Re-alignment Of Whole Genome Sequencing (WGS) Data And Meta-analysis Of Multiple Traits Identified Novel Genetic Risk Loci On The X Chromosome For Alzheimer's Disease

Xu Zhang[1], Aayush Rastogi[1], Samantha Hanus[1], Hari Varshan Dharmendra Mohan Prabu[2], Maitry Ronakbhai Trivedi[2], Jay Shah[2], Yinghua Chen[3], Benjamin P. Readhead[1], Teresa Wu[2], Eric M. Reiman[3], Yi Su[3], Melissa A. Wilson[4], and Qi Wang[1]

1) ASU-Banner Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ, USA; 2) School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA; 3) Banner Alzheimer's Institute, Phoenix, AZ, USA; 4) Center for Genomics and Data Science Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

Introduction

The X chromosome remains understudied in its contributions to Alzheimer's Disease (AD). Genotyping data from WGS are prone to technical artifacts on the sex chromosomes due to the pipelines used for high-throughput processing. The X and Y chromosomes differ from autosomes in several key features, including hemizygosity in males, the almost complete inactivation of one copy of the X chromosome in females, and unique patterns of recombination. All these features require special attention in genetic studies pursuing sex differences in large AD research cohorts, such as X-chromosome-wide association studies (XWAS). Furthermore, previous XWAS have focused on case-control designs, which do not leverage rich clinical and neuropathological traits. Addressing these challenges is critical to identify novel genetic risk loci on the X chromosome.

Project Workflow

We adopted a sex chromosome complement (SCC)-informed WGS data analysis workflow (Figure 1) to specifically address sex chromosome complementary regions and variant calling.

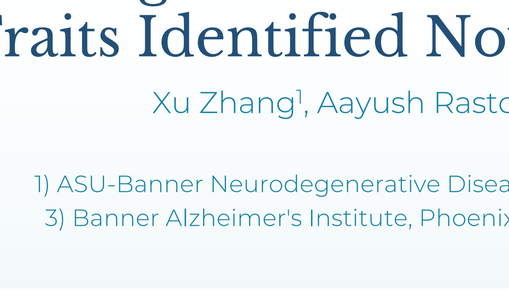


Figure 1. SCC-informed variant calling workflow

The whole workflow (Figure. 2) handles female/male samples specifically to account for different ploidies at the PARs and nonPARs.

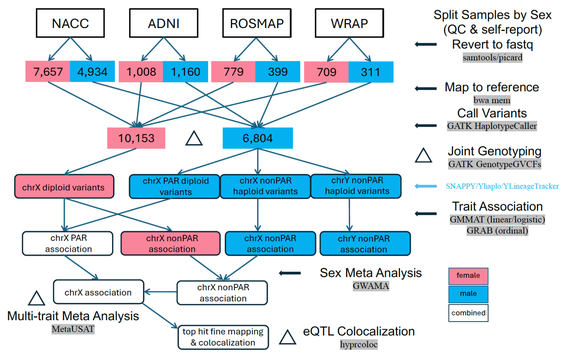


Figure 2. Whole project workflow

Materials and Methods

Raw WGS data were obtained from ADSP[1]. Autosome variant data from ADSP-Release 5 (ADSP-R5)[2] were used to generate sample kinship matrix and population principal components (PCs). Phenotypical data were obtained from ADSP-PHC[3]. Additional phenotypical data were from individual studies[4-7].

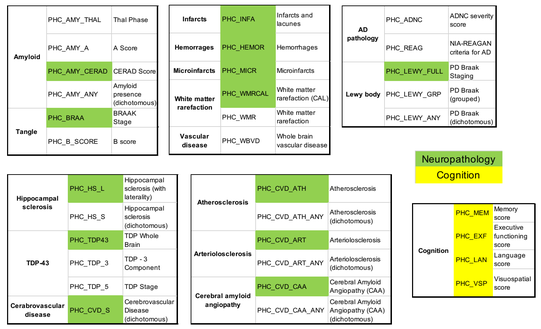
We reprocessed data from 16,957 samples (10,153 female/ 6,804 male), performed quality control by checking coverage on chrX/Y against chr8/19 to identify outliers with sex aneuploidy or poor mapping quality, removed duplicated samples, and kept 15,237 (9,148/6,089) samples for XWAS.

Only common (minor allele frequency (MAF) >= 0.01) single nucleotide polymorphisms (SNPs) were evaluated for association with longitudinal cognitive assessments in four domains and thirteen neuropathological assessments, by the following formula (sex stratified) with kinship matrix included:

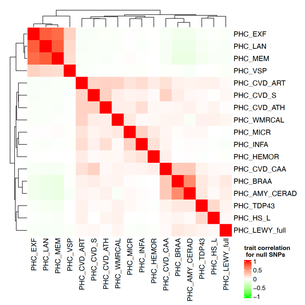
$$\textit{linear mixed effects model:}$$
$$Congition \sim allele\ dosage + age_{visit} + apoe4 + apoe2 + edu + site + 10PC + (1|ID)$$
$$\textit{ordinal or logistic model:}$$
$$Neuropathlogy \sim allele\ dosage + age_{death} + apoe4 + apoe2 + edu + pmi + site + 10PC$$

Framework for Multi-trait Meta Analysis

we first clustered the twenty-nine neuropathological traits and chose the most representative trait for each category, then conducted the univariate XWAS for each trait type, and performed meta analysis for the two sexes (nonPAR), followed by applying multivariate method on the summary stats to get a single omnibus P value for each SNP.

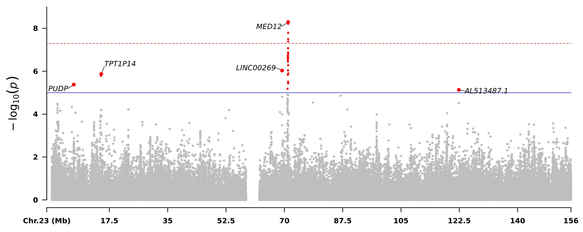


We performed the meta analysis in metaUSAT[8], which takes the XWAS association statistics (e.g., Z-scores) for a SNP across multiple traits, accounts for the correlation among those traits and any sample overlap from null SNPs, and performs a single omnibus test.

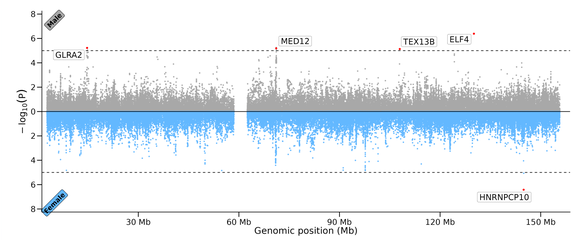


Results

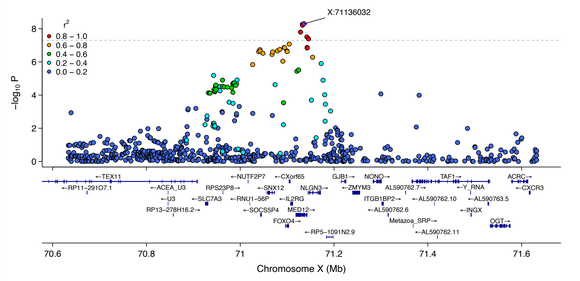
Manhattan plot of the XWAS meta-analysis. One locus exceeded the genome-wide significance cutoff ($P < 5 \times 10^{-8}$).



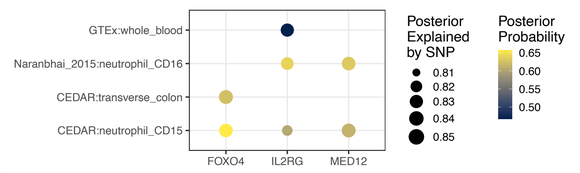
Miami plot of sex-stratified XWAS meta-analysis results, showing concordant lead-SNP effects across sexes and a subset of loci with sex-specific heterogeneity.



Regional association plot with linkage disequilibrium (LD) information, highlighting the lead SNP rs12839587 (X:71,136,032 AD risk AF = 0.276; omnibus $P = 5.10 \times 10^{-9}$). SNPs are colored according to their LD (r^2) with the lead variant.



Fine-mapping and colocalization analyses prioritized X:71,136,032 as a candidate causal SNP shared by multiple traits (PHC_MEM, PHC_EXF, PHC_LAN, and PHC_CVD_S) and colocalized with the expression of several genes, including MED12, IL2RG, and FOXO4, in peripheral immune-related tissues.



Discussion

By applying an SCC-informed WGS analysis workflow and conducting a meta-analysis of AD-related traits, we identified a novel X-chromosome locus associated with AD. To our knowledge, this locus has not been reported in previous large-scale AD case-control XWAS. These findings underscore the importance of addressing the technical challenges associated with sex chromosome analysis in sequencing data and highlight the value of leveraging consortium harmonized quantitative endophenotypes in genetic studies of AD.

Colocalization analyses prioritized X:71,136,032 as a candidate causal variant shared across multiple AD-related traits and linked its effects to the expression of MED12, IL2RG, and FOXO4 in peripheral immune-related tissues. Although these genes have distinct biological functions, they have been studied in the context of X-linked disorders, genomic regulation, cancer immunology, and were implicated in AD previously. Their potential roles in AD pathogenesis warrant further functional investigation.

Conclusions

Scaling our analytical framework to larger cohorts will be important for further elucidating the genetic mechanisms underlying sex differences in AD. Future studies should also evaluate the population specificity of the identified variants and investigate sex-specific effects through sex-stratified analyses and the association of chrY haplotypes with AD related traits to evaluate their potential contribution to AD.

References

1. Beecham GW, et. al. The Alzheimer's Disease Sequencing Project: Study design and sample selection. Neurol Genet. 2017;3(5):e194. doi: 10.1212/NXG.0000000000000194.
2. Carter L, et. al. ADSP Whole Genome Sequencing (WGS) Release 5 data update from Genome Center for Alzheimer's Disease. Alzheimer's & Dementia. 2025;20(S1). doi: 10.1002/alz.087495.
3. Hohman TJ. ADSP Phenotype Harmonization Consortium. Alzheimer's & Dementia. 2023;19(S12). doi: 10.1002/alz.077713.
4. Kukull WA. The National Alzheimer's Coordinating Center (NACC) 1999-2025: Personal history and recollections. Alzheimer's & Dementia. 2025;21(10):e70836. doi: 10.1002/alz.70836.
5. Petersen RC, et. al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology. 2010;74(3):201-9. doi: 10.1212/WNL.0b013e3181cb3e25.
6. Bennett DA, et. al. Religious Orders Study and Rush Memory and Aging Project. Journal of Alzheimer's Disease. 2018;64:S161-S89. doi: 10.3233/JAD-179939.
7. Johnson SC, et. al. The Wisconsin Registry for Alzheimer's Prevention: A review of findings and current directions.. Alzheimer's & Dement (Amst). 2018;10:130-42. doi: 10.1016/j.dadm.2017.11.007.
8. Ray D, et. al. Methods for meta-analysis of multiple traits using GWAS summary statistics. Genet Epidemiol. 2018;42(2):134-45. doi: 10.1002/gepi.22105.

Acknowledgements

NIH National Institute on Aging
ADSP PHC Alzheimer's Disease Sequencing Project
adsp
NIAGADS Data Sharing Service
ASU Research Computing
Arizona Alzheimer's Consortium
ASU Biodesign Institute Arizona State University