An integrated genome and phenome-wide association study approach to understanding Alzheimer's disease predisposition

Archita S. Khaire¹, Courtney E. Wimberly¹, Eleanor C. Semmes^{2,3}, Jillian H. Hurst PhD³, Kyle M. Walsh^{1,3,4*}

¹ Division of Neuro-epidemiology, Department of Neurosurgery, Duke University, Durham, NC, USA

² Medical Scientist Training Program, Duke University, Durham, NC, USA

³ Children's Health and Discovery Initiative, Department of Pediatrics, Duke University, Durham, NC, USA

⁴ Center for the Study of Aging and Human Development, Duke University, Durham, NC, USA

Running title: Leveraging GWAS and PheWAS in Alzheimer's disease

Keywords: Alzheimer's disease, GWAS, PheWAS, EGFR, genetic epidemiology

***Corresponding author**

Email: kyle.walsh@duke.edu

Mailing address: DUMC Box 3050, Durham, NC 27710

Phone: (919) 684-8732

Fax: (919) 684-5207

Conflict of interest disclosure: The authors declare no competing interests.

Word count: 232 (abstract), 3205 (manuscript excluding abstract and references), 3 tables, 1 figure

Financial support: This study was supported by the Neuroscience Initiative of Translating Duke Health (CEW, KMW), the Infectious Diseases Society of America Foundation's "Microbial Pathogenesis in Alzheimer's Disease" grant (KMW, CEW, ECS), and the National Institute on Aging of the National Institutes of Health under Award Number P30AG072958 (ASK, KMW). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ABSTRACT

Background: Genome-wide association studies (GWAS) have identified common, heritable alleles that increase late-onset Alzheimer’s disease (LOAD) risk. We recently published an analytic approach to integrate GWAS and phenome-wide association study (PheWAS) data, enabling identification of candidate traits and trait-associated variants impacting disease risk, and apply it here to LOAD. **Methods:** PheWAS was performed for 23 known LOAD-associated single nucleotide polymorphisms (SNPs) and 4:1 matched control SNPs using UK Biobank data. Traits enriched for association with LOAD SNPs were ascertained and used to identify trait-associated candidate SNPs to be tested for association with LOAD risk (17,008 cases; 37,154 controls). **Results:** LOAD-associated SNPs were significantly enriched for associations with 6/778 queried traits, including three platelet traits. The strongest enrichment was for platelet distribution width (PDW) ($P=1.2 \times 10^{-5}$), but no consistent direction of effect was observed between increased PDW and LOAD susceptibility across variants or in Mendelian randomization analysis. Of 384 PDW-associated SNPs identified by prior GWAS, 36 were nominally associated with LOAD risk and 5 survived false-discovery rate correction for multiple testing. Associations confirmed known LOAD risk loci near *PICALM*, *CD2AP*, *SP11*, and *NDUFAF6*, and identified a novel risk locus in the epidermal growth factor receptor (*EGFR*) gene. **Conclusions:** Through integration of GWAS and PheWAS data, we identify substantial pleiotropy between genetic determinants of LOAD and of platelet morphology, and for the first time implicate EGFR – a mediator of β -amyloid toxicity – in Alzheimer’s disease susceptibility.

INTRODUCTION

Late-onset Alzheimer's disease (LOAD) is defined by the development of symptoms after age 65 and is the most common form of dementia. LOAD affects an estimated 5.7 million Americans, a number projected to grow rapidly with demographic trends.(Rabinovici 2019) Genome-wide association studies (GWAS) have greatly enhanced our understanding of the etiology of LOAD.(Bruni, Bernardi, and Gabelli 2020; Lewcock et al. 2020) A recent GWAS meta-analysis identified common heritable risk alleles at 24 statistically independent genetic loci, implicating genes involved in immunity, lipid metabolism, tau and amyloid- β (A β) in LOAD pathogenesis.(Kunkle et al. 2019) Although collaborative GWAS meta-analyses including >90k LOAD patients have enabled identification of low-frequency risk alleles, gains in statistical power are increasingly limited. Furthermore, functional genomics studies to reveal the biologic consequences of associated single nucleotide polymorphisms (SNPs) through *in silico* and *in vivo* analyses remain limited.(Andrews, Fulton-Howard, and Goate 2020) The use of polygenic score and Mendelian randomization (MR) approaches,(Shi et al. 2021; Zhang, Wang, and Liu 2021; Li et al. 2021; Handy et al. 2021) gene-pathway analyses,(Gao et al. 2018) and phenome-wide association studies (PheWAS)(Lumsden et al. 2020) can augment traditional GWAS approaches and expand our understanding of the underlying biology behind LOAD pathogenesis.

PheWAS are an approach to analyze a collection of phenotypes for their association with a particular genetic variant of interest, and they have recently been made feasible through the integration of electronic health record (EHR) data with large-scale genomic datasets.(Denny, Bastarache, and Roden 2016) A PheWAS essentially inverts the classic GWAS approach, testing a large collection of phenotypes (rather than a single disease or trait) for association with a single genetic variant (rather than millions of variants distributed across the genome). However, both GWAS and PheWAS typically test SNP-trait associations through the same multivariable regression-based frameworks, including adjustment for subject ancestry to control for population stratification.(Bush, Oetjens, and Crawford 2016)

By expanding the list of phenotypes associated with a variant of interest, performing PheWAS on a known disease-associated SNP can identify potential co-morbid conditions,(Byun et al. 2021) as well as traits that may mediate the SNP-disease association.(Pettit et al. 2021) For example, a PheWAS of the lung cancer risk-variant rs16969968 reveals numerous strong associations with various smoking-related behaviors (*e.g.*, cigarettes per day) that mediate the variant’s association with lung cancer risk.(Millard et al. 2019) Identifying such “intermediate traits” connecting LOAD risk alleles to causal mechanisms of LOAD pathogenesis has clear relevance for disease prevention and population-based risk reduction.(Semmes, Zhang, and Walsh 2018) Furthermore, by identifying traits that have overlapping genetic architecture with a disease of interest, investigators can curate the genetic determinants of such traits to generate an empirical list of candidate-SNPs and test these for pleiotropic association with the disease under study.(Semmes et al. 2020)

The strongest genetic risk factor for sporadic LOAD is the APOE $\epsilon 4$ allele, and a PheWAS of *APOE* gene alleles ($\epsilon 4$, $\epsilon 3$ and $\epsilon 2$) was recently performed.(Lumsden et al. 2020) Although that study identified 14 non-neurologic diseases associated with *APOE* alleles, including hypercholesterolemia and ischemic heart disease, a comprehensive PheWAS analysis of all known LOAD risk loci has not been conducted. We recently published an analytic strategy that integrates GWAS and PheWAS data for a disease of interest in order to identify both epidemiologic traits and trait-associated genetic variants that influence disease risk.(Semmes et al. 2020) By applying our approach to LOAD, we attempt to leverage genetic pleiotropy between LOAD and other traits to identify new LOAD risk factors and to nominate biological pathways through which LOAD risk alleles may exert their pathogenic effects.

MATERIALS AND METHODS

Selection of LOAD risk alleles. Hundreds of studies have investigated the genetic epidemiology of LOAD, with associated alleles replicating to highly variable degrees in subsequent research. Associated variants

discovered by GWAS are historically more likely to replicate in future studies due to their use of more stringent thresholds for determining statistical significance, lack of an *a priori* hypothesis, and tendency to externally replicate associated variants within the primary publication.(Walsh et al. 2013) We compiled a list of common (minor allele frequency ≥ 0.01 in European-ancestry populations) LOAD-associated genetic variants for PheWAS analysis from the most recent and largest GWAS meta-analysis of LOAD published at the time, which included 35,274 clinical and autopsy-documented LOAD cases and 59,163 controls genotyped or imputed for 9,456,058 SNPs.(Kunkle et al. 2019) Using a list of LOAD-associated variants reaching canonical genome-wide statistical significance (*i.e.* $P < 5.0 \times 10^{-8}$) in their study, we performed linkage disequilibrium (LD) pruning ($R^2 \leq 0.15$ in European-ancestry populations) using LDlink and retained the variant with the lowest P-value.(Machiela and Chanock 2015) A total of 24 independent LOAD-associated variants were carried forward for PheWAS analyses, including 19 SNPs identified by prior GWAS of LOAD and 5 novel variants that were successfully replicated within that publication.

Control SNP Set. We compiled a comparison SNP-set to serve as controls for PheWAS analyses. A set of unlinked control SNPs (1000 Genomes Project) was generated using SNPsnap (Broad Institute).(Pers, Timshel, and Hirschhorn 2015) Four control SNPs were matched to the LOAD-associated SNPs on: minor allele frequency ($\pm 5\%$), surrounding gene density ($\pm 50\%$), distance to nearest gene ($\pm 50\%$) and, as a proxy for haplotype block size, the number of other SNPs in LD at $R^2 \geq 0.50$ ($\pm 50\%$). For several LOAD risk SNPs, we could not generate more than 4 control SNPs without loosening our matching parameters, but the gain in statistical power achieved beyond a case-to-control ratio of 1:4 is typically minimal.(Hennessy et al. 1999) For one LOAD-associated SNP in the MHC class II region on chromosome 6 (rs9270816), no appropriate control SNPs could be matched. Therefore, rs9270816 was excluded from downstream analyses and a total of 23 LOAD-associated SNPs and 92 control SNPs were carried forward for PheWAS analyses.

PheWAS analyses. The UK Biobank atlas of genetic associations (<http://geneatlas.roslin.ed.ac.uk/>) was constructed by genotyping 452,264 European-ancestry individuals for 805,426 genetic variants, performing genome-wide SNP imputation and quality-controls, and linking genetic data to EHR data. GeneATLAS contains data for 778 traits (118 quantitative, 660 binary) and associations with 9,113,133 genetic variants (genotyped or imputed). GeneATLAS is searchable and can be queried for genetic (*e.g.*, SNPs) or phenotypic (*e.g.*, height) data to assess genotype-phenotype associations.(Canela-Xandri, Rawlik, and Tenesa 2018)

We queried GeneATLAS for trait associations with 23 known LOAD risk SNPs and the control SNP-set (92 matched SNPsnap controls). Summary statistics for traits associated with each queried variant were downloaded from GeneATLAS for downstream analyses. Nominally significant SNP-trait associations ($P < 0.01$) were carried forward to trait-enrichment analyses, as previously described.(Semmes et al. 2020) Although a more stringent p-value threshold for carrying SNP-trait associations forward was considered (*e.g.* 0.05/778), this was determined to be too conservative because many of the 778 traits in GeneATLAS have high genetic correlations with each other (*e.g.*, weight and hip circumference, 0.909; reticulocyte percentage and reticulocyte count, 0.952). Additionally, these individual SNP-trait associations were carried forward for trait-enrichment comparisons between LOAD-associated SNPs and the control SNP-set, and as such the PheWAS significance threshold is somewhat arbitrary so long as it is uniform across all SNP-sets.

PheWAS associations of the 23 LOAD SNPs with 778 phenotypic traits were compared to PheWAS results for the control SNP-set using the Python Programming Environment (<https://www.python.org>, version 3.6.9). For traits associated with >1 LOAD SNP, we used Fisher's exact tests to compare the frequency at which individual traits were associated with LOAD SNPs versus control SNPs to determine if traits were enriched for association with known LOAD risk variants. Fisher's exact p-values for trait enrichment underwent FDR correction to adjust for multiple testing.

SNPs associated with PheWAS-identified trait (platelet distribution width). We identified 546 unlinked ($R^2 \leq 0.15$ in European-ancestry populations) SNPs with minor allele frequency (MAF) >0.01 that were independently associated with platelet distribution width (PDW) at genome-wide statistical significance in a recent GWAS (N=563,085 subjects).(Vuckovic et al. 2020) Variants, effect alleles, effect sizes, and standard errors were extracted for downstream single-SNP association testing with LOAD and for two-sample Mendelian randomization analysis with LOAD.

Mendelian randomization analyses. To assess for a causal relationship between PDW and LOAD risk, we performed formal two-sample MR analyses with the R package “MendelianRandomization”.(Yavorska and Burgess 2017; Burgess and Thompson 2013) Using summary statistics of SNP-exposure (*i.e.*, PDW) and SNP-outcome (*i.e.*, LOAD) associations, we used the (1) inverse-variance weighted (IVW), (2) MR-Egger, and (3) weighted median methods to test for a causal relationship between PDW and LOAD risk.(Bowden et al. 2016; Bowden, Davey Smith, and Burgess 2015)

LOAD case-control data. To determine whether PDW-associated SNPs impact risk of LOAD at the single-variant level or under a Mendelian randomization framework, we leveraged SNP summary-level data from a recent GWAS of LOAD. This GWAS included 17,008 European-ancestry LOAD patients and 37,154 controls from the International Genomics of Alzheimer’s Project (I-GAP).(Lambert et al. 2013) Genotypes were imputed with the use of European population reference (EUR) haplotype data from the 1000 Genomes Project, and SNPs that did not pass study-level genotyping quality-control (QC) measures or had a MAF <0.01 were excluded, as were SNPs that did not pass study-level imputation QC measures (INFO score ≥ 0.30) or were not successfully genotyped/imputed in at least 40% of subjects across studies. The association of LOAD with genotype dosage was analyzed using logistic regression, adjusting for age, sex and ancestry-informative principal components (for three studies with incident LOAD data, Cox proportional hazards

models were used). A total of 7,055,881 SNPs with $MAF > 0.01$ are included in the final I-GAP meta-analysis results. Of 546 PDW-associated SNPs with $MAF > 0.01$ from Vuckovic, *et al.*, (Vuckovic et al. 2020) 384 had data available from I-GAP after QC filtering.

eQTL and *in silico* SNP functional analyses. We characterized SNPs in novel LOAD risk loci using HaploReg to annotate chromatin state and regulatory motifs. (Ward and Kellis 2016) We examined whether variants were protein-binding, located in DNase hypersensitive sites, promoter or enhancer histone marks, predicted to change transcription factor binding motifs, or functioned as expression quantitative trait loci (eQTL) in diverse tissues, (Ward and Kellis 2016) including the brain. (Zou et al. 2012)

RESULTS

An overview of the published methodology adapted to our study of LOAD is displayed in **Figure 1**. (Semmes et al. 2020) We first identified 24 common LOAD-associated SNPs from prior GWAS of European-ancestry populations, of which 23 were successfully matched to 4 control SNPs each (**Table 1 and Supplementary Table 1**). PheWAS analyses were then performed with the UK Biobank GeneATLAS database to test each LOAD-associated variant and control variant for association with 778 traits. After determining which PheWAS-identified traits were enriched for association with the LOAD SNP-set ($n = 23$) compared to the control SNP-set ($n = 92$), we identified SNPs associated with the enriched PheWAS-identified trait in prior GWAS. Using MR and candidate-SNP approaches, we tested whether enriched traits or trait-associated SNPs conferred LOAD risk in I-GAP data.

PheWAS analyses. 216 of the 778 traits in the GeneATLAS database were nominally associated ($P < 0.01$) with > 1 of the 23 LOAD-associated SNPs and were tested for enrichment versus the control SNP-set (**Table S2**). A total of eight traits were more likely to be associated with LOAD SNPs than with control SNPs at

nominal statistical significance ($P < 0.05$), including: platelet distribution width (PDW), platelet crit, platelet count, leukocyte count, immature reticulocyte fraction, cheese intake, high light scatter reticulocyte percentage, and current tobacco smoking (**Table 2**). PDW was the PheWAS-identified trait most enriched for association with LOAD SNPs, as 11 out of 23 (47.8%) LOAD SNPs were associated with platelet distribution width at $P < 0.01$ compared to 6 of 92 (6.5%) control SNPs (Fisher's exact $P = 1.2 \times 10^{-5}$). Moreover, this association survived both FDR correction ($P_{\text{FDR}} = 2.6 \times 10^{-3}$) and Bonferroni correction for all 778 traits examined ($P = 9.3 \times 10^{-3}$). We did not observe any consistent direction of effect between PDW and LOAD risk, as five LOAD risk SNPs were associated with increased PDW and six LOAD risk SNPs were associated with decreased PDW. Although the other seven traits appearing in Table 2 did not demonstrate significant enrichment for association with LOAD SNPs after FDR correction, it is noteworthy that they include two additional platelet-related traits (platelet crit and platelet count).

Mendelian randomization (MR) analyses. Although the eleven LOAD SNPs associated with PDW did not demonstrate associations in a consistent direction (*i.e.*, increased versus decreased PDW), we further evaluated the relationship between PDW and LOAD risk using a larger set of variants underlying PDW biology. To test for a causal relationship between PDW and LOAD, we used several MR analytical approaches leveraging 384 known PDW-associated SNPs from prior GWAS as instrumental variables and summary statistics for these SNPs from the I-GAP GWAS of LOAD. Estimates from IVW ($P_{\text{IVW}} = 0.27$), MR-Egger ($P_{\text{MR-Egger}} = 0.69$, $P_{\text{MR-intercept}} = 0.19$), and median-weighted ($P_{\text{MR-median}} = 0.64$) methods were all non-significant, indicating that PDW does not mediate LOAD risk and that there is likely no causal relationship between the traits.

PDW-associated SNPs as candidate LOAD risk loci. To examine whether individual PDW-associated variants might have pleiotropic effects on LOAD risk, we evaluated single-SNP associations of 384 PDW-

associated SNPs with LOAD risk in 17,008 LOAD cases and 37,154 controls from I-GAP. Thirty-six SNPs were nominally associated ($P < 0.05$) with LOAD case-control status (**Table S3**), of which five remained associated after FDR correction (**Table 3**). These associations confirmed known LOAD risk loci near *PICALM*, *CD2AP*, *SPII*, and *NDUFAF6*, and identified a novel risk locus at rs12535226 in the epidermal growth factor receptor (*EGFR*) gene. The rs12535226 variant is located the first intron of *EGFR* and has a minor allele frequency of 48% in European-ancestry populations. *In silico* functional analysis provided strong evidence of this variant having a regulatory role, as it is located in an enhancer region enriched for H3K27Ac histone marks in both brain tissues and neurosphere cultures, (Ward and Kellis 2016) and was identified as an *EGFR* expression quantitative trait locus (eQTL) in both the cerebellum and temporal cortex. (Zou et al. 2012)

DISCUSSION

We apply an innovative and previously-validated framework for leveraging existing GWAS and PheWAS data to uncover traits associated with Alzheimer's risk and to identify trait-associated variants as candidate LOAD risk loci. We first curated SNPs associated with LOAD from large GWAS meta-analyses and matched these LOAD risk variants to control SNPs on multiple parameters, including MAF, gene density, distance to nearest gene, and the number of other SNPs in LD. Performing PheWAS on these LOAD and control SNPs using UK Biobank data, we identified platelet distribution width (PDW) as being significantly enriched for association with LOAD risk loci. Although MR analysis did not implicate PDW as having a causal role in LOAD pathogenesis, a number of PDW-associated SNPs demonstrated pleiotropic associations with LOAD risk, including a novel association between LOAD and an *EGFR* eQTL variant.

Although PDW was significantly enriched for association with LOAD risk SNPs, MR analyses did not suggest the presence of vertical pleiotropy, wherein PDW directly contributes to LOAD risk. However, the large number of pleiotropic loci observed to influence both PDW and LOAD suggest that substantial

horizontal pleiotropy exists between these traits, which occurs when a variant affects LOAD directly or through multiple pathways not limited to platelet morphology. This appears reasonable given the emerging role of immune function in LOAD pathogenesis,(Lewcock et al. 2020) and is supported by our single-SNP association results that identified known and novel LOAD risk loci using PDW-associated variants as empirically-derived candidate SNPs. Confirming the validity of this approach, our analyses re-identified variants in *SP11*, a transcriptional activator of myeloid and B cell developmental gene expression previously linked to LOAD pathogenesis. They also identified a PDW-associated variant upstream of *NDUFAF6*, a region that has approached genome-wide statistical significance in a prior GWAS of LOAD ($P=9.2 \times 10^{-8}$) but has not yet emerged as a confirmed risk locus and was therefore not included among the LOAD SNPs used in our PheWAS analysis.

A novel LOAD risk locus in *EGFR* was implicated by our integrated analyses and the risk allele has been associated with elevated EGFR expression levels in human brain tissues.(Zou et al. 2012) EGFR, a member of the Human Epidermal Growth Factor Receptor 1 (ErbB) family, is a transmembrane tyrosine kinase receptor that initiates cell differentiation and proliferation. EGFR expression increases during neurogenesis and gliogenesis, and it activates pathways involved in oncogene expression.(Voldborg et al. 1997) EGFR plays a crucial role in the development of cancer, especially non-small lung cell carcinoma and glioblastoma, and extensive research has been completed evaluating it as a therapeutic target in the oncology setting.(da Cunha Santos, Shepherd, and Tsao 2011; Congdon et al. 2014) However, its function in the pathology of neurodegenerative diseases, such as LOAD, has received comparatively little attention.(Mansour et al. 2021)

A strongly supported theory for LOAD pathogenesis is the amyloid cascade hypothesis, positing that A β plaque accumulation leads to “synaptic dysfunction, tau pathology, and neuronal loss”.(Saraceno et al. 2013) The A β 42 isoform increases in quantity in forms of AD, forming the characteristic amyloid plaques associated with Alzheimer’s neuropathology.(Schmidt et al. 2009) In *Drosophila* model systems, research

suggests that A β -induced memory loss results from activation of EGFR by A β 42 oligomers, providing evidence that EGFR may function as a cell membrane receptor of A β peptides.(Wang et al. 2012) Following brain injury, EGFR expression increases, leading to the conversion of quiescent astrocytes to reactive astrocytes.(Liu and Neufeld 2007) While quiescent astrocytes have a constructive function in the developing CNS, reactive astrocytes are destructive and lead to neuronal loss. Recent tissue-based analyses have also identified significant differences in EGFR expression and protein levels between LOAD patients and control subjects,(He et al. 2021) and single-cell sequencing analyses implicate EGFR and MAP4K4 upregulation as a biomarker of altered ligand-receptor axis communication between neurons and microglia in AD patient brains.(Jian et al. 2021) These emerging data complement our findings and lend support to ongoing work that seeks to repurpose EGFR-targeted cancer therapies to inhibit EGFR as a potential pathway to rescuing memory loss and preventing neuronal loss in LOAD patients.(Tavassoly, Sato, and Tavassoly 2020)

There are several limitations to our study and valid concerns of our hybrid GWAS-PheWAS approach. For the PheWAS analyses, we used a p-value threshold of <0.01 to carry a SNP-trait association forward to enrichment analyses, rather than a Bonferroni-corrected threshold (*i.e.*, 0.05/778 traits). However, we found that this threshold maximized study power in our previous application of this GWAS-PheWAS approach to leukemia etiology.(Semmes et al. 2020) Additionally, this threshold is only for carrying traits forward to subsequent enrichment analyses, and associations from these enrichment comparisons were subjected to FDR correction to account for the number of traits evaluated for potential enrichment with LOAD SNPs. Another limitation is that our analysis relied upon traits evaluable in the U.K. Biobank, and therefore does not permit us to assess enrichment for potential disease-related traits that may have relevance to LOAD pathology, such as cerebral microbleeds or history of viral infections. Our PheWAS analyses also relied upon well-validated LOAD risk loci that had reached genome-wide statistical significance and been replicated in external datasets. Inclusion of additional putative risk alleles meeting less stringent cutoffs could potentially improve study power, but may also dilute effects by including false-positive LOAD

associations. Since completing our analyses, an additional GWAS meta-analysis of LOAD has been published, identifying several additional risk alleles.(Wightman et al. 2021) However, more than 50% of LOAD cases included in this study were “proxy cases” with a positive family history of LOAD, which may increase statistical power at the expense of precision. Such studies are useful adjuncts to study LOAD etiology, but do not lend themselves to MR approaches due to potential misspecification of SNP effect sizes.

Although multiple GWAS in the past decade have contributed to our understanding of inherited susceptibility to LOAD, there remains significant missing heritability.(Wightman et al. 2021; Kunkle et al. 2019; Lambert et al. 2013) A recent review on the benefits and pitfalls GWAS emphasized the need for novel analytic approaches to enhance our understanding of genotype-phenotype associations in the post-GWAS era and the utility of large biorepository databases linking EHR and genotyping data, polygenic scores, and innovative study designs.(Wang, Li, and Hakonarson 2010) While increasing GWAS sample size may reveal more associations, new methods for analyzing the wealth of existing data are essential, particularly in unraveling the underlying biological pathways involved in SNP-disease associations. In summary, by integrating GWAS and PheWAS datasets to decipher the complex etiology of LOAD, we identify horizontal pleiotropy with PDW and implicate EGFR variation in disease predisposition. These results should stimulate further mechanistic studies on the role of EGFR in AD pathogenesis to dissect their potential relationship. Such findings will have utility in informing the rational design of future prevention and treatment strategies for an aging population.

REFERENCES

- Andrews, S. J., B. Fulton-Howard, and A. Goate. 2020. 'Interpretation of risk loci from genome-wide association studies of Alzheimer's disease', *Lancet Neurol*, 19: 326-35.
- Bowden, J., G. Davey Smith, and S. Burgess. 2015. 'Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression', *Int J Epidemiol*, 44: 512-25.
- Bowden, J., G. Davey Smith, P. C. Haycock, and S. Burgess. 2016. 'Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator', *Genet Epidemiol*, 40: 304-14.
- Bruni, A. C., L. Bernardi, and C. Gabelli. 2020. 'From beta amyloid to altered proteostasis in Alzheimer's disease', *Ageing Res Rev*, 64: 101126.
- Burgess, S., and S. G. Thompson. 2013. 'Use of allele scores as instrumental variables for Mendelian randomization', *Int J Epidemiol*, 42: 1134-44.
- Bush, W. S., M. T. Oetjens, and D. C. Crawford. 2016. 'Unravelling the human genome-phenome relationship using phenome-wide association studies', *Nat Rev Genet*, 17: 129-45.
- Byun, J., Y. Han, Q. T. Ostrom, J. Edelson, K. M. Walsh, R. W. Pettit, M. L. Bondy, R. J. Hung, J. D. McKay, C. I. Amos, and Integral-Ilcco for. 2021. 'The Shared Genetic Architectures Between Lung Cancer and Multiple Polygenic Phenotypes in Genome-Wide Association Studies', *Cancer Epidemiol Biomarkers Prev*, 30: 1156-64.
- Canela-Xandri, O., K. Rawlik, and A. Tenesa. 2018. 'An atlas of genetic associations in UK Biobank', *Nat Genet*, 50: 1593-99.
- Congdon, K. L., P. C. Gedeon, C. M. Suryadevara, H. G. Caruso, L. J. Cooper, A. B. Heimberger, and J. H. Sampson. 2014. 'Epidermal growth factor receptor and variant III targeted immunotherapy', *Neuro Oncol*, 16 Suppl 8: viii20-5.
- da Cunha Santos, G., F. A. Shepherd, and M. S. Tsao. 2011. 'EGFR mutations and lung cancer', *Annu Rev Pathol*, 6: 49-69.
- Denny, J. C., L. Bastarache, and D. M. Roden. 2016. 'Phenome-Wide Association Studies as a Tool to Advance Precision Medicine', *Annu Rev Genomics Hum Genet*, 17: 353-73.
- Gao, S., A. E. Casey, T. J. Sargeant, and V. P. Makinen. 2018. 'Genetic variation within endolysosomal system is associated with late-onset Alzheimer's disease', *Brain*, 141: 2711-20.
- Handy, A., J. Lord, R. Green, J. Xu, D. Aarsland, L. Velayudhan, A. Hye, R. Dobson, P. Proitsi, initiative Alzheimer's Disease Neuroimaging, AddNeuroMed, and Gerad Consortium the. 2021. 'Assessing Genetic Overlap and Causality Between Blood Plasma Proteins and Alzheimer's Disease', *J Alzheimers Dis*, 83: 1825-39.
- He, Z., Y. Le Guen, L. Liu, J. Lee, S. Ma, A. C. Yang, X. Liu, J. Rutledge, P. M. Losada, B. Song, M. E. Belloy, R. R. Butler, 3rd, F. M. Longo, H. Tang, E. C. Mormino, T. Wyss-Coray, M. D. Greicius, and I. Ionita-Laza. 2021. 'Genome-wide analysis of common and rare variants via multiple knockoffs at biobank scale, with an application to Alzheimer disease genetics', *Am J Hum Genet*, 108: 2336-53.
- Hennessy, S., W. B. Bilker, J. A. Berlin, and B. L. Strom. 1999. 'Factors influencing the optimal control-to-case ratio in matched case-control studies', *Am J Epidemiol*, 149: 195-7.

- Jian, C., L. Wei, R. Mo, R. Li, L. Liang, L. Chen, C. Zou, Y. Meng, Y. Liu, and D. Zou. 2021. 'Microglia Mediate the Occurrence and Development of Alzheimer's Disease Through Ligand-Receptor Axis Communication', *Front Aging Neurosci*, 13: 731180.
- Kunkle, B. W., B. Grenier-Boley, R. Sims, J. C. Bis, V. Damotte, A. C. Naj, A. Boland, M. Vronskaya, S. J. van der Lee, A. Amlie-Wolf, C. Bellenguez, A. Frizatti, V. Chouraki, E. R. Martin, K. Sleegers, N. Badarinarayan, J. Jakobsdottir, K. L. Hamilton-Nelson, S. Moreno-Grau, R. Olaso, R. Raybould, Y. Chen, A. B. Kuzma, M. Hiltunen, T. Morgan, S. Ahmad, B. N. Vardarajan, J. Epelbaum, P. Hoffmann, M. Boada, G. W. Beecham, J. G. Garnier, D. Harold, A. L. Fitzpatrick, O. Valladares, M. L. Moutet, A. Gerrish, A. V. Smith, L. Qu, D. Bacq, N. Denning, X. Jian, Y. Zhao, M. Del Zompo, N. C. Fox, S. H. Choi, I. Mateo, J. T. Hughes, H. H. Adams, J. Malamon, F. Sanchez-Garcia, Y. Patel, J. A. Brody, B. A. Dombroski, M. C. D. Naranjo, M. Daniilidou, G. Eiriksdottir, S. Mukherjee, D. Wallon, J. Uphill, T. Aspelund, L. B. Cantwell, F. Garzia, D. Galimberti, E. Hofer, M. Butkiewicz, B. Fin, E. Scarpini, C. Sarnowski, W. S. Bush, S. Meslage, J. Kornhuber, C. C. White, Y. Song, R. C. Barber, S. Engelborghs, S. Sordon, D. Voijnovic, P. M. Adams, R. Vandenberghe, M. Mayhaus, L. A. Cupples, M. S. Albert, P. P. De Deyn, W. Gu, J. J. Himali, D. Beekly, A. Squassina, A. M. Hartmann, A. Orellana, D. Blacker, E. Rodriguez-Rodriguez, S. Lovestone, M. E. Garcia, R. S. Doody, C. Munoz-Fernandez, R. Sussams, H. Lin, T. J. Fairchild, Y. A. Benito, C. Holmes, H. Karamujic-Comic, M. P. Frosch, H. Thonberg, W. Maier, G. Roschupkin, B. Ghetti, V. Giedraitis, A. Kawalia, S. Li, R. M. Huebinger, L. Kilander, S. Moebus, I. Hernandez, M. I. Kamboh, R. Brundin, J. Turton, Q. Yang, M. J. Katz, L. Concar, J. Lord, A. S. Beiser, C. D. Keene, S. Helisalmi, I. Kloszewska, W. A. Kukull, A. M. Koivisto, A. Lynch, L. Tarraga, E. B. Larson, A. Haapasalo, B. Lawlor, T. H. Mosley, R. B. Lipton, V. Solfrizzi, M. Gill, W. T. Longstreth, Jr., T. J. Montine, V. Frisardi, M. Diez-Fairen, F. Rivadeneira, R. C. Petersen, V. Deramecourt, I. Alvarez, F. Salani, A. Ciarrella, E. Boerwinkle, E. M. Reiman, N. Fievet, J. I. Rotter, J. S. Reisch, O. Hanon, C. Cupidi, A. G. Andre Uitterlinden, D. R. Royall, C. Dufouil, R. G. Maletta, I. de Rojas, M. Sano, A. Brice, R. Cecchetti, P. S. George-Hyslop, K. Ritchie, M. Tsolaki, D. W. Tsuang, B. Dubois, D. Craig, C. K. Wu, H. Soininen, D. Avramidou, R. L. Albin, L. Fratiglioni, A. Germanou, L. G. Apostolova, L. Keller, M. Koutroumani, S. E. Arnold, F. Panza, O. Gkatzima, S. Asthana, D. Hannequin, P. Whitehead, C. S. Atwood, P. Caffarra, H. Hampel, I. Quintela, A. Carracedo, L. Lannfelt, D. C. Rubinsztein, L. L. Barnes, F. Pasquier, L. Frolich, S. Barral, B. McGuinness, T. G. Beach, J. A. Johnston, J. T. Becker, P. Passmore, E. H. Bigio, J. M. Schott, T. D. Bird, J. D. Warren, B. F. Boeve, M. K. Lupton, J. D. Bowen, P. Proitsi, A. Boxer, J. F. Powell, J. R. Burke, J. S. K. Kauwe, J. M. Burns, M. Mancuso, J. D. Buxbaum, U. Bonuccelli, N. J. Cairns, A. McQuillin, C. Cao, G. Livingston, C. S. Carlson, N. J. Bass, C. M. Carlsson, J. Hardy, R. M. Carney, J. Bras, M. M. Carrasquillo, R. Guerreiro, M. Allen, H. C. Chui, E. Fisher, C. Masullo, E. A. Crocco, C. DeCarli, G. Bisceglia, M. Dick, L. Ma, R. Duara, N. R. Graff-Radford, D. A. Evans, A. Hodges, K. M. Faber, M. Scherer, K. B. Fallon, M. Riemenschneider, D. W. Fardo, R. Heun, M. R. Farlow, H. Kolsch, S. Ferris, M. Leber, T. M. Foroud, I. Heuser, D. R. Galasko, I. Giegling, M. Gearing, M. Hull, D. H. Geschwind, J. R. Gilbert, J. Morris, R. C. Green, K. Mayo, J. H. Growdon, T. Feulner, R. L. Hamilton, L. E. Harrell, D. Drichel, L. S. Honig, T. D. Cushman, M. J. Huentelman, P. Hollingworth, C. M. Hulette, B. T. Hyman, R. Marshall, G. P. Jarvik, A. Meggy, E. Abner, G. E. Menzies, L. W. Jin, G. Leonenko, L. M. Real, G. R. Jun, C. T. Baldwin, D. Grozeva, A. Karydas, G. Russo, J. A. Kaye, R. Kim, F. Jessen, N. W. Kowall, B. Vellas, J. H. Kramer, E. Vardy, F. M. LaFerla, K. H. Jockel, J. J. Lah, M. Dichgans, J. B. Leverenz, D. Mann, A. I. Levey, S. Pickering-Brown, A. P. Lieberman, N. Klopp, K. L. Lunetta, H. E. Wichmann, C. G. Lyketsos, K. Morgan, D. C. Marson, K. Brown, F. Martiniuk, C. Medway, D. C. Mash, M. M. Nothen, E. Masliah, N. M. Hooper, W. C. McCormick, A. Daniele, S. M. McCurry, A. Bayer, A. N. McDavid, J. Gallacher, A. C. McKee, H.

van den Bussche, M. Mesulam, C. Brayne, B. L. Miller, S. Riedel-Heller, C. A. Miller, J. W. Miller, A. Al-Chalabi, J. C. Morris, C. E. Shaw, A. J. Myers, J. Wiltfang, S. O'Bryant, J. M. Olichney, V. Alvarez, J. E. Parisi, A. B. Singleton, H. L. Paulson, J. Collinge, W. R. Perry, S. Mead, E. Peskind, D. H. Cribbs, M. Rossor, A. Pierce, N. S. Ryan, W. W. Poon, B. Nacmias, H. Potter, S. Sorbi, J. F. Quinn, E. Sacchinelli, A. Raj, G. Spalletta, M. Raskind, C. Caltagirone, P. Bossu, M. D. Orfei, B. Reisberg, R. Clarke, C. Reitz, A. D. Smith, J. M. Ringman, D. Warden, E. D. Roberson, G. Wilcock, E. Rogaeva, A. C. Bruni, H. J. Rosen, M. Gallo, R. N. Rosenberg, Y. Ben-Shlomo, M. A. Sager, P. Mecocci, A. J. Saykin, P. Pastor, M. L. Cuccaro, J. M. Vance, J. A. Schneider, L. S. Schneider, S. Slifer, W. W. Seeley, A. G. Smith, J. A. Sonnen, S. Spina, R. A. Stern, R. H. Swerdlow, M. Tang, R. E. Tanzi, J. Q. Trojanowski, J. C. Troncoso, V. M. Van Deerlin, L. J. Van Eldik, H. V. Vinters, J. P. Vonsattel, S. Weintraub, K. A. Welsh-Bohmer, K. C. Wilhelmsen, J. Williamson, T. S. Wingo, R. L. Woltjer, C. B. Wright, C. E. Yu, L. Yu, Y. Saba, Consortium Alzheimer Disease Genetics, Initiative European Alzheimer's Disease, Heart Cohorts for, Consortium Aging Research in Genomic Epidemiology, Genetic, Polygenic Environmental Risk in Ad/Defining Genetic, Consortium Environmental Risk for Alzheimer's Disease, A. Pilotto, M. J. Bullido, O. Peters, P. K. Crane, D. Bennett, P. Bosco, E. Coto, V. Boccardi, P. L. De Jager, A. Lleo, N. Warner, O. L. Lopez, M. Ingelsson, P. Deloukas, C. Cruchaga, C. Graff, R. Gwilliam, M. Fornage, A. M. Goate, P. Sanchez-Juan, P. G. Kehoe, N. Amin, N. Ertekin-Taner, C. Berr, S. Debette, S. Love, L. J. Launer, S. G. Younkin, J. F. Dartigues, C. Corcoran, M. A. Ikram, D. W. Dickson, G. Nicolas, D. Campion, J. Tschanz, H. Schmidt, H. Hakonarson, J. Clarimon, R. Munger, R. Schmidt, L. A. Farrer, C. Van Broeckhoven, C. O' Donovan M, A. L. DeStefano, L. Jones, J. L. Haines, J. F. Deleuze, M. J. Owen, V. Gudnason, R. Mayeux, V. Escott-Price, B. M. Psaty, A. Ramirez, L. S. Wang, A. Ruiz, C. M. van Duijn, P. A. Holmans, S. Seshadri, J. Williams, P. Amouyel, G. D. Schellenberg, J. C. Lambert, and M. A. Pericak-Vance. 2019. 'Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing', *Nat Genet*, 51: 414-30.

Lambert, J. C., C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, A. L. DeStafano, J. C. Bis, G. W. Beecham, B. Grenier-Boley, G. Russo, T. A. Thorton-Wells, N. Jones, A. V. Smith, V. Chouraki, C. Thomas, M. A. Ikram, D. Zelenika, B. N. Vardarajan, Y. Kamatani, C. F. Lin, A. Gerrish, H. Schmidt, B. Kunkle, M. L. Dunstan, A. Ruiz, M. T. Bihoreau, S. H. Choi, C. Reitz, F. Pasquier, C. Cruchaga, D. Craig, N. Amin, C. Berr, O. L. Lopez, P. L. De Jager, V. Deramecourt, J. A. Johnston, D. Evans, S. Lovestone, L. Letenneur, F. J. Moron, D. C. Rubinsztein, G. Eiriksdottir, K. Sleegers, A. M. Goate, N. Fievet, M. W. Huentelman, M. Gill, K. Brown, M. I. Kamboh, L. Keller, P. Barberger-Gateau, B. McGuinness, E. B. Larson, R. Green, A. J. Myers, C. Dufouil, S. Todd, D. Wallon, S. Love, E. Rogaeva, J. Gallacher, P. St George-Hyslop, J. Clarimon, A. Lleo, A. Bayer, D. W. Tsuang, L. Yu, M. Tsolaki, P. Bossu, G. Spalletta, P. Proitsi, J. Collinge, S. Sorbi, F. Sanchez-Garcia, N. C. Fox, J. Hardy, M. C. Deniz Naranjo, P. Bosco, R. Clarke, C. Brayne, D. Galimberti, M. Mancuso, F. Matthews, Initiative European Alzheimer's Disease, Genetic, Disease Environmental Risk in Alzheimer's, Consortium Alzheimer's Disease Genetic, Heart Cohorts for, Epidemiology Aging Research in Genomic, S. Moebus, P. Mecocci, M. Del Zompo, W. Maier, H. Hampel, A. Pilotto, M. Bullido, F. Panza, P. Caffarra, B. Nacmias, J. R. Gilbert, M. Mayhaus, L. Lanefelt, H. Hakonarson, S. Pichler, M. M. Carrasquillo, M. Ingelsson, D. Beekly, V. Alvarez, F. Zou, O. Valladares, S. G. Younkin, E. Coto, K. L. Hamilton-Nelson, W. Gu, C. Razquin, P. Pastor, I. Mateo, M. J. Owen, K. M. Faber, P. V. Jonsson, O. Combarros, M. C. O'Donovan, L. B. Cantwell, H. Soininen, D. Blacker, S. Mead, T. H. Mosley, Jr., D. A. Bennett, T. B. Harris, L. Fratiglioni, C. Holmes, R. F. de Bruijn, P. Passmore, T. J. Montine, K. Bettens, J. I. Rotter, A. Brice, K. Morgan, T. M. Foroud, W. A. Kukull, D. Hannequin, J. F. Powell, M. A. Nalls, K. Ritchie, K. L. Lunetta, J. S. Kauwe, E. Boerwinkle, M. Riemenschneider, M. Boada, M. Hiltunen, E. R. Martin, R. Schmidt, D.

- Rujescu, L. S. Wang, J. F. Dartigues, R. Mayeux, C. Tzourio, A. Hofman, M. M. Nothen, C. Graff, B. M. Psaty, L. Jones, J. L. Haines, P. A. Holmans, M. Lathrop, M. A. Pericak-Vance, L. J. Launer, L. A. Farrer, C. M. van Duijn, C. Van Broeckhoven, V. Moskvina, S. Seshadri, J. Williams, G. D. Schellenberg, and P. Amouyel. 2013. 'Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease', *Nat Genet*, 45: 1452-8.
- Lewcock, J. W., K. Schlepckow, G. Di Paolo, S. Tahirovic, K. M. Monroe, and C. Haass. 2020. 'Emerging Microglia Biology Defines Novel Therapeutic Approaches for Alzheimer's Disease', *Neuron*, 108: 801-21.
- Li, M., J. Lin, S. Liang, Z. Chen, Y. Bai, X. Long, S. Huang, and Z. Mo. 2021. 'The role of age at menarche and age at menopause in Alzheimer's disease: evidence from a bidirectional mendelian randomization study', *Aging (Albany NY)*, 13: 19722-49.
- Liu, B., and A. H. Neufeld. 2007. 'Activation of epidermal growth factor receptors in astrocytes: from development to neural injury', *J Neurosci Res*, 85: 3523-9.
- Lumsden, A. L., A. Mulugeta, A. Zhou, and E. Hypponen. 2020. 'Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank', *EBioMedicine*, 59: 102954.
- Machiela, M. J., and S. J. Chanock. 2015. 'LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants', *Bioinformatics*, 31: 3555-7.
- Mansour, H. M., H. M. Fawzy, A. S. El-Khatib, and M. M. Khattab. 2021. 'Potential Repositioning of Anti-cancer EGFR Inhibitors in Alzheimer's Disease: Current Perspectives and Challenging Prospects', *Neuroscience*, 469: 191-96.
- Millard, L. A. C., M. R. Munafo, K. Tilling, R. E. Wootton, and G. Davey Smith. 2019. 'MR-pheWAS with stratification and interaction: Searching for the causal effects of smoking heaviness identified an effect on facial aging', *PLoS Genet*, 15: e1008353.
- Pers, T. H., P. Timshel, and J. N. Hirschhorn. 2015. 'SNPsnap: a Web-based tool for identification and annotation of matched SNPs', *Bioinformatics*, 31: 418-20.
- Pettit, R. W., J. Byun, Y. Han, Q. T. Ostrom, J. Edelson, K. M. Walsh, M. L. Bondy, R. J. Hung, J. D. McKay, and C. I. Amos. 2021. 'The shared genetic architecture between epidemiological and behavioral traits with lung cancer', *Sci Rep*, 11: 17559.
- Rabinovici, G. D. 2019. 'Late-onset Alzheimer Disease', *Continuum (Minneapolis)*, 25: 14-33.
- Saraceno, C., S. Musardo, E. Marcello, S. Pelucchi, and M. Di Luca. 2013. 'Modeling Alzheimer's disease: from past to future', *Front Pharmacol*, 4: 77.
- Schmidt, M., C. Sachse, W. Richter, C. Xu, M. Fandrich, and N. Grigorieff. 2009. 'Comparison of Alzheimer Abeta(1-40) and Abeta(1-42) amyloid fibrils reveals similar protofilament structures', *Proc Natl Acad Sci U S A*, 106: 19813-8.
- Semmes, E. C., J. Vijayakrishnan, C. Zhang, J. H. Hurst, R. S. Houlston, and K. M. Walsh. 2020. 'Leveraging Genome and Phenome-Wide Association Studies to Investigate Genetic Risk of Acute Lymphoblastic Leukemia', *Cancer Epidemiol Biomarkers Prev*, 29: 1606-14.

- Semmes, E. C., C. Zhang, and K. M. Walsh. 2018. 'Intermediate phenotypes underlying osteosarcoma risk', *Oncotarget*, 9: 37345-46.
- Shi, Y., R. Liu, Y. Guo, Q. Li, H. Zhou, S. Yu, H. Liang, and Z. Li. 2021. 'An Updated Mendelian Randomization Analysis of the Association Between Serum Calcium Levels and the Risk of Alzheimer's Disease', *Front Genet*, 12: 731391.
- Tavassoly, O., T. Sato, and I. Tavassoly. 2020. 'Inhibition of Brain Epidermal Growth Factor Receptor Activation: A Novel Target in Neurodegenerative Diseases and Brain Injuries', *Mol Pharmacol*, 98: 13-22.
- Voldborg, B. R., L. Damstrup, M. Spang-Thomsen, and H. S. Poulsen. 1997. 'Epidermal growth factor receptor (EGFR) and EGFR mutations, function and possible role in clinical trials', *Ann Oncol*, 8: 1197-206.
- Vuckovic, D., E. L. Bao, P. Akbari, C. A. Lareau, A. Mousas, T. Jiang, M. H. Chen, L. M. Raffield, M. Tardaguila, J. E. Huffman, S. C. Ritchie, K. Megy, H. Pongstingl, C. J. Penkett, P. K. Albers, E. M. Wigdor, S. Sakaue, A. Moscati, R. Manansala, K. S. Lo, H. Qian, M. Akiyama, T. M. Bartz, Y. Ben-Shlomo, A. Beswick, J. Bork-Jensen, E. P. Bottinger, J. A. Brody, F. J. A. van Rooij, K. N. Chitralla, P. W. F. Wilson, H. Choquet, J. Danesh, E. Di Angelantonio, N. Dimou, J. Ding, P. Elliott, T. Esko, M. K. Evans, S. B. Felix, J. S. Floyd, L. Broer, N. Grarup, M. H. Guo, Q. Guo, A. Greinacher, J. Haessler, T. Hansen, J. M. M. Howson, W. Huang, E. Jorgenson, T. Kacprowski, M. Kahonen, Y. Kamatani, M. Kanai, S. Karthikeyan, F. Koskeridis, L. A. Lange, T. Lehtimäki, A. Linneberg, Y. Liu, L. P. Lyytikäinen, A. Manichaikul, K. Matsuda, K. L. Mohlke, N. Mononen, Y. Murakami, G. N. Nadkarni, K. Nikus, N. Pankratz, O. Pedersen, M. Preuss, B. M. Psaty, O. T. Raitakari, S. S. Rich, B. A. T. Rodriguez, J. D. Rosen, J. I. Rotter, P. Schubert, C. N. Spracklen, P. Surendran, H. Tang, J. C. Tardif, M. Ghanbari, U. Volker, H. Volzke, N. A. Watkins, S. Weiss, V. A. Million Veteran Program, N. Cai, K. Kundu, S. B. Watt, K. Walter, A. B. Zonderman, K. Cho, Y. Li, R. J. F. Loos, J. C. Knight, M. Georges, O. Stegle, E. Evangelou, Y. Okada, D. J. Roberts, M. Inouye, A. D. Johnson, P. L. Auer, W. J. Astle, A. P. Reiner, A. S. Butterworth, W. H. Ouwehand, G. Lettre, V. G. Sankaran, and N. Soranzo. 2020. 'The Polygenic and Monogenic Basis of Blood Traits and Diseases', *Cell*, 182: 1214-31 e11.
- Walsh, K. M., E. Anderson, H. M. Hansen, P. A. Decker, M. L. Kosel, T. Kollmeyer, T. Rice, S. Zheng, Y. Xiao, J. S. Chang, L. S. McCoy, P. M. Bracci, J. L. Wiemels, A. R. Pico, I. Smirnov, D. H. Lachance, H. Sicotte, J. E. Eckel-Passow, J. K. Wiencke, R. B. Jenkins, and M. R. Wrensch. 2013. 'Analysis of 60 reported glioma risk SNPs replicates published GWAS findings but fails to replicate associations from published candidate-gene studies', *Genet Epidemiol*, 37: 222-8.
- Wang, K., M. Li, and H. Hakonarson. 2010. 'Analysing biological pathways in genome-wide association studies', *Nat Rev Genet*, 11: 843-54.
- Wang, L., H. C. Chiang, W. Wu, B. Liang, Z. Xie, X. Yao, W. Ma, S. Du, and Y. Zhong. 2012. 'Epidermal growth factor receptor is a preferred target for treating amyloid-beta-induced memory loss', *Proc Natl Acad Sci U S A*, 109: 16743-8.
- Ward, L. D., and M. Kellis. 2016. 'HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease', *Nucleic Acids Res*, 44: D877-81.
- Wightman, D. P., I. E. Jansen, J. E. Savage, A. A. Shadrin, S. Bahrami, D. Holland, A. Rongve, S. Borte, B. S. Winsvold, O. K. Drange, A. E. Martinsen, A. H. Skogholt, C. Willer, G. Brathén, I. Bosnes, J. B. Nielsen, L. G. Fritsche, L. F. Thomas, L. M. Pedersen, M. E. Gabrielsen, M. B. Johnsen, T. W.

Meisingset, W. Zhou, P. Proitsi, A. Hodges, R. Dobson, L. Velayudhan, Team and Me Research, J. M. Sealock, L. K. Davis, N. L. Pedersen, C. A. Reynolds, I. K. Karlsson, S. Magnusson, H. Stefansson, S. Thordardottir, P. V. Jonsson, J. Snaedal, A. Zettergren, I. Skoog, S. Kern, M. Waern, H. Zetterberg, K. Blennow, E. Stordal, K. Hveem, J. A. Zwart, L. Athanasiu, P. Selnes, I. Saltvedt, S. B. Sando, I. Ulstein, S. Djurovic, T. Fladby, D. Aarsland, G. Selbaek, S. Ripke, K. Stefansson, O. A. Andreassen, and D. Posthuma. 2021. 'A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease', *Nat Genet*, 53: 1276-82.

Yavorska, O. O., and S. Burgess. 2017. 'MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data', *Int J Epidemiol*, 46: 1734-39.

Zhang, Z., M. Wang, and X. Liu. 2021. 'C-reactive protein and risk of Alzheimer's disease', *Neurobiol Aging*.

Zou, F., H. S. Chai, C. S. Younkin, M. Allen, J. Crook, V. S. Pankratz, M. M. Carrasquillo, C. N. Rowley, A. A. Nair, S. Middha, S. Maharjan, T. Nguyen, L. Ma, K. G. Malphrus, R. Palusak, S. Lincoln, G. Bisceglia, C. Georgescu, N. Kouri, C. P. Kolbert, J. Jen, J. L. Haines, R. Mayeux, M. A. Pericak-Vance, L. A. Farrer, G. D. Schellenberg, Consortium Alzheimer's Disease Genetics, R. C. Petersen, N. R. Graff-Radford, D. W. Dickson, S. G. Younkin, and N. Ertekin-Taner. 2012. 'Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants', *PLoS Genet*, 8: e1002707.

FIGURE LEGENDS:

Figure 1. Methodology for integrative analysis of GWAS and PheWAS data. This figure illustrates our approach for investigating phenotype associations with known disease risk-variants in order to identify novel candidate risk loci and/or intermediate phenotypes for subsequent analysis in case-control cohorts. Applying this approach to late-onset Alzheimer's disease (LOAD) identified platelet distribution width (PDW) as a phenotype enriched for association with LOAD GWAS hits and informed downstream analysis of the role of PDW-associated variants in conferring LOAD risk in a case-control sample. Created with Biorender.com.

^a Genome-wide meta-analysis of 35,274 clinical and autopsy-documented LOAD cases and 59,163 controls (Kunkle, *et al.*)

^b PheWAS catalog (UK Biobank GeneATLAS) - <http://geneatlas.roslin.ed.ac.uk/phewas/>

^c Four control SNPs were matched to each LOAD-associated SNP on minor allele frequency ($\pm 5\%$), surrounding gene density ($\pm 50\%$), distance to nearest gene ($\pm 50\%$) and the number of other SNPs in LD at $R^2 \geq 0.50$ ($\pm 50\%$) using the SNPsnap software (Broad Institute) - <https://data.broadinstitute.org/mpg/snpnap/>

^d Genome-wide meta-analysis of platelet traits in 563,085 subjects (Vuckovic, *et al.*)

^e Summary-level SNP association data available from the I-GAP genome-wide meta-analysis LOAD patients and controls (Lambert, *et al.*)

Abbreviations: GWAS, genome-wide association study; PheWAS, phenome-wide association study; LOAD, late-onset Alzheimer's disease; SNPs, single nucleotide polymorphisms.

Table 1. Common genetic variants associated with late-onset Alzheimer's disease (LOAD) in prior GWAS and included in our PheWAS-based enrichment analyses.

Position (Chr:BP) ^a	rsID	MAF ^b	Function	Nearest Gene
1:207802552	rs4844610	0.19	intronic	<i>CR1</i>
2:127892810	rs6733839	0.40	intergenic	<i>BIN1</i>
2:233981912	rs10933431	0.22	intergenic	<i>INPP5D</i>
6:47431284	rs9473117	0.27	intergenic	<i>CD2AP</i>
7:100091795	rs12539172	0.32	3' UTR	<i>NYAP1</i>
7:143099133	rs10808026	0.23	intronic	<i>EPHA1</i>
8:27219987	rs73223431	0.35	intronic	<i>PTK2B</i>
8:27467686	rs9331896	0.40	intronic	<i>CLU</i>
10:11720308	rs7920721	0.36	intergenic	<i>ECHDC3</i>
11:47380340	rs3740688	0.46	intronic	<i>SPI1</i>
11:59936926	rs7933202	0.40	intergenic	<i>MS4A6A</i>
11:85868640	rs3851179	0.38	intergenic	<i>PICALM</i>
11:121435587	rs11218343	0.04	intronic	<i>SORL1</i>
14:53391680	rs17125924	0.08	intronic	<i>FERMT2</i>
14:92932828	rs12881735	0.22	intronic	<i>SLC24A4</i>
15:59045774	rs593742	0.31	intergenic	<i>ADAM10</i>
16:19808163	rs7185636	0.15	intronic	<i>IQCK</i>
16:79355857	rs62039712	0.10	intergenic	<i>WWOX</i>
17:61538148	rs138190086	0.01	intergenic	<i>ACE</i>
19:1056492	rs3752246	0.18	missense	<i>ABCA7</i>
19:45411941	rs429358	0.14	missense	<i>APOE</i>
20:54997568	rs6024870	0.08	intronic	<i>CASS4</i>
21:28156856	rs2830500	0.30	intergenic	<i>ADAMTS1</i>

^aPosition in GRCh37/hg19

^bMinor Allele Frequency in European-ancestry subjects from the 1000 Genomes Project

Table 2: Traits more likely to be associated with late-onset Alzheimer’s disease (LOAD) risk SNPs than with control SNPs in PheWAS enrichment analyses.

Trait	Associated LOAD SNPs	Associated control SNPs	P-value	P-value _(FDR)
Platelet distribution width	11/23	6/92	1.2×10^{-5}	2.6×10^{-3}
Platelet crit	10/23	11/92	1.4×10^{-3}	0.15
Leukocyte count	7/23	5/92	2.2×10^{-3}	0.16
Cheese intake	4/23	1/92	5.5×10^{-3}	0.25
Platelet count	8/23	9/92	5.9×10^{-3}	0.25
Immature reticulocyte fraction	5/23	3/92	7.9×10^{-3}	0.28
High light scatter reticulocyte percentage	6/23	6/92	0.014	0.39
Current tobacco smoking	4/23	2/92	0.014	0.39

Table 3: Variants achieving genome-wide statistical significance in a prior GWAS of platelet distribution width (Vuckovic, *et al.*) and their association with late-onset Alzheimer's disease (LOAD) risk (17,008 patients; 37,154 controls). A total of 384 sentinel variants from Vuckovic, *et al.* were assessed for association with LOAD risk and appear in Supplementary Table 3, of which five remained associated after FDR correction and appear here.

Variant ID	Position (Chr:BP) ^a	P-value	P-value _(FDR)	Nearest gene(s)	Previously-identified LOAD risk locus
rs3851179	11:85868640	2.8x10 ⁻¹⁵	1.1x10 ⁻¹²	<i>EED, PICALM</i>	Yes
rs4711890	6:47606712	4.9x10 ⁻⁶	9.4x10 ⁻⁴	<i>CD2AP</i>	Yes
rs7107356	11:47676170	8.4x10 ⁻⁶	1.1x10 ⁻³	<i>MTCH2, AGBL2</i>	Yes
rs12535226	7:55156419	4.3x10 ⁻⁴	0.034	<i>EGFR</i>	No
rs4735333	8:95955074	4.5x10 ⁻⁴	0.034	<i>TP53INP1, NDUFAF6</i>	Yes ^b

^aPosition in GRCh37/hg19

^bThis region approached genome-wide statistical significance in a prior GWAS of LOAD(Kunkle *et al.* 2019) (lead SNP rs4735340, P=9.2x10⁻⁸), but was not independently replicated and thus is not yet considered a validated LOAD risk locus.

Figure 1