

Fine-mapping of Candidate Causal Variants for Alzheimer Disease in the Midwestern Amish

Yeunjoo E. Song^{1,2}, Penelope Benchek¹, Michelle Grunin¹, Renee A. Laux¹, Sarada L. Fuzzell¹, Sherri D. Hochstetler¹, Dawn Miller¹, Kristy Miskimen¹, Audrey Lynn¹, Yining Liu¹, Noel Moore¹, Dana Jian¹, Alex V. Gulyayev³, Daniel A. Dorfman³, Larry D. Adams³, Laura J. Caywood⁴, Jason E. Clouse³, Sharlene D. Herington³, Jeffery M. Vance^{3,4}, Michael L. Cuccaro^{3,4}, Paula K.Ogrocki^{5,6}, Alan J. Lerner^{5,6}, William K. Scott^{3,4}, Margaret A. Pericak-Vance^{3,4}, William S. Bush^{1,2}, Jonathan L. Haines^{1,2}

¹Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; ²Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA; ³John P Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; ⁴Dr. John T Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA, CA; ⁵Brain Health and Memory Center, University Hospital, Cleveland, OH, USA ; ⁶Department of Neurology, Case Western Reserve University, Cleveland, OH, USA



INTRODUCTION

There are numerous genetic loci for Alzheimer Disease (AD), yet resolving these signals into specific causal variants remains a challenge. Fine-mapping aims to identify the underlying potentially causal variants. We examined whole genome sequencing data and performed fine-mapping in the Midwestern U.S. Amish, a founder population which is better powered for discovery of previously unknown rare alleles that have risen to higher frequency due to genetic drift, as part of the Collaborative Amish Aging and Memory Project (CAAMP).

METHODS

Study population

- Amish families in Holmes County, OH and Adams, Elkhart and LaGrange Counties, IN (Figure 1)
- Participants were identified from published community directories, referral from other community members or due to close relationship with other participants. assays²

Phenotype definition

- The AD status was assigned via consensus review of medical history and neuropsychological testing.
- Covariates: age, sex, 2 principal components.

Genome-wide association study

- Whole genome sequencing data from Illumina NovaSeq 6000 platform
- Only bi-allelic common variants (MAF >= 0.05) were included after extensive QC.
- A linear mixed model was used for the association test, adjusting for covariates and relatedness.

Fine-mapping

- In-sample LD matrices were constructed using the common variants (>5%).
- For each GWAS locus (± 1 MB from the most associated variant), used the SuSiE² Bayesian framework to prioritize variants based on Posterior Inclusion Probabilities (PIP>0.5) and Combined Annotation Dependent Depletion (CADD)² v1.7 scores.
- Prioritized variants were further integrated with GTEx³ v10 eQTL data and epigenomic annotations to identify putative regulatory mechanisms utilizing HaploReg⁴ v4.2 and AlphaGenome.

ACKNOWLEDGEMENTS

- We would like to thank the Amish communities of Holmes County, Ohio, and specifically the patients and families for their time and participation in this study.
- This work was funded by NIH-NIA (AG058066, AG058654).



RESULTS

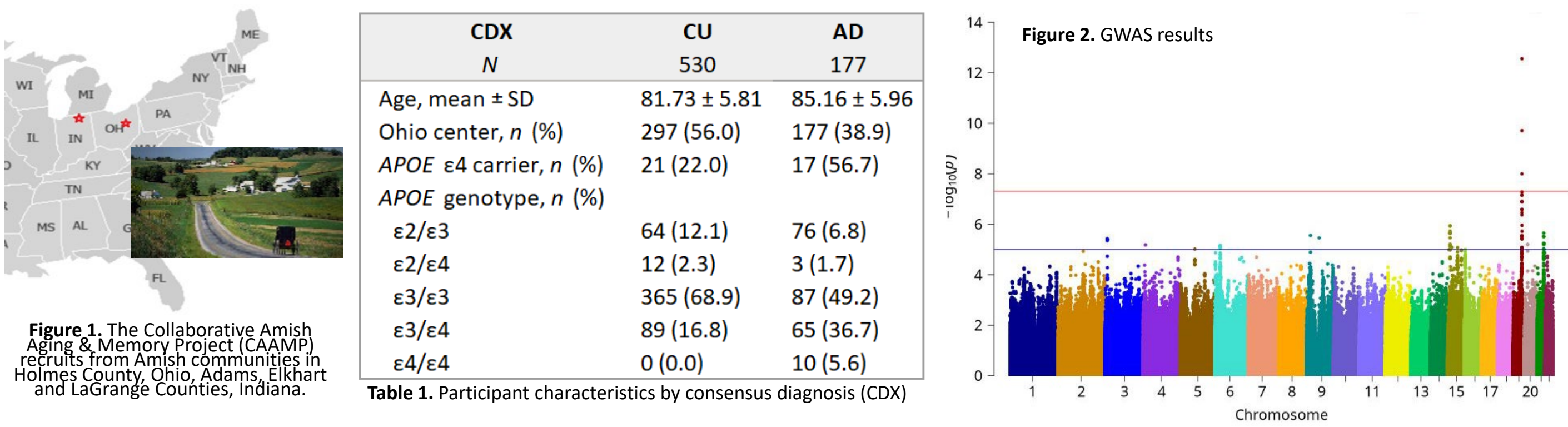


Figure 1. The Collaborative Amish Aging & Memory Project (CAAMP) recruits from Amish communities in Holmes County, Ohio, Adams, Elkhart and LaGrange Counties, Indiana.

CDX	CU	AD
N	530	177
Age, mean $\pm$ SD	81.73 $\pm$ 5.81	85.16 $\pm$ 5.96
Ohio center, n (%)	297 (56.0)	177 (38.9)
APOE $\epsilon 4$ carrier, n (%)	21 (22.0)	17 (56.7)
APOE genotype, n (%)		
$\epsilon 2/\epsilon 3$	64 (12.1)	76 (6.8)
$\epsilon 2/\epsilon 4$	12 (2.3)	3 (1.7)
$\epsilon 3/\epsilon 3$	365 (68.9)	87 (49.2)
$\epsilon 3/\epsilon 4$	89 (16.8)	65 (36.7)
$\epsilon 4/\epsilon 4$	0 (0.0)	10 (5.6)

Table 1. Participant characteristics by consensus diagnosis (CDX)

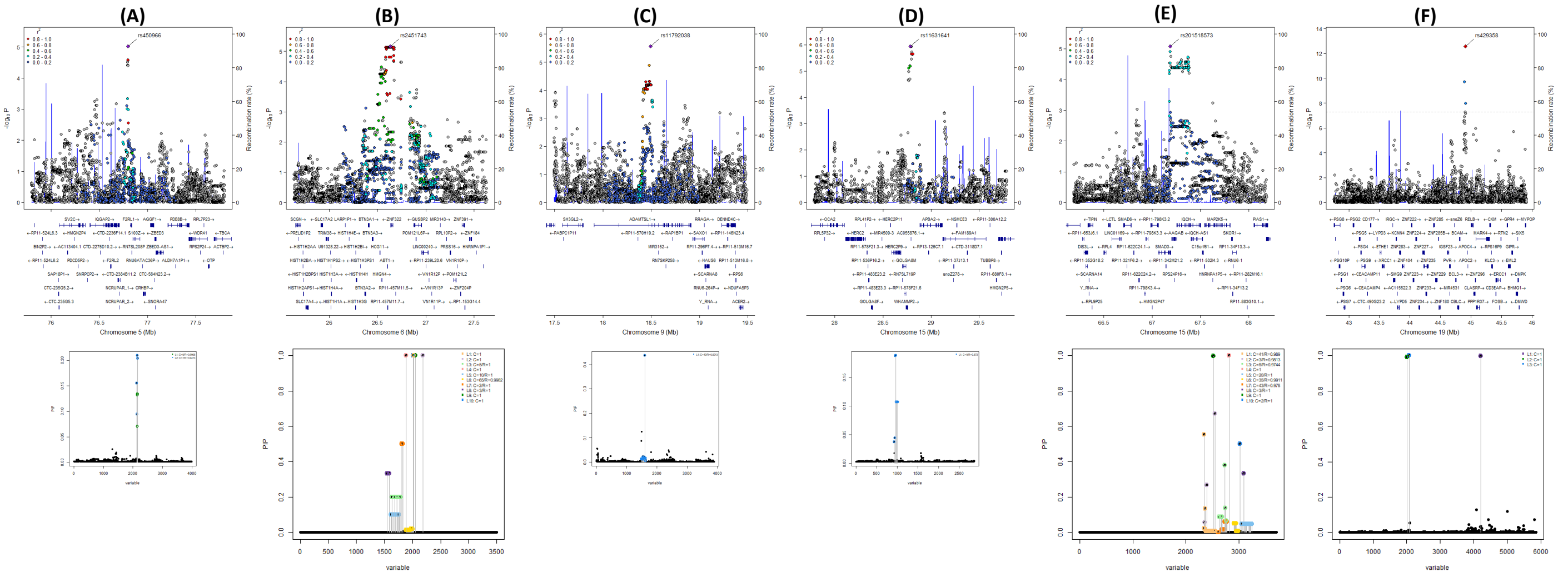


Figure 3. Regional Manhattan plots for loci fine-mapped. (A) F2RL1, (B) ZNF322, (C) ADAMTSL1, (D) PDCD6IPP2, (E) SMAD3, (F) SMG3

GWAS Lead							MAF			SuSiE Credible Set				CADD GRCh38-v1.7		GTEx v10 p < 0.05 in brain		ENSEMBL		VEP Consequence	
Variant	Chrom	Position	Ref	Alt	p-value	CAAMP	dbSNP	gnomAD		Position	Ref	Alt	distance	pip	PHRED	eQTLs	sQTLs	GeneName		priority selected by potential impact	
rs9871377	3	11895327	G	A	3.89E-06	0.1553	0.3165	0.2553										FANCD2P2-NUP210P2		DOWNSTREAM, UPSTREAM	
rs10002690	4	11420321	C	T	6.57E-06	0.1783	0.2240	0.1804										HS3ST1		INTRONIC	
rs450966	5	76798023	G	T	9.61E-06	0.4011	0.4211	0.4173										between LOC124901199 & F2RL1		INTERGENIC	
rs2451743	6	26632698	G	T	6.89E-06	0.3869	0.523	0.4566										ZNF322		DOWNSTREAM	
																		LOC124901288		INTRONIC	
																		BTN3A2		INTRONIC	
																		BTN3A2		INTRONIC	
																		BTN3A2		INTRONIC	
																		BTN3A2		INTRONIC	
																		BTN3A2		INTRONIC	
																		ADAMTSL1		INTRONIC	
																		DUX4L50:FRG1JP:MIR4477B		DOWNSTREAM, UPSTREAM	
																		PDCD6IPP2		UPSTREAM	
																		LOC105370777		NONCODING_CHANGE, INTRONIC	
																		SMAD3		NONCODING_CHANGE, INTRONIC	
																		between SMASR & SMAD3-DT		INTERGENIC	
																		SMAD3-DT		INTRONIC	
																		LOC102723493		INTRONIC	
																		SMAD3		INTRONIC	
																		AAGAB		INTRONIC	
																		RBFOX1		NONCODING_CHANGE, UPSTREAM	
																		SMG9		NONCODING_CHANGE, UPSTREAM, INTRONIC	
																		PSG4		INTRONIC	
																		PSG4		INTRONIC	
																		ZNF45-AS1		NONCODING_CHANGE, UPSTREAM	

Table 2. CADD and GTEx eQTL and sQTL results

Demographic analysis (Table 1)

- After extensive QC, 1,138 samples were available to estimate overall allele frequency.
- 707 participants had both AD status and WGS genotype.
- There were 572 sibships of size 1 to 11 (mean size 1.7 ± 1.1).

Genome-wide association study (Figure 2)

- Significant ($p=2.8 \times 10^{-13}$) associations in APOE region
- 13 other loci had suggestive associations ($p < 1 \times 10^{-5}$)

Fine-mapping

- SuSiE identified at least one 95% credible set in 6 loci (F2RL1, ZNF322, ADAMTSL1, PDCD6IPP2, SMAD3, SMG3) (Figure 4 upper row)
- 3 loci (ZNF322, SMAD3, SMG3) contained at least one high-confidence variant with a PIP > 0.5 (Figure 3 lower row)
- Among all 14 prioritized variants from these 3 loci, 9 had eQTL expression in brain tissues and 5 variants overlap with active regulatory elements in human brain (Table 2)

CONCLUSION

We narrowed 13 broad signals down to 3 high-confidence defined sets of prioritized variants with supporting functional evidence of their role as causal drivers of AD. All 3 loci have been reported with some links to AD via immune system, synaptic function, intracellular membrane trafficking, and neuronal damage, though more studies are needed to fully understand their precise roles in AD. Use of this founder population further illuminates the genetic architecture of AD as we note that the mean frequency of effect alleles is increased 1% across all 14 prioritized variants and 22% in the 5 variants in ZNF322 loci compared to TOPMED.

REFERENCES

- Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. J R Stat Soc Series B Stat Methodol. 2020 Dec;82(5):1273-1300. doi: 10.1111/rssb.12388. Epub 2020 Jul 10. PMID: 37220626; PMCID: PMC10201948.
- Schubach M, Maass T, Nazaretyan L, Röner S, Kircher M. CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. Nucleic Acids Res. 2024 Jan 5;52(D1):D1143-D1154. doi: 10.1093/nar/gkad989. PMID: 38183205; PMCID: PMC10767851.
- GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580-5. doi: 10.1038/ng.2653. PMID: 23715323; PMCID: PMC4010069..
- Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 2016 Jan 4;44(D1):D877-81. doi: 10.1093/nar/gkv1340. Epub 2015 Dec 10. PMID: 26657631; PMCID: PMC4702929.