although no polygenic resilience scores in design 2 demonstrated significant risk-buffering effects, we cannot rule out the possibility that common variants might reduce risk penetrance in normal controls with enriched risk from both APOE and PRSs. In fact, among APOE-e4 carriers, higher resilience scores in design 1 at the \(P < 0.1\) threshold (\(OR = 1.64, 95\% CI = 1.08–2.50, P = 0.021\)) and the \(P < 0.2\) threshold (\(OR = 1.98, 95\% CI = 1.24–3.15, P = 3.9e-03\)) were associated with higher odds of being a “resilient” high-PRS normal control than a risk-matched LOAD case. Among “resilient” high-PRS controls, higher resilience scores in design 1 (\(P < 0.2\) threshold) were significantly associated with increased odds of carrying at least one APOE-e4 allele (\(OR = 1.57, 95\% CI = 1.07–2.29, P = 0.021\)). A similar trend was observed when the \(P < 0.1\) threshold was used, although this was not significant (\(OR = 1.30, 95\% CI = 0.90–1.89, P = 0.16\)) (Supplementary Results). We, therefore, conclude that polygenic resilience scores may moderate the risk effects of the LOAD PRS generally, and the APOE-e4 allele specifically. However, these analyses were carried out in relatively small studies (ADC7, AddNeuroMed, and ADNI-GO/2/3), and need to be repeated in larger, more powerful, replication samples.

Interplay of polygenic effects and APOE

In design 2, we hypothesized that a two-stage selection of individuals (with both higher PRSs and one or more APOE-e4 alleles) would enrich for individuals with the absolute highest genetic risk for LOAD \((59, 60)\); yet, there was a substantial reduction in the performance of design 2 in contrast to design 1. The lack of significant replication of association with resilience in design 2 simply might be due to lower statistical power in both the resilience score development and replication stages, considering the total sample size of design 2 is approximately half that of design 1. Alternatively, resilience-promoting variants may be found among APOE-e4 carriers through broader exploration of the model-parameter space \((e.g.,\ PRS\ threshold\ in\ particular)\), separate evaluation of APOE-e4 homozygotes and various heterozygote combinations, and more accurate modeling of the genetic architecture of resilience \((see\ limitations\ below)\). An important question future studies should address is to what extent common variants may influence the penetrance of genetic risk in larger samples of APOE-e4 carriers, or whether the prevalence of risk-modifying common variants differs between APOE-e4 carriers and noncarriers.

On the other hand, multiple studies \([22–25, 28, 54, 61–63]\) have revealed that PRSs capture independent risk effects beyond APOE alone, while few studies have explored the risk-predictive performance of PRSs stratified by APOE status. Higher PRSs were found to be associated with increased susceptibility for LOAD in APOE-e4 noncarriers \([25, 29, 59]\). Furthermore, the risk effects of PRS deciles across APOE status could be dependent on the ages of participants \([29, 59, 62, 64]\). Further mining of the complex relationship between the risk effects of PRSs and APOE is outside the scope of the current study; however, further investigations on the penetrance of high PRSs among APOE-e4 carriers and noncarriers seem warranted.

Strengths and limitations

Our approach has identified candidate resilience loci that may ultimately serve as targets for the promotion of resilience. We
examined the performance of two polygenic resilience scores: design 1 selected participants with the highest polygenic risk regardless of APOE-ε4 status, while design 2 restricted analyses to APOE-ε4 carriers. To our knowledge, this is the first study to identify a polygenic resilience score for genetic LOAD risk, comprising thousands of risk-independent common variants that partially offset the genetic risk conferred by a relatively high PRS. An important distinction of the current study relative to prior work on genetic resilience to LOAD is that we accounted not only for the risk from APOE but also the aggregate effect of thousands of additional risk variants throughout the genome via the LOAD PRS.

A conservative variant-filtering strategy was applied, which resulted in the removal of common variants associated with LOAD risk variants (risk association $P < 0.5$) and those in liberal LD ($r^2 > 0.2$) with LOAD risk variants. A strength of this approach is that we ensured the polygenic resilience scores derived in the current study are independent of the risk scores so that the SNPs comprising the polygenic resilience score are not sub-threshold risk SNPs. Our design of defining “resilient” groups from the same risk background instead of contrasting high-risk normal controls with low-risk LOAD cases also helped avoid re-discovering variants merely associated with risk. In addition, the resilience alleles of these risk-residual SNPs are not simply protective alleles defined in a risk framework, where each biallelic locus is defined by both a risk allele and a corresponding and opposing protective allele. Thus, this strategy helps identify resilience effects that are conditioned on net risk effects, owing to the combination of risk and protective alleles summed in polygenic risk scores. Yet, although our approach is conservative, it is limited in the identification of a better-performing resilience score because most of the genome has been discarded from the analysis. Biologically, it is plausible that variants nearby risk loci, such as those in the same LD block or in the same gene with risk SNPs, could exert modifying functions [65]. Our conservative strategy, discarding all SNPs with any semblance of risk association, and those in liberal LD threshold with such SNPs, consequently leads to lower power in uncovering variants with potentially higher biological functionality. This notion is borne out in the fact that no significant gene-ontology pathways were enriched by resilience-related common variants identified in this study (results not shown). With larger samples, resilience-conferring SNPs may be investigated using a stricter LD threshold (e.g., $r^2 > 0.1$) to further restrain the “hitchhiking” of risk variants. More importantly, Mendelian randomization, conditional association testing, or simulation analysis may be better suited to evaluate the hypothesis that resilience signals are more likely to co-localize with risk loci or genes. In addition, filtering variants by LD with risk SNPs results in a low LD structure among the remaining SNPs as demonstrated previously [10], which diminishes our capacity to examine the genetic correlation of resilience to LOAD with other risk- or resilience-related phenotypes (e.g., via LD score regression). A high priority should be placed on the design of new methods that can detect resilience-associated SNPs that may reside in regions of strong LD with risk variants.

Resilience was defined by discrete groups in our analysis, which truncated effective sample sizes to the upper tail of the risk distribution. Choosing a lower percentile cutoff would increase the sample size available for the resilience analysis, while potentially diminishing the signal of resilience genes. In the future, when larger samples are available, higher risk thresholds may be applied and subgroups at more extreme risk could be leveraged to increase power. It is an important task for future studies to investigate which of these factors (i.e., sample size, signal: noise) would have the greater effect on power, and to better understand the prevalence of resilience to AD in the population. Theoretically, resilience may be a continuous measure; thus, our resilience approach might also be improved by leveraging all study samples and modeling the continuity of resilience using either linear or non-linear analysis. Despite the restricted sample sizes in the current study, two resilience scores in design 1 were sufficiently robust to replicate significantly in fully independent studies. Further replication would be key to testing the validity of these resilience scores. It is expected that the strength of our results (in terms of variance explained and the significance of associations) will only increase with the addition of more samples.

Several studies [9, 22, 54, 55] indicated that polygenic risk scores of $P$-value thresholds less than 0.5 (i.e., $5 \times 10^{-5}$, $1 \times 10^{-5}$ 0.1) might show better performance in predicting LOAD risk. Therefore, it may be valuable to compare the performance of resilience scores developed from risk scores at other $p$-value thresholds. In addition, it is likely that a subgroup of “resilient” normal controls identified in this work will eventually develop LOAD, but with later onset. Thus, all resilient participants demonstrate resilience against high levels of genetic risk for LOAD, but only those who never develop LOAD are additionally resistant against the disease itself. Lastly, the participants in our analysis were of European ancestry, so the degree of generalization of our results to non-European populations is presently unknown.

Future directions

Two analysis designs were deployed in the current study to select individuals with a high genetic risk burden from both PRSs and APOE, and other methods could be devised to expand the capabilities of our resilience approach in LOAD. It has been suggested, for example, that using a PRS with the APOE region removed and adding APOE alleles as a covariate may boost the performance of LOAD risk prediction [66], compared with incorporating APOE alleles as weighted SNPs in PRSs. In addition, it could be important to include the number of APOE-ε4 or ε2 alleles as covariates in resilience analysis models to better reflect the relative risk levels among individuals. In our study, we consider it important to utilize the most comprehensive risk profile of LOAD to identify resilient individuals, i.e., normal controls with the highest genetic risk from all sources. Future studies may be interested in examining the resilience effects that moderate a portion of the LOAD risk. For example, resilience to the risk effects of APOE-ε4 alone can be studied by defining all APOE-ε4 carrying (or homozygotic) normal controls as resilient. To examine whether the resilience scores remain predictive if the APOE region is excluded from the PRS, the resilience to residual polygenic risk effects excluding APOE can be investigated in normal controls without APOE-ε4, who land in the top percentiles of PRSs (excluding the APOE region). Previous studies [42, 67] demonstrated that women carrying APOE-ε4 alleles were at greater risk of developing AD than men with the same APOE-ε4 dosages, especially between the ages of 65 and 75. When larger sample sizes are available, limiting our analysis to females in design 2 may further enrich for high-risk individuals and increase resilience signals.

Potentially, polygenic resilience scores from the current study could be applied to other resilience-related questions. For example, it would be instrumental in discovering the extent to which polygenic resilience score is associated with other phenotypes that have been associated with resilience to LOAD risk (e.g., education, general cognitive ability in early life, and other indices of cognitive reserve, brain reserve, or brain maintenance) [31, 32, 68–72]. In follow-up studies, it might be illuminating to investigate whether these resilience-promoting genetic factors show protective effects for cognitive impairment or LOAD-related pathophysiological changes.

Conclusio

We found evidence to support the hypothesis that thousands of risk-independent common variants underlie resilience among unaffected individuals with higher genetic risk for LOAD. We conclude that common variants not in LD with known LOAD risk
variants exert a protective effect on LOAD risk. Our findings provide a significant and novel contribution to the existing understanding of genetic resilience to LOAD risk. This novel approach highlights a window of opportunity for identifying risk-modifying biological mechanisms and potential pathways for intervention in populations at the highest risk for LOAD.

DATA AVAILABILITY
The Alzheimer’s Disease Genetics Consortium, the Alzheimer’s Disease Neuroimaging Initiative, and the AddNeuroMed data used in this study were provided under restricted access by NIAGADS (https://www.niagads.org), ADNI (http://adni.loni.usc.edu), and Synapse platform (https://www.synapse.org/#!/Synapse syn4907804), respectively. Only summary statistics were made available to us from the European Alzheimer’s Disease Initiative, the Genetic and Environmental Risk in Alzheimer’s Disease, the Cohorts for Heart and Aging Research in Genomic Epidemiology, the Psychiatric Genomic Consortium, the Australian Imaging, Biomarker & Lifestyle Study, and the Sydney Memory and Ageing Study. The resilience-scoring weights of design 1 are available from the corresponding author upon request.

CODE AVAILABILITY
The code for data analysis and, and the resilience-scoring weights of design 1 are available from the corresponding author upon request.

REFERENCES
1. Alzheimer’s Association. Alzheimer’s disease facts and figures. Alzheimers Dement. 2021;17:372–406.
2. Ryan NS, Nicholas JM, Weston PSJ, Liang Y, Lashley T, Guerreiro R, et al. Clinical phenotype and genetic associations in autosomal dominant familial Alzheimer’s disease: a case series. Lancet Neurol. 2016;15:1326–35.
3. Rossor MN, Fox NC, Mummery CJ, Schott JM, Warren JD. The diagnosis of young-onset dementia. Lancet Neurol. 2010;9:793–806.
4. Edwards ii GA, Gamez N, Escobedo G Jr, Calderon O, Moreno-Gonzalez I. Modifiable risk factors for Alzheimer’s disease. Front Aging Neurosci. 2019;11:146.
5. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry. 2006;63:168–74.
6. Lee SH, Harold D, Nyholt DR, Consortium AN, International Endogene C, Genetic. et al. Estimation and partitioning of polygenic variance captured by common SNPs for Alzheimer’s disease, multiple sclerosis and endometriosis. Hum Mol Genet. 2013;22:832–41.
7. Ridge PG, Mukherjee S, Crane PK, Kauwe JS. Alzheimer’s disease genetics C. Alzheimer’s disease: analyzing the missing heritability. PLoS ONE. 2013;8:e79771.
8. Brainstorm C, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer’s disease. The Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020;396:201–8.
9. Zhang Q, Sidorenko J, Couvy-Duchesne B, Marioni RE, Wright MJ, Goate AM, et al. A polygenic resilience score moderates the genetic risk for Alzheimer’s disease pathology in late onset families. Science. 1993;261:921–3.
10. Altmann A, Scelsi MA, Shoai M, de Silva E, Aksman LM, Cash DM, et al. A comprehensive analysis of methods for assessing polygenic burden on Alzheimer’s disease pathology and risk beyond APOE. Brain. 2020;2:fcz047.
11. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genotype-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat Genet. 2019;51:404–13.
12. Kunkle BW, Grenier-Bolye B, Sims R, Bis JC, Damotte V, Naj AC, 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;51:414–30.
13. Brainstorm C, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease. Nat Genet. 2021;53:1276–82.
14. Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer’s disease risk genes. Nat Genet. 2021;53:392–402.
15. Corruble E, Saunders AM, Strittmatter WJ, Schmeichel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of the Alzheimer’s disease in late onset families. Science. 1993;261:921–3.
16. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet. 2015;47:702–9.
17. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun. 2018;9:2098.

Translational Psychiatry (2022)12:296
57. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association studies. Mol Genet Genomic. 2010;283:203–13.

58. Clarke TK, Fox HC, Patterson N, Cardon LR, Rivas-Munoz A, Spurlock BB, et al. De novo missense mutations in pick-up genes associated with increased risk of Alzheimer disease. Hum Mol Genet. 2019;28:1333–46.

59. DeStefano AL, Fulker DW, DeStefano MS, Gigerenzer G, Keene CD, Martin A, et al. A 22-single nucleotide polymorphism Alzheimer disease risk score identifies a new Alzheimer disease susceptibility locus. Nat Genet. 2014;46:524–7.

60. Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, et al. A polygenic risk score based on genome-wide association scans identifies a new Alzheimer disease susceptibility locus. Nat Neurosci. 2014;17:1401–9.

61. Marioni RE, Campbell A, Hagenaars SP, Nagy R, Amador C, Hayward C, et al. A multivariate analysis of cognition in two-stage screening. Stat Med. 2015;34:3281–307.

62. Bellou E, Baker E, Leonenko G, Scherzer-Smith M, Daunt P, Menzies G, et al. Age-dependent effect of APOE and polygenic component on Alzheimer's disease. Neurobiol Aging. 2020;93:69–77.

63. Stocker H, Berner K, Perusse L, Rieder MJ, Thorleifsson G, Mieszkowski P, et al. Prediction of clinical diagnosis of Alzheimer's disease, vascular disease, and all-cause dementia by a polygenic risk score and APOE status in a community-based cohort prospectively followed over 17 years. Mol Psychiatry. 2022;26:5812–22.

64. Fulton-Howard B, Goate AM, Adelson RP, Koppel J, Gordon ML, Alzheimer Disease Genetics C, et al. Greater effect of polygenic risk score for Alzheimer disease polygenic scores in the Health and Retirement Study: a longitudinal study. Transl Psychiatry. 2022;12:296.

65. Concorde to the first draft of the manuscript and APOE provided critical edits. JLH, BGB, IKA, EKO, AS, AT, and QY performed the statistical analyses in non-local datasets and provided summary statistics. JLH, JLV, TEP, PH, SS, and WSK were involved in the acquisition of data. JH performed the analyses of this study. JCB, BGB, GL, EKO, AS, AT, and QY performed the statistical analyses in non-local datasets and provided summary statistics. JLH, JLV, TEP, PH, SS, and WSK were involved in the interpretation of data. JH wrote the first draft of the manuscript and APOE provided critical edits. JH, BGB, IKA, EKO, AS, OA, MG, SM, KAM, PSS, JWF, CN, TEP, PH, and WSK provided the critical review of the manuscript. All authors gave final approval of the version of the manuscript submitted for peer review.

66. Neu SC, Pa J, Kukull W, Buekly D, Kuzma A, Gangadharam P, et al. Apolipoprotein E genotype and sex risk factors for Alzheimer disease: a meta-analysis. JAMA Neurol. 2017;74:1178–89.

67. Mukherjee S, Kim S, Ribbens LE, Nho K, Risacher SL, Glymour MM, et al. Genetic architecture of resilience of executive functioning. Brain Imaging Behav. 2012;6:621–33.

68. Mukherjee S, Kim S, Ramanan VK, Ribbens LE, Nho K, Glymour MM, et al. Gene-based GWAS and biological pathway analysis of the resilience of executive functioning. Brain Imaging Behav. 2014:8:110–8.

69. Folsky D, Xu J, Chibnik LB, Schneider JA, Knight J, Kennedy JL, et al. Genetic epistasis regulates amyloid deposition in resilient aging. Alzheimers Dement. 2017:13:1107–16.

70. Ridge PG, Karch CM, Hsu S, Arano I, Teerlink CC, Ebbert MTW, et al. Linkage, whole genome sequence, and biological data implicate variants in RAB10 in Alzheimer’s disease resilience. Genome Med. 2017;9:100.

ACKNOWLEDGEMENTS

We thank all the participants of this study for their contributions. Our effort on this project was supported by the U.S. National Institute on Aging [grant numbers R01AG064955 and R01AG054002]. Additional acknowledgments and detailed acknowledgments of funding sources for the study are provided in Supplementary acknowledgments.

AUTHOR CONTRIBUTIONS

JH, JLH, CCF, SVF, CFN, SJP, VEP, PH, SS, MTT, WSK, and JG5 contributed to the conception and design of the study. The Alzheimer’s Disease Neuroimaging Initiative (ADNI), NA, JCB, BGB, IKA, GL, KN, EKO, AS, AT, QY, OAA, HB, MG, NAK, JCL, SLM, CLM, KAM, NLP, DP, PSS, JWF, PH, SS, and WSK were involved in the acquisition of data. JH performed the analyses of this study. JCB, BGB, GL, EKO, AS, AT, and QY performed the statistical analyses in non-local datasets and provided summary statistics. JLH, JLV, TEP, PH, SS, and WSK were involved in the interpretation of data. JH wrote the first draft of the manuscript and APOE provided critical edits. JH, BGB, IKA, EKO, AS, OA, MG, SM, KAM, PSS, JWF, SVF, CNJ, TEP, PH, and WSK provided the critical review of the manuscript. All authors gave final approval of the version of the manuscript submitted for peer review.

COMPETING INTERESTS

OA is a consultant to HealthLytics. SVF in the past year, received income, potential income, travel expenses continuing education support and/or research support from Aardvark, Akili, Genomind, Ironshore, KemPharm/Corium, Noven, Ondosis, Otsuka, Rhodes, Supernus, Takeda, Tris, and Vallon. With his institution, he has US patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of Alzheimer's Disease. In previous years, he received support from: Alcoba, Arbor, Aveksam, CoqCubed, Eli Lilly, Enzymotec, Impact, Janssen, Lundbeck/Takeda, Merck, NeuracelSciences, Novartis, Pfizer, Shire, and Sunovion. He also receives royalties from books published by Guilford Press: Straight Talk about Your Child's Mental Health; Oxford University Press: Schizophrenia: The Facts; and Elsevier: ADHD: Non-Pharmacologic Interventions. He is also Program Director of www.adhdinadults.com. He is supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 965381; NIMH grants U01AR076092-01A1, 1R21MH1264940, R01MH116037; Oregon Health and Science University, Otsuka Pharmaceuticals, Noven Pharmaceuticals Incorporated, and Supernus Pharmaceutical Company. The remaining authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-022-02055-0.

Correspondence and requests for materials should be addressed to Stephen J. Glatt.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
