**Novel and future insights into the complex genetic etiology of Alzheimer’s disease**

Shea J Andrews1; Brian-Fulton Howard2; Alan E Renton2; Alison Goate2

1Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, California, USA

2Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai New York, NY, USA

**SYNOPSIS**

Genome-wide association studies (GWAS) over the last decade and a half have made enormous contributions to our understanding of Alzheimer’s disease (AD) by identifying microglia as the key cell type and brain lipid metabolism as the core biological pathway underlying how the brain responds to amyloid deposition. However, compared to other neurodegenerative and neuropsychiatric diseases, until recently there were relatively few genetic loci associated with AD. This has been addressed with the publication of two new GWAS of AD (Wightman et al 2021; Bellenguez et al 2022) that dramatically increase the sample size (n = 1,126,563 and n = 788,989) and double the number of known loci associated with AD (n = 38 and n = 75). GWAS have now identified 103 independent AD-association signals across 82 loci (Fig. 1). Gene prioritization has further highlighted that many candidate causal genes are expressed in microglia or other myeloid-lineage cells (Bellenguez et al 2022; Novikova et al 2021), implicating microglia in AD pathogenesis, potentially through efferocytosis, which is the phagocytosis of lipid rich cellular debris by innate immune cells. The summary statistics from these GWAS will also be critical for conducting downstream polygenic risk score analyses (PRS) for disease classification and prediction, and using genetic correlation & Mendelian randomization to evaluate shared genetic architecture between traits (Bellenguez et al 2022; Wu et al 2021; Andrews et al 2021).

Heritability estimates of AD from GWAS, however, have highlighted a problem - as sample sizes have increased, the heritability of AD has decreased. Similarly, the number of loci discovered does not scale with sample size as expected. This has been attributed to poor diagnostic accuracy and lack of age matching, both of which are exacerbated by the inclusion of proxy-case/control cohorts, and suggests that continually increasing AD GWAS sample sizes will lead to diminishing returns (Escott-Price & Hardy et al 2022). However, we note that the discrepancy between loci discovered and total sample size can be attributed to AD GWAS having imbalanced case/control designs and that the number of loci discovered does scale with effective sample size (Fig. 2). Additionally, current heritability estimates have remained stable (5-7%) after accounting for differences in effective sample size between cohorts and misspecifications in models that combine clinically-diagnosed cases with proxy cases (Fuente et al 2022; Grotzinger et al 2022). As such while diagnostic accuracy, age matching, and data transparency are ongoing issues that need to be addressed, increasing effective sample size in conjunction with improved study designs will remain important in the context of multiple discovery gaps.

There are multiple discovery gaps that GWAS can focus on to further explore the genetic underpinnings of AD pathogenesis. First, current GWAS are not adequately powered or designed to detect rare variants which likely explain some of the missing heritability. To identify rare variants, integration of next generation sequencing studies and larger GWAS imputed using modern ancestrally diverse reference panels will be critical and this approach has been successfully applied in discovery of two rare protective *APOE* variants (Guen et al 2022). Second, the majority of participants in AD GWAS are of European ancestry, despite the prevalence of AD being higher in minority populations. GWAS of AD in African American (Kunkle et al 2021) and East Asian (Shigemizu et al 2021) populations remain severely underpowered on their own, but have discovered novel loci, ancestry-specific rare alleles, and ancestry-specific effects of known loci such as *APOE e4* — which has a relatively weaker effect on AD risk in African populations than in European populations despite a higher allele frequency. Although European sample sizes for microarray-based case/control AD GWAS are already large, increasing ancestrally diverse sample sizes will ensure study populations are globally representative, maximizing opportunities for scientific and medical advances to be shared equitably and improving discovery in all populations. Finally, plasma biomarkers offer new opportunities for using genomics to understand the genetic etiology of AD. Plasma biomarkers will allow for improved diagnostic accuracy of cases/controls, improving the power to detect loci specific to AD and potentially ameliorate the reported trend towards lower heritability in larger GWAS (Escott-Price & Hardy 2022). Identifying genetic loci associated with specific biomarkers will also aid in the understanding of the core pathophysiological changes that underpin AD (Damotte et al 2021; Lord et al 2021). Addressing these discovery gaps will require robust funding for and an institutional focus on high quality phenotyping and genotyping across all ancestries. This will be facilitated by effective data sharing to make sure data are discoverable, easy to access, consistently phenotyped, and available in forms that are easy to process in a homogenous manner for consistency across studies. The Alzheimer’s Disease Sequencing Project (ADSP) embodies these principles and will significantly contribute to achieving these goals.

In summary, the largest AD GWAS to date have substantially improved our understanding of the genetic architecture underpinning AD. Increasing the effective sample sizes of GWAS investigating rare variants, diverse populations, and plasma biomarkers, facilitated by effective data sharing, are now required to further address the discovery gaps in AD genetic architecture.

**OUTLINE**

***Introduction***

* What is Alzheimer’s disease
* Brief overview of AD genetics including autosomal AD genetics and previous GWAS, summarizing previous rapid review (Andrews et al 2020)
* Is there more to discover using AD GWAS?

***Recent developments***

*Advances in AD GWAS*

* Overview of findings from Wightman et al 2021
* Overview of findings from Bellenguez et al 2022
* Discussion of candidate causal genes that have been identified for AD (Bellenguez et al 2022; Novikova et al 2021)
* Candidate causal AD genes play a central role in efferocytosis, suggesting that it is a core pathway underlying AD pathogenesis
* Figure 1: Loci discovered to be associated with clinically- and proxy-defined Alzheimer’s disease
* Supplementary table 1: summary statistics for AD loci

*Extensions of AD GWAS*

* Polygenic risk scores for predicting AD (Bellenguez et al 2022)
* Using cross-trait PRS, genetic correlation, and Mendelian randomization to investigate shared genetic architecture between traits (Wu et al 2021; Andrews 2020)

***Where next?***

*Rare variants*

* Current GWAS study designs are underpowered to discover rare variants
* Integrating next generation sequencing and GWAS to identify rare variants has identified rare protective variants in *APOE* (Guen et al 2022)
* SVs, indels, decomposed multi-allelic variants, and rare non-coding variation are active areas of investigation (eg ADSP)

*AD genetics in diverse populations*

* A lack of diversity in GWAS of AD coupled with health disparities in AD limits the opportunities for scientific and medical advances to be shared equitably
* Recent findings from AD GWAS in diverse populations highlighting the discovery of novel loci and ancestry-specific effects of *APOE* (Kunkle et al 2021; Shigemizu et al 2021)
* Our current view of AD genetic architecture across ancestries is characterized by a combination of shared loci/genes (sometimes driven by different effect sizes and often driven by different rare alleles), non-shared loci/genes, and mostly shared pathways. This has important and differing implications for risk prediction and therapeutic development
* OMICs data for microglia and other myeloid cells particularly from diverse populations to enable integrative analyses using data from the most relevant cell types and populations.

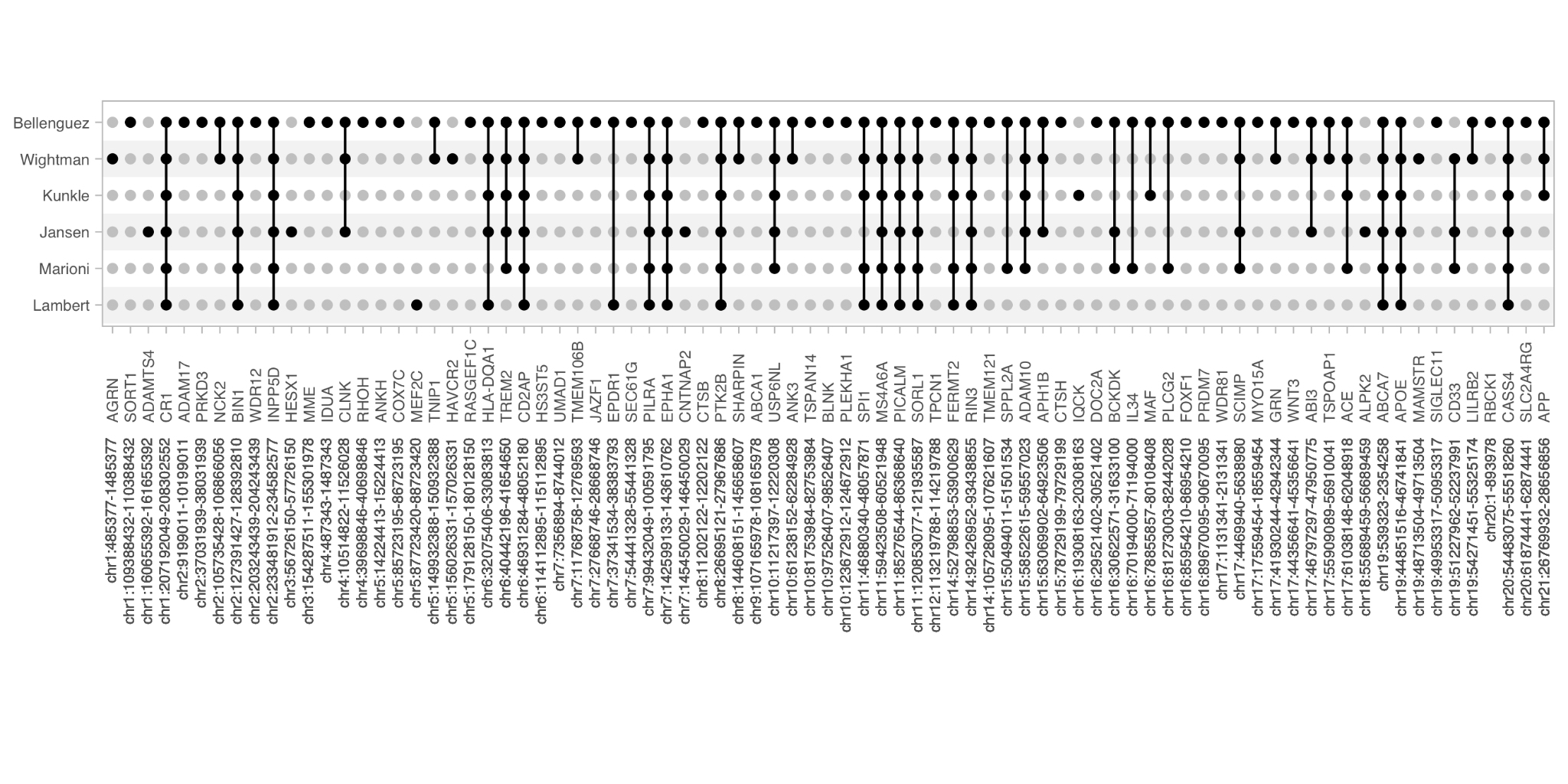
*Genetics of AD plasma biomarkers*

* AD plasma biomarkers for amyloid, tau, and neurodegeneration are approaching clinical use and will offer more accurate diagnosis of cases and controls
* GWAS of plasma biomarkers can offer new insights into genetics of AD biology (Damotte et al 2021, Lord et al 2021)

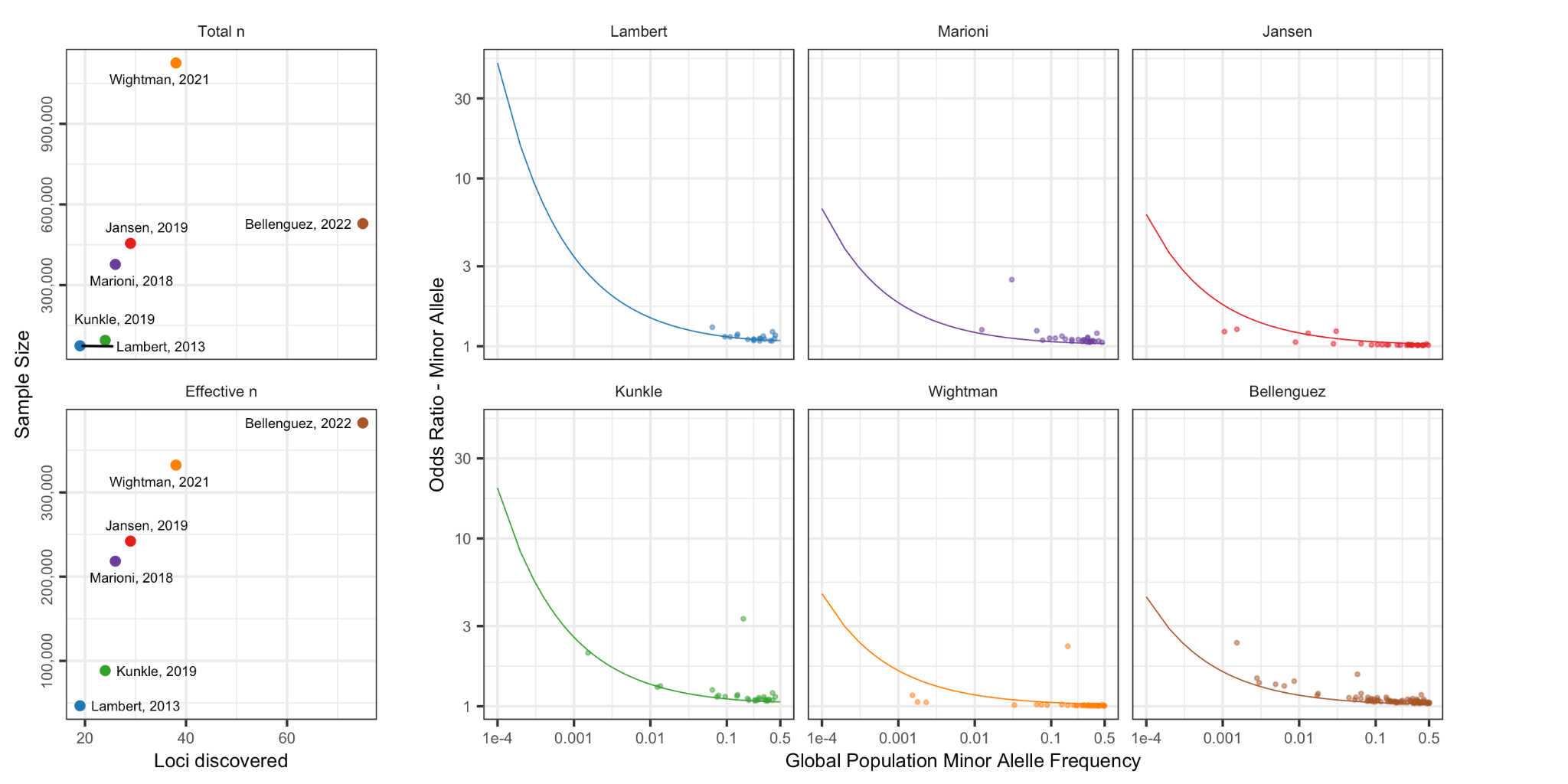
***Conclusion***

* Summary of recent advances in AD GWAS and future research directions
* Figure 3: Genetic architecture of clinically- and proxy-defined Alzheimer’s disease:

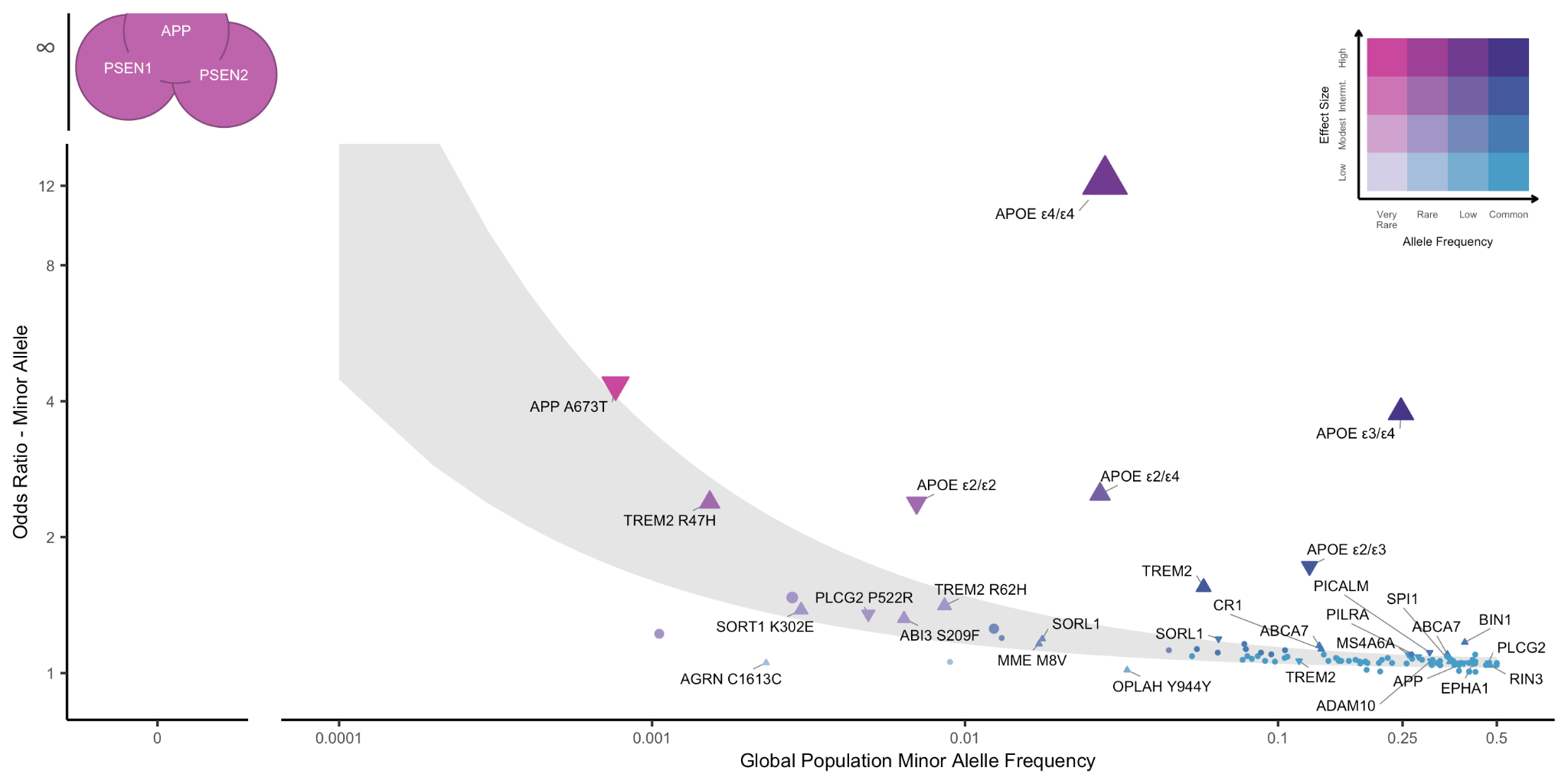
**FIGURES**

****

**Figure 1: Loci discovered to be associated with clinically- and proxy-defined Alzheimer’s disease.** UpSet plot of the intersection of loci discovered for six genome-wide association studies for Alzheimer’s disease highlighting the most replicated findings across studies. Dark circles represent genome-wide significant loci in each study, while light circles represent non-significant loci in each study. Loci were defined by merging overlapping regions +/- 500kb around each lead variant to obtain non-overlapping regions. LD pruning (EUR reference, r2 = 0.1, maf = 0.001) using LDlink snpclip was used to define independent variants in each loci. Loci with candidate causal genes were obtained from the Alzheimer’s Disease Genetics Consortium Gene Verification Committee (https://adsp.niagads.org/index.php/gvc-top-hits-list/).



**Figure 2: Estimated statistical power for Alzheimer’s disease genome-wide association studies.** Imbalances in the proportion of cases and controls in discovery cohorts reduces statistical power, with effective sample size representing the sample size for an equiviently power GWAS with balanced sample (i.e. 50% cases, 50% controls). The number of loci discovered has increased with increasing effective sample size, not total sample size. Additionally, the power to detect rare variants with moderate effect sizes at genome-wide significance increases.

****

**Figure 3: Genetic architecture of clinically- and proxy-defined Alzheimer’s disease:** Genetic variants associated with disease are often conceptualized along two dimensions: minor allele frequency and effect size. Highligly penetrant alleles associated with autosomal dominant Alzheimer’s disease are extremely rare with large effect sizes, while variants discovered by genome wide association studies are mostly common with small effect sizes. Family studies have identified mutations in *APP*, *PSEN1*, and *PSEN2*. Genome-wide association studies of Alzheimer’s disease in non-European populations have identified 82 loci and 103 independent variants (Lambert et al 2013, Marioni et al 2018, Jansen et al 2019, Kunkle et al 2019, Wightman et al 2021, Bellenguez et al 2022) that are mostly common variants with small effect size, however, increasing sample sizes and better imputation have also enabled the detection of rare variants with moderate effect sizes and common variants with low effect sizes. The shaded area represents the 80% power curves for Lambert et al 2013 (top) and Bellenguez et al 2022 (Bottom) indicating that existing GWAS are underpowered for detecting rare variants. Effect sizes are on absolute scale, triangles indicate the direction of effect for candidate causal variants.

**REFERENCES**

Andrews, S. J., Fulton-Howard, B. & Goate, A. Interpretation of risk loci from genome-wide association studies of Alzheimer’s disease. *Lancet Neurology* **19**, 326–335 (2020).

Andrews, S. J. *et al.* Causal Associations Between Modifiable Risk Factors and the Alzheimer’s Phenome. *Ann Neurol* **89**, 54–65 (2021).

Bis, J. C. *et al.* Whole exome sequencing study identifies novel rare and common Alzheimer’s-Associated variants involved in immune response and transcriptional regulation. *Mol Psychiatr* **25**, 1859–1875 (2020).

Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nat Genet* **54**, 412–436 (2022).

Damotte, V. *et al.* Plasma amyloid β levels are driven by genetic variants near APOE, BACE1, APP, PSEN2: A genome-wide association study in over 12,000 non-demented participants. *Alzheimer’s Dementia* (2021) doi:10.1002/alz.12333.

Escott-Price, V. & Hardy, J. Genome-wide association studies for Alzheimer’s disease: bigger is not always better. *Brain Commun* **4**, fcac125 (2022).

Fuente, J. de la, Grotzinger, A. D., Marioni, R. E., Nivard, M. G. & Tucker-Drob, E. M. Integrated analysis of direct and proxy genome wide association studies highlights polygenicity of Alzheimer’s disease outside of the APOE region. *Plos Genet* **18**, e1010208 (2022).

Guen, Y. L. *et al.* Association of Rare APOE Missense Variants V236E and R251G With Risk of Alzheimer Disease. *Jama Neurol* **79**, (2022).

Grotzinger, A. D., Fuente, J. de la, Privé, F., Nivard, M. G. & Tucker-Drob, E. M. Pervasive Downward Bias in Estimates of Liability-Scale Heritability in GWAS Meta-Analysis: A Simple Solution. *Biol Psychiat* (2022) doi:10.1016/j.biopsych.2022.05.029.

Lord, J. et al. A genome-wide association study of plasma phosphorylated tau181. Neurobiol Aging 106, 304.e1-304.e3 (2021).

Kunkle, B. W. *et al.* Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel. *Jama Neurol* **78**, (2021)

Marca-Ysabel, M. V. *et al.* Dissecting the role of Amerindian genetic ancestry and the ApoE ε4 allele on Alzheimer disease in an admixed Peruvian population. *Neurobiol Aging* **101**, 298.e11-298.e15 (2021).

Novikova, G. *et al.* Integration of Alzheimer’s disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat Commun* **12**, 1610 (2021).

Shigemizu, D. *et al.* Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer’s disease risk. *Transl Psychiat* **11**, 151 (2021).

Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease. *Nat Genet* **53**, 1276–1282 (2021).

Wu, H. M., Goate, A. M. & O’Reilly, P. F. Heterogeneous effects of genetic risk for Alzheimer’s disease on the phenome. *Transl Psychiat* **11**, 406 (2021).

**CUTS**

*An end to GWAS of AD?*

* Diagnostic accuracy of clinical AD, diagnostic accuracy of proxy AD, age mismatching, and insufficient data transparency/sharing are valid concerns with AD GWAS and suggests that current AD GWAS may be capturing the genetic architecture of Alzheimer’s disease and related dementias, and not specifically AD. This is an important caveat when interpreting the results from PRS, genetic correlation, and Mendelian randomization(Escott-Price & Hardy 2022).
* With increasing sample size, there have not been corresponding increases in the observed heritability of AD or number of loci discovered, suggesting that larger and larger GWAS will experience diminishing returns (Escott-Price & Hardy 2022).
* However, this is an artifact due to imbalanced study designs
  + Number of loci discovered scales with effective sample size
  + SNP based heritability estimates are biased downwards due to misspecification of heritability models, with heritability of AD higher after accounting for differences in effective sample size and proxy case-control designs (Fuente et al 2022; Grotzinger et al 2022).
* However, there remains substantial missing heritability, increased samples sizes and novel study designs are required to address this discovery gap.
* Figure 2: Estimated statistical power for Alzheimer’s disease genome-wide association studies.

Hi Elanor,

We are fine with dropping the section on methodological issues regarding the relationship between heritability estimates and loci discovered. We still think there should be a discussion regarding heritability of AD and why it seems so low, especially given that clinicians may interpret low heritability to mean that there is little genetic contribution to AD, but we can include this in the section discussing Whightman as they report heritability and focus on broad reasons for why heritability estimates can be low.