Development/Plasticity/Repair

Transcription Factor-Wide Association Studies to Identify Functional SNPs in Alzheimer's Disease

Jessica Dunn,[1*] Cedric Moore,[1*] Nam-Shik Kim,[2*] Tianshun Gao,[3] Zhiqiang Cheng,[1] Peng Jin,[4] Guo-li Ming,[2] Jiang Qian,[3] Yijing Su,[2,5] Hongjun Song,[2] and Heng Zhu[1]

[1]Department of Pharmacology, Johns Hopkins University, Baltimore, Maryland 21205, [2]Department of Neuroscience and Mahoney Institute for Neurosciences, University of Pennsylvania, Philadelphia, Pennsylvania 19104, [3]Department of Ophthalmology, Johns Hopkins University, Baltimore, Maryland 21205, [4]Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, and [5]Department of Oral Medicine, School of Dental Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Alzheimer's disease (AD) is a progressive neurodegenerative disorder with profound global impact. While genome-wide association studies (GWAS) have revealed genomic variants linked to AD, their translational impact has been limited due to challenges in interpreting the identified genetic associations. To address this challenge, we have devised a novel approach termed transcription factor-wide association studies (TF-WAS). By integrating the GWAS, expression quantitative trait loci, and transcriptome analyses, we selected 30 AD single nucleotide polymorphisms (SNPs) in noncoding regions that are likely to be functional. Using human transcription factor (TF) microarrays, we have identified 90 allele-specific TF interactions with 53 unique TFs. We then focused on several interactions involving SMAD4 and further validated them using electrophoretic mobility shift assay, luciferase, and chromatin immunoprecipitation on engineered genetic backgrounds (female cells). This approach holds promise for unraveling the intricacies of not just AD, but any complex disease with available GWAS data, providing insight into underlying molecular mechanisms and clues toward potential therapeutic targets.

Key words: Alzheimer's disease; SNPs; TF-WAS

Significance Statement

We introduce a powerful platform for better understanding the genetic contribution of Alzheimer's disease (AD) and other complex diseases. Through genome-wide association studies (GWAS), many statistically significant single nucleotide polymorphisms (SNPs) associated with AD have been identified, but their functionality remains unknown. By screening >85% of human proteome transcription factors and cofactors for allele-specific binding preferences with GWAS SNPs, we can comprehensively elucidate the functionality of these SNPs in disease etiology. Using this strategy, we have identified and validated several allele-specific interactions with AD-associated GWAS SNPs that have potential implications in processes relevant to AD. By leveraging available GWAS data, we can identify functional SNPs not just in AD but in essentially all other complex diseases.

Introduction

Alzheimer's disease (AD) is a devastating neurodegenerative condition responsible for 60–70% of dementia cases worldwide (Song et al., 2019). Despite its profound impact, there are currently no treatments available to halt or reverse its progression, primarily because our understanding of its intricate molecular mechanisms remains incomplete. Genome-wide association studies (GWAS) and familial linkage analysis have successfully identified genetic loci that confer risk for AD, including *APP*, *APOE*, *PSEN1*, and *PSEN2*, among many others (Strittmatter et al., 1993; Bekris et al., 2010). This information has provided critical insight into some of the biological processes involved in the disease, such as cholesterol and lipid metabolism, immune responses, and endosomal vesicle cycling (Van Cauwenberghe

Received Sept. 20, 2024; revised Nov. 1, 2024; accepted Nov. 8, 2024.

Author contributions: G.M., J.Q., H.S., and H.Z. designed research; J.D., C.M., N.-S.K., and Y.S. performed research; P.J. contributed unpublished reagents/analytic tools; T.G. and Z.C. analyzed data; J.D., Y.S., H.S., and H.Z. wrote the paper.

This work was supported by National Institutes of Health grants R01AG061852 (H.Z., H.S., and J.Q.), R01MH122451 (H.Z.), R01EY030475 (H.Z. and J.Q.), P50AA027054 (H.Z.), R35NS097370 (G.M.), R35NS116843, and RF1AG079557 (H.S.) and by the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (G.M.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

*J.D., C.M., and N.-S.K. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Heng Zhu at hzhu4@jhmi.edu, Hongjun Song at shongjun@pennmedicine.upenn.edu, Yijing Su at yijingsu@upenn.edu, or Jiang Qian at qianjiang007@gmail.com.

https://doi.org/10.1523/JNEUROSCI.1800-24.2024

et al., 2016). Despite this progress, the functional basis for many of these associations remains elusive.

In the post-GWAS era, identifying causal variants among numerous significant variants remains challenging, particularly in linkage disequilibrium regions. Expression quantitative trait loci (eQTLs) have been useful in this pursuit by associating genetic variants with variation in gene expression levels. For example, based on whole-genome single nucleotide polymorphism (SNP) genotyping and whole-transcriptome expression profiling in cortical samples, many significant associations between inherited variants and transcripts expressed in the brain were identified (Myers et al., 2007; Webster et al., 2009). While certainly successful in identifying SNP-transcript connections, eQTL analysis is limited because it must be performed in physiologically relevant cells and tissues, and it neglects RNA-level processes beyond gene expression, such as RNA splicing, degradation, and transport.

Another challenge has been uncovering the mechanisms through which these variants are able to confer disease risk and influence pathology. This is underscored by the fact that many GWAS variants are SNPs located in noncoding genomic regions, further obfuscating their functionality. This phenomenon suggests functional SNPs are located within *cis*-regulatory elements, thus affecting phenotype on the level of transcriptional regulation. There is substantial evidence in the literature that SNPs within regulatory regions can alter canonical transcription factor binding and consequently impact gene expression (Maurano et al., 2012; Li et al., 2019).

To this end, ChIP-seq is a useful tool for connecting GWAS SNPs with transcription factor binding sites (Landt et al., 2012; Reddy et al., 2012), but this method requires a priori knowledge of relevant TFs and an extensive collection of ChIP-grade antibodies. Quantitative mass spectrometry has also been utilized to identify differential transcription factor binding to GWAS SNPs (Butter et al., 2012); however this approach is difficult to scale up and often suffers from false negatives due to the low binding affinity of typical TF–DNA interactions. Others have predicted allele-specific binding sites using genomic footprints obtained via mapping DNase I hypersensitivity sites or ATAC-seq (Neph et al., 2012; Buenrostro et al., 2013), but this approach can only be utilized for a limited number of TFs with known consensus motifs, and not all TFs cause genomic footprints due to weak or transient binding interactions. SNP-SELEX represents an alternative strategy for efficiently detecting protein–DNA interactions in a high-throughput manner. Nonetheless, it has limited coverage of SNPs within the human genome, and the method exhibits a subtle bias toward risk-associated loci (Yan et al., 2021).

To overcome these challenges, we introduce transcription factor-wide association studies (TF-WAS). In this method, we employ human transcription factor protein microarrays (TF arrays) containing ∼1,700 full-length purified transcription factors (TFs) and cofactors (1,265 unique factors, greater than 85% coverage of those in the human proteome) spotted in duplicate to screen GWAS SNPs for differential TF binding (Hu et al., 2009). In doing so, we can discern allele-specific binding interactions in an unbiased and high-throughput manner. This approach, coupled with bioinformatic analyses and orthogonal validation assays, can shed light on SNP functionality and TFs involved, thus providing clues to underlying mechanisms in AD. By harnessing available GWAS data, TF-WAS can enable valuable insights into any complex disease or trait, offering a comprehensive understanding of their genetic underpinnings and potential therapeutic targets.

Materials and Methods

Bioinformatic SNP selection. We obtained 2,750 AD-associated SNPs from GWAS (v1.0.2) database (Sollis et al., 2023). We have two strategies to enrich the SNPs that are likely to cause gene expression changes. We selected (1) the SNPs located in the enhancers and (2) the SNPs that were tested to affect gene expression through expression quantitative trait locus (eQTL) studies. Approximately 400,000 enhancers were obtained from EnhancerAtlas (Gao and Qian 2020) and SEA 3.0 (Chen et al., 2020) databases. Meanwhile, ∼290,000 brain-/nerve-related eQTLs were identified from the eQTL data portal GTEx (v8; Strober et al., 2020). By the overlapping analysis between AD-associated SNPs and these enhancers/eQTLs, we obtained a total of 418 AD-associated SNPs that are likely to affect gene expression.

Protein microarray fabrication. Protein microarrays were fabricated as described previously (Hu et al., 2009). From our previously generated collection of ∼21,000 full-length human ORFs expressible as N-terminal GST-His_6 fusion proteins (Jeong et al., 2012), ∼1,700 transcription factor and cofactor proteins were selected as a subcollection for printing on the TF arrays. Each protein was expressed in 8 ml of yeast culture in a 96-well format, with protein expression induced for 6 h by addition of galactose in glucose-free media. Yeast cells were lysed mechanically in lysis buffer (50 mM Tris-Cl at pH 7.5, containing 100 mM NaCl, 1 mM EGTA, 1 mM PMSF, 10% glycerol, 0.1% Triton X-100, 0.1% beta-mercaptoethanol, and Roche protease inhibitor tablet). Protein was purified from the lysates though binding with glutathione sepharose beads (GE HealthCare, GE17-0756-04) overnight at 4°C. After incubation overnight, the beads were washed three times with wash buffer I (50 mM Tris-Cl at pH 7.5, containing 500 mM NaCl, 1 mM EGTA, 1 mM PMSF, 10% glycerol, 0.1% Triton X-100, and 0.1% beta-mercaptoethanol) and three times with wash buffer II (50 mM HEPES at pH 8.0, containing 500 mM NaCl, 1 mM EGTA, 1 mM PMSF, 10% glycerol, and 0.1% beta-mercaptoethanol) to remove any nonspecifically bound proteins from the beads and equilibrate to the elution buffer, respectively. Proteins were eluted from the beads using 80 µl of elution buffer (50 mM HEPES at pH 8.0, containing 100 mM NaCl, 40 mM reduced glutathione, pH 8.0, 30% glycerol, and 0.1% beta-mercaptoethanol). Purified proteins were rearrayed into a 384-well format. Proteins were printed in duplicate at 200 pL per spot on PATH Protein Microarray slides (Grace Bio-Labs, 805020). Quality was ensured by probing with Anti-GST antibody to verify adequate protein loading.

Fluorescent and biotinylated DNA probe generation. Each of the selected SNPs for this study were synthesized (Integrated DNA Technologies) with 15 bp flanking contextual sequences on either side, and an additional common modified T7 priming sequence (5′-ACCCT ATAGTGAGTGCTATTA – 3′) at the 3′ end. Cy3, Cy5, and biotin primers complementary to the T7 priming site were also synthesized (Integrated DNA Technologies). Fluorescent/biotin primers were incubated at a 1:1 molar ratio in 1× NEB II Buffer (NEB). Mixtures were boiled at 95°C for 10 min and then cooled slowly to room temperature to allow for annealing to occur. Once the mixture was fully cooled, 3 U of Klenow Large Fragment 3′-5′ exo- (NEB M0210) and dNTPs (final concentration 1.5 mM) was added to each reaction. Reactions were incubated at 37°C for 20 min to generate double-stranded probes.

Dye-swap protein microarray screening. SNPs are probed to the TF arrays in pairs as a competition assay, with one allele labeled with Cy3 and the other labeled with Cy5. For each pair, the arrays are performed in duplicate, with the allele colors swapped. For example, on one array the Cy3 risk and Cy5 nonrisk are probed, while the Cy5 risk and Cy3 nonrisk are probed to another array. Prior to the competition assay, TF arrays are blocked with blocking buffer (25 mM HEPES at pH 7.5, containing 50 mM potassium glutamate, 8 mM magnesium acetate, 3 mM DTT, 10% glycerol, 0.1% Triton X-100, and 3% BSA) for 3 h at 4°C. After blocking, Cy5 and Cy3 alleles are mixed in 1× hybridization buffer (10 mM Tris-Cl at pH 8.0, containing 50 mM potassium chloride, 1 mM magnesium chloride, 1 mM DTT, 5% glycerol, 10 mM zinc chloride, and 3 mg/ml BSA) to a final concentration of 40 nM of each allele. Blocking

buffer is removed from the arrays, and then the mixed hybridization reaction is added and incubated overnight at 4°C. After overnight incubation, arrays are washed once with 3 ml of ice-cold TBST for 5 min, briefly rinsed with water, and then dried via centrifugation. Arrays are scanned in both the Cy5 (635 nm) and Cy3 (532 nm) channels separately at 1,000 PMT using the GenePix 4000B scanner (Molecular Devices). GenePix Pro 7 software determines the foreground and local background intensities for each detected fluorescent spot at every location on the corresponding alignment grid (GAL file). For each image, a .TIFF file and a .GPR file are generated and saved for analysis.

Protein microarray analysis. For each spot on the alignment grid, the background and foreground intensities were determined by the GenePix Pro 7 software. Using these values, ratiometric binding analysis was performed as follows for each spot on the TF arrays using RStudio:

$$R* = \mathrm{Log}_2\sqrt{\frac{\text{Cy3Non-Rish} * \text{Cy5 Non-Rish}}{\text{Cy3Rish} * \text{Cy5Risk}}}. \quad (1)$$

Ratiometric binding values that are highly positive or highly negative indicate strong preferential binding for the nonrisk or risk allele of the given SNP, respectively. Values above +/−1 were considered significant differential binding events for the purposes of this study. Raw GPR files and processed data obtained in this study can be accessed through the GEO database with access number GSE280753.

Electrophoretic mobility shift assay. Cy5 probes for the alleles of each SNP were generated for detection of interactions with purified SMAD4 protein. We mixed 10 nM Cy5 SNP allele, 1 μM cold unlabeled competitor allele, and purified SMAD4 protein in 1× hybridization buffer (10 mM Tris-Cl at pH 8.0, containing 50 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 10 μM $ZnCl_2$, and 3 mg/ml BSA). Reactions were incubated for 1 h at room temperature and then at 4°C overnight. After incubation, reactions mixtures were analyzed using gel electrophoresis with a 5% TBE PAGE gel in cold 1× TBE running buffer at 100 V for 1 h. Gels were visualized on Odyssey CLx (LI-COR Biosciences) using the Cy5 channel.

OCTET. Kinetic measurements were obtained using OCTET QK (Molecular Devices) with High Precision Streptavidin (SAX) biosensors. Biotinylated probes were generated for each of the SNP alleles to be tested. Prior to kinetic analysis, the biosensors were incubated in 1× hybridization buffer (10 mM Tris-Cl at pH 8.0, containing 50 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 10 μM $ZnCl_2$, 3 mg/ml BSA, and 0.1 mg/ml salmon sperm DNA) for 10 min to hydrate the sensors and equilibrate to the buffer. A baseline measurement was taken in 1× hybridization buffer for 120 s, and then biosensors were moved to a well containing 500 nM biotinylated DNA in 1× hybridization buffer for 600 s to allow for DNA loading. After DNA was loaded onto the biosensors, another baseline measurement was taken for 120 s in 1× hybridization buffer. Next, the biosensors were immersed in solutions containing a range of concentrations of purified SMAD4 protein (195, 172, 144, 115, 86, and 0 nM) and allowed to incubate until a binding equilibrium was reached. Finally, the biosensors were immersed back into hybridization buffer to allow for the dissociation of SMAD4 from the sensors. Binding curves and kinetic values were generated by ForteBio Data Analysis software.

Luciferase assay. Luciferase constructs were generated for each of the selected SNP allele pairs. Each allele was synthesized (Integrated DNA Technologies) with four repeats of the SNP (with 7 bp flanking contextual sequence on either side) and *Nhe*I and *Hin*DIII restriction sites on the ends. These sequences were cloned upstream of the luc2P luciferase reporter gene in pGL4.32 (Promega), replacing the NF-kB response element which served to drive luciferase reporter gene expression. A clone of the transcription factor SMAD4 (IOH3638) within a pDONR221 entry vector was obtained from the ChemCORE at Johns Hopkins University and gateway cloned into the CMV driven pcDNA DEST40 to act as an overexpression vector.

For the luciferase assay, HEK293t cells (female) were plated in 24-well dishes and allowed to grow until ∼70–90% confluent. Cells were then transfected with 62.5 ng of SMAD4 overexpression vector, 62.5 ng of SNP allele luciferase vector, and 6.25 ng of Renilla control vector using Lipofectamine 3000 (Thermo Fisher Scientific) according to manufacturer's instructions. Luciferase assays were performed as described by the Dual Luciferase Reporter Assay system from Promega (E1910). Following 2 d of incubation, cells were passively lysed using Passive Lysis Buffer at room temperature for 15 min. Lysates were transferred to a 96-well plate to facilitate luminescence signal reading. LARII reagent was added to each of the wells, and then luminescence measurements were recorded for the luciferase signals for each well. Following this measurement, Renilla substrate was added to each well, and luminescence readings were recorded again to determine the Renilla signal for each well. Raw luciferase signals for each sample were normalized to Renilla control signals to account for well-to-well variability in cell count. Sample signals were further normalized to no TF control wells to account for any background interaction with the SNP luciferase vectors. Error is reported as the standard error across at least two experiments, each experiment consisting of three replicates.

CRISPR/Cas9 generation of PDE1A *and* CACNA2D3 *point mutations in HEK293 cells.* CRISPR/Cas9 system was used to generate point mutations at the loci of *PDE1A* and *CACNA2D3* in HEK 293 cell lines (female) as previously described (Ran et al., 2013). The donor templates for *PDE1A* or *CACNA2D3* point mutations were designed to harbor point mutation in defined loci and synthesized as single-strand oligo donors (ssODNs) at IDT (Integrated DNA Technologies) without cloning. To achieve high HDR (homologous DNA repair) efficiencies, ssODNs contain flanking sequences of 40 bp on each side that are homologous to the target region. For the gRNA design, we utilized an online CRISPR Design Tool (https://benchling.com) and selected the 20 nt guide sequence within or near by the point mutation sites (Table 1). Then the designed gRNAs were cloned into the gRNA Cloning Vector (Addgene, plasmid #41824). The functionality and efficacy of designed gRNAs were assessed by SURVEYOR nuclease assay.

HEK 293 cells were plated in 100 mm dishes 1 d before transfection and transfected with Cas9 expression vector, pSpCas9(BB)-2A-Puro (px459; Addgene plasmid # 48139), cloned gRNA expression vector, and ssODN using Lipofectamine 2000 according to manufacturer's instruction (Thermo Fisher Scientific, catalog #11668019). Transfected

Table 1. Guide sequences for gRNA design and screening primer sets

gRNA sequence	*PDE1A*	5′ TTAGCTTTTGAAACTCACTTAGG 3′
	CACNA2D3	5′ ACACACCCTTCCTCGAGTCAAGG 3′
ssODN sequence	*PDE1A*	5′TTAAAGGAAAGCCAAAATGTTTGCATCATTTTACAGATTATTTTTTCCAAGTTTTAGCTTTTGAAACTCACTTAGGTAGACAGTTAAAATCAATATTCTACAGTTAATTTGTCCTATTTATAAA GATAAAGAATATGTGAATTTTTGGCATTTCTCTCACA 3′
	CACNA2D3	5′GAACAAGGAAGGGAAGCATTTTGTGACACACACACACACACACACACACACACACACACACACCCTTCCTCGAGACAAGGAGAAACTGAGCCTCCCAATTTACATTTTGAAGGT ATAGGATGGGAGGAATGAATGGAGAGAAAATGTAGATTTAT 3′
ScrF primer	*PDE1A*	5′ CATTCAGGCACAGAAATGGA 3′
	CACNA2D3	5′ TCCTAGAACACATGGCCAGA 3′
ScrR primer	*PDE1A*	5′ GGATGAAAAATGGGGTGAAA 3′
	CACNA2D3	5′ CCAAGCTCTTACCCAGGGAA 3′

cells were incubated for 48 h after transfection then treated with puromycin (3 μg/ml) for 7 d. Surviving colonies were picked from 96-well plates and expanded until confluent. Then 10% of cells were cultured for further use and 90% of cells were lysed with DirectPCR Lysis Reagent (cell) (Viagen Biotech, catalog #301-C) for 6 h at 56°C for PCR screening with designed screening primer sets (Table 1). For primary screening, PCR products of *PDE1A* and *CACNA2D3* were cut with AccI and HinfI, respectively, and then selected clones were performed with Sanger sequencing or MiSeq for final confirmation.

Results

Bioinformatic selection of Alzheimer's disease SNPs for probing to the TF arrays

In this study, we first procured a dataset of 1,046 SNPs associated with AD from the GWAS database (v1.0.2; Sollis et al., 2023). To refine our selection of GWAS-identified SNPs to those that have the potential to influence gene expression, particularly in AD, we employed overlapping analysis with two additional datasets, as visually depicted in Figure 1*A*,*B*. The first dataset contains a selection of ∼290,000 expression quantitative trait loci (eQTLs) relevant to the brain and nervous system, sourced from the GTEx Portal (v8; Strober et al., 2020). These loci have demonstrated the ability to modulate gene expression in relevant cell types, and thus have increased potential to play a role in the etiology of AD. The second dataset consists of ∼400,000 enhancer elements obtained from EnhancerAtlas (Gao and Qian 2020) and SEA 3.0 (Chen et al., 2020) databases. As mentioned previously, functional SNPs have been hypothesized to reside within *cis*-regulatory elements, where they can exert regulatory control over nearby genes by influencing transcription or other essential cellular processes. Using multimodality analysis with these three datasets, we were able to achieve a list of 418 prioritized SNPs associated with AD with increased potential for functionality (Table 2). In this set of 418 SNPs, 200 were annotated to be within enhancers, 155 were associated with changes on transcription of a target gene expressed in the brain or nervous system via eQTL, and 63 had a combination of the two annotations (Fig. 1*A*). The identified SNPs span various positions relative to the coding regions, including intronic, intergenic, downstream, upstream, and 3′-UTR regions, among others (Fig. 1*A*).

Identification of allele-specific SNP–TF interactions using human TF arrays

From the list of 418 SNPs, a set of 30 SNPs were selected in this study for screens on the TF arrays (Table 2, bold entries). The selected SNPs were either located in an enhancer region, implicated in transcription of a target gene through eQTL, or had a combination of these annotations. SNPs were strategically chosen to cover diverse regions in relation to the target genes, while also covering a spectrum of chromosomal positions (Extended

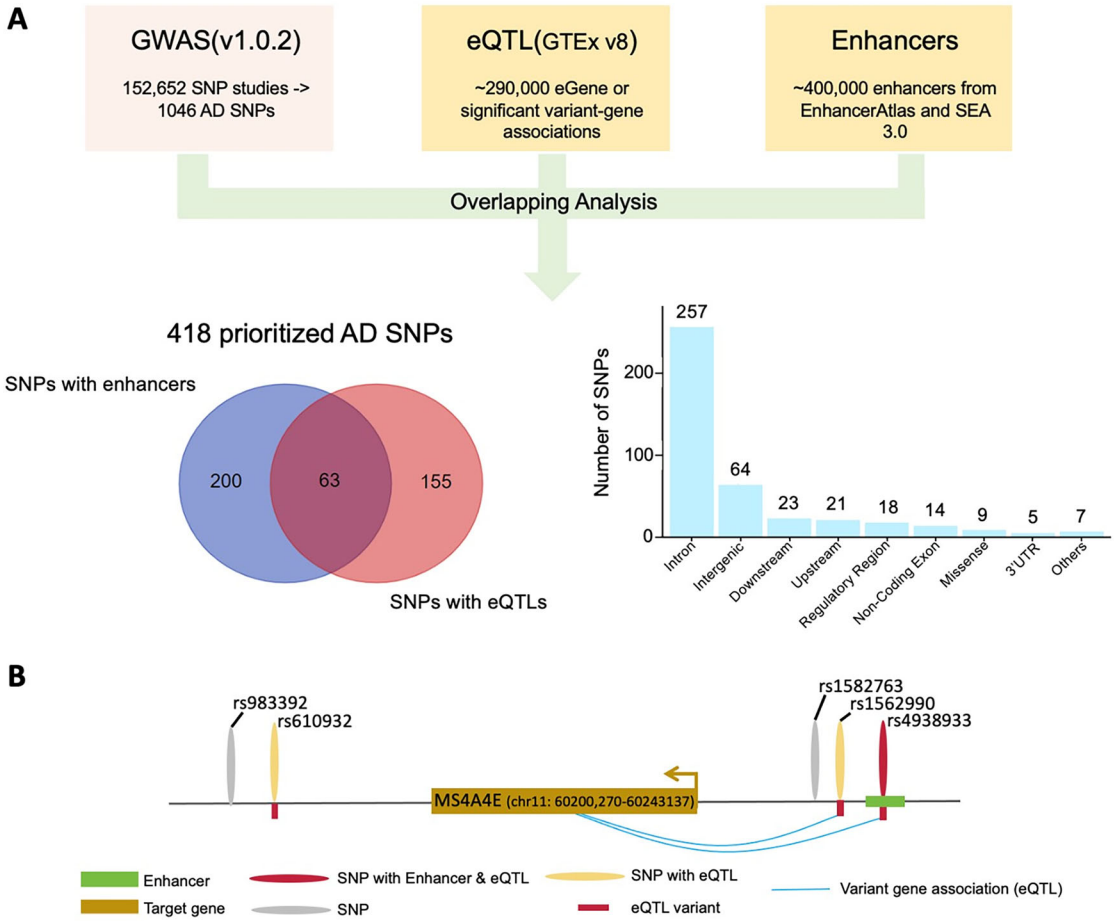

Figure 1. Selection of AD-associated SNPs for probing to the TF arrays. ***A***, 1,046 AD-associated GWAS SNPs were further filtered using overlapping analysis with ∼290,000 brain and nervous system-related eQTLs from the GTEx portal (v8), as well as ∼400,000 enhancer elements from EnhancerAtlas and SEA 3.0. This resulted in a list of 418 prioritized SNPs to be tested using the dye-swap approach on the TF arrays. SNPs within this list span various regions relative to the target gene, within intronic being the region the majority of the SNPs reside in. See Extended Data Figure 1-1 for more details. ***B***, An illustration of the SNPs identified through overlapping analysis.

Table 2. Full list of 418 bioinformatically selected SNPs

rs ID	Chromosome	Position	SNP loci	Mapped genes	PMID
rs17767225	chr14	71199580	14q24.2	LOC105370706	22881374
rs11168036	chr5	140327854	5q31.3	PFDN1, HBEGF	25778476; 28183528
rs679515	chr1	207577223	1q32.2	CR1	25778476
rs9304861	chr19	34780984	19q13.11	LOC105372375, LOC100419834	26830138
rs2075650	chr19	44892362	19q13.32	TOMM40	26993346; 20885792; 20460622; 24770881; 24755620; 19734902; 20061627; 19734903; 21123754; 20100581; 28641921; 20932310
rs9271192	chr6	32610753	6p21.32	HLA-DRB1, LOC107986589	24162737
rs4420638	chr19	44919689	19q13.32	APOC1, APOC1P1	17975299; 26830138; 17998437; 22005931; 17474819; 22832961; 26421299; 28641921
rs6738181	chr2	204263298	2q33.3	LOC100419685, DSTNP5	22881374
rs17366218	chr2	182154019	2q32.1	PDE1A	25778476
rs2279590	chr8	27598736	8p21.1	CLU, CLU	25778476; 19734903
rs7920721	chr10	11678309	10p14	LOC105376413, LOC105376412	25778476; 28183528; 24162737
rs9331896	chr8	27610169	8p21.1	CLU	25778476; 24162737
rs10792832	chr11	86156833	11q14.2	RNU6-560P, LOC107984426	25778476; 24162737
rs56131196	chr19	44919589	19q13.32	APOC1, APOC1P1	26830138; 23419831
rs143083071	chr1	99242198	1p21.3	LOC100129620, LPPR4	26830138
rs143638193	chr4	140579551	4q31.1	RN7SL152P, TBC1D9	26830138
rs145049847	chr16	22190681	16p12.2	SDR42E2, TRL-TAG3-1	26830138
rs857551	chr21	43410112	21q22.3	LOC107987301	26830138
rs10273775	chr7	147200311	7q35	CNTNAP2	22159054
rs4676049	chr2	109018801	2q13	RANBP2	20885792
rs769449	chr19	44906745	19q13.32	APOE	23562540; 26421299; 28247064; 28641921
rs519113	chr19	44873027	19q13.32	PVRL2	23565137
rs7039300	chr9	14064742	9p23	RPL3P11, ATP5HP3	23419831
rs7431992	chr3	54319213	3p21.1	CACNA2D3	26339675
rs75635567	chr8	103186989	8q22.3	BAALC	27770636
rs186588455	chr17	38287606	17q12	LOC105371760	27770636
rs116530595	chr9	110213589	9q31.3	C9orf152, TXN	27770636
rs73239797	chr7	96446404	7q21.3	LOC105375410, LOC105375411	27770636
rs6665019	chr1	25001518	1p36.11	RUNX3, MIR4425	25188341
rs35862341	chr14	55895020	14q22.3	LINC00520, LOC105370511	25188341
rs7638995	chr3	69124224	3p14.1	LMOD3, FRMD4B	22881374
rs6468852	chr8	102963761	8q22.3	LOC100506753	22881374
rs11848070	chr14	71040884	14q24.2	PCNX	22881374
rs4663105	chr2	127133851	2q14.3	LOC105373605	25778476
rs382216	chr5	131351444	5q31.1	CDC42SE2	25778476
rs6890695	chr5	131552731	5q31.1	RAPGEF6	25778476
rs758324	chr5	131773852	5q31.1	FNIP1	25778476
rs142958719	chr5	131897266	5q31.1	MEIKIN	25778476
rs476428	chr5	131965922	5q31.1	ACSL6	25778476
rs75045569	chr7	143412115	7q35	EPHA1-AS1	25778476
rs3851179	chr11	86157598	11q14.2	RNU6-560P, LOC107984426	25778476; 19734902
rs7207400	chr17	45746994	17q21.31	MGC57346-CRHR1	25778476
rs2732703	chr17	46275856	17q21.31	ARL17B, LRRC37A	25778476
rs199499	chr17	46788132	17q21.31	WNT3	25778476
rs9869689	chr3	121607829	3q13.33	FBXO40	25778476
rs1129187	chr6	42964462	6p21.1	PEX6	25778476
rs2854437	chr15	45065212	15q21.1	SORD	25778476
rs2271920	chr8	27458600	8p21.2	PTK2B	25778476
rs11218343	chr11	121564878	11q24.1	SORL1	25778476; 23565137; 24162737
rs59043219	chr1	209797265	1q32.2	IRF6	25778476
rs1936246	chr6	58045670	6p11.2	LOC101927293	25778476
rs116139393	chr7	6732029	7p22.1	LOC107986695, PMS2CL	25778476
rs1347297	chr2	178380259	2q31.2	OSBPL6	26830138
rs12041233	chr1	37287106	1p34.3	RNA5SP43, RPS29P6	26830138
rs182798940	chr1	43453815	1p34.2	SZT2, HYI	26830138
rs7609954	chr3	61650482	3p14.2	PTPRG	26830138
rs116300850	chr5	77556409	5q13.3	WDR41	26830138
rs61142792	chr5	177556462	5q35.3	LOC105377750	26830138
rs12374991	chr7	11188315	7p21.3	PHF14	26830138
rs147213018	chr9	77060964	9q21.2	LOC105376096, RFC5P1	26830138
rs117792039	chr10	105050857	10q25.1	SORCS3	26830138
rs11220271	chr11	125907477	11q24.2	DDX25	26830138
rs61960582	chr13	52557892	13q14.3	LOC105370208, TPTE2P3	26830138

(*Table continues.*)

Table 2. Continued

rs ID	Chromosome	Position	SNP loci	Mapped genes	PMID
rs56146971	chr14	91453757	14q32.12	LOC105370625, SMEK1	26830138
rs189794920	chr16	2608384	16p13.3	LOC652276	26830138
rs79480822	chr17	63576864	17q23.3	DCAF7	26830138
rs34111070	chr17	63716838	17q23.3	STRADA	26830138
rs6714710	chr2	97728623	2q11.2	ZAP70	26993346
rs4965006	chr12	131934988	12q24.33	PUS1	26993346
rs11637445	chr15	67699268	15q23	MAP2K5	26993346
rs8038734	chr15	72519061	15q24.1	ARIH1	26993346
rs433852	chr19	48613847	19q13.33	FAM83E	26993346
rs12134133	chr1	207284500	1q32.2	C4BPAP2, CD55	26993346
rs12044355	chr1	231708601	1q42.2	DISC1, TSNAX-DISC1	19118814
rs2061333	chr19	44110055	19q13.31	LOC100379224	19118814
rs340849	chr1	213944747	1q32.3	PROX1-AS1	22159054
rs17511627	chr13	26150190	13q12.13	RNF6, ATP8A2P3	22159054
rs912330	chr13	98479040	13q32.2	STK24	22159054
rs7364180	chr22	41822852	22q13.2	CCDC134	21123754
rs11782819	chr8	10477271	8p23.1	PRSS52P, LINCR-0001	20452100
rs11055612	chr12	13770394	12p13.1	GRIN2B	20197096
rs78022502	chr2	127638592	2q14.3	LIMS2	23535033
rs538867	chr3	39471787	3p22.1	MOBP	23535033
rs340635	chr4	87010252	4q21.3	AFF1	23535033
rs143954261	chr5	127393758	5q23.2	MEGF10	23535033
rs4794202	chr17	47853173	17q21.32	SP6	23535033
rs117964204	chr17	50614721	17q21.33	CACNA1G	23535033
rs17169634	chr7	34054385	7p14.3	BMPER	24770881
rs1552244	chr3	10093893	3p25.3	FANCD2OS, FANCD2	24755620
rs6857	chr19	44888997	19q13.32	PVRL2	23419831; 28183528; 25188341
rs4968782	chr17	63471115	17q23.3	CYB561, LOC342541	25340798
rs6808835	chr3	46408373	3p21.31	CCRL2	25340798
rs2228467	chr3	42864624	3p22.1	ACKR2	25340798
rs3743162	chr15	84887738	15q25.3	SLC28A1	22005931
rs733175	chr4	10048517	4p16.1	SLC2A9, WDR1	22005930
rs16970672	chr17	77948568	17q25.3	LOC105371909, TNRC6C	22005930
rs4038131	chr2	17593765	2p24.2	VSNL1	22005930
rs10207628	chr2	127094445	2q14.3	BIN1	22005930
rs9811423	chr3	113103475	3q13.2	LOC107986114, LOC101929717	22005930
rs11006923	chr10	28216015	10p12.1	MPP7	22005930
rs6509701	chr19	52880932	19q13.41	ZNF320	22005930
rs11252926	chr10	520439	10p15.3	DIP2C	22005930
rs1800795	chr7	22727026	7p15.3	LOC541472, IL6	26545630
rs1925690	chr6	87157345	6q14.3	ZNF292	21116278
rs10937470	chr3	191283019	3q28	UTS2B	21116278
rs9846480	chr3	138306554	3q22.3	NME9	21116278
rs9899728	chr17	75022679	17q25.1	ICT1, RNU6-362P	26913989
rs394819	chr19	44901322	19q13.32	TOMM40	26339675
rs840163	chr12	56926927	12q13.3	SDR9C7	25649651
rs4474465	chr11	78493334	11q14.1	LOC105369403, NARS2	25649651
rs314277	chr6	104959787	6q16.3	LIN28B	28560309
rs2632516	chr17	58331728	17q22	BZRAP1-AS1	28183528
rs2373115	chr11	78380104	11q14.1	GAB2	17553421
rs6656401	chr1	207518704	1q32.2	CR1, CR1	19734903; 24162737
rs3818361	chr1	207611623	1q32.2	CR1, CR1	19734903; 21460840
rs62209	chr10	10958376	10p14	CELF2	21379329
rs9349407	chr6	47485642	6p12.3	CD2AP	21460841
rs4938933	chr11	60266956	11q12.2	MS4A4E, MS4A4A	21460841
rs3865444	chr19	51224706	19q13.41	CD33	21460841; 24162737
rs6701713	chr1	207612944	1q32.2	CR1	21460841
rs3752246	chr19	1056493	19p13.3	ABCA7	21460841
rs1357692	chr2	107062032	2q12.3	LOC105373535, LOC105373536	22832961
rs10948363	chr6	47520026	6p12.3	CD2AP	24162737
rs11771145	chr7	143413669	7q35	EPHA1-AS1	24162737
rs4147929	chr19	1063444	19p13.3	ABCA7	24162737
rs1476679	chr7	100406823	7q22.1	ZCWPW1	24162737
rs10838725	chr11	47536319	11p11.2	CELF1	24162737
rs72807343	chr5	179811261	5q35.3	SQSTM1	24162737

(*Table continues.*)

Table 2. Continued

rs ID	Chromosome	Position	SNP loci	Mapped genes	PMID
rs9381040	chr6	41186912	6p21.1	LOC107986595, TREML2	24162737
rs7818382	chr8	95041772	8q22.1	NDUFAF6	24162737
rs10751667	chr11	941941	11p15.5	AP2A2	24162737
rs8035452	chr15	50748601	15q21.2	SPPL2A	24162737
rs7225151	chr17	5233752	17p13.2	LOC100130950, SCIMP	24162737
rs3764650	chr19	1046521	19p13.3	ABCA7	21460840
rs610932	chr11	60171834	11q12.2	MS4A6A	21460840
rs142076474	chr3	134501479	3q22.2	CEP63	27770636
rs62179067	chr2	179270456	2q31.2	SESTD1, LOC644776	27770636
rs846858	chr19	44625528	19q13.31	IGSF23	27770636
rs5771225	chr22	50201044	22q13.33	SELO	27770636
rs79926713	chr6	33427745	6p21.32	SYNGAP1	27770636
rs62177277	chr2	179139721	2q31.2	SESTD1	27770636
rs1031261	chr2	32640454	2p22.3	TTC27	22745009
rs3820201	chr1	53115998	1p32.3	SLC1A7	22745009
rs2298948	chr2	75699439	2p12	GCFC2	22745009
rs959695	chr8	99822954	8q22.2	VPS13B	22745009
rs2838923	chr21	45427029	21q22.3	COL18A1	22745009
rs16912145	chr10	58322908	10q21.1	CISD1, UBE2D1	20100581
rs6835098	chr4	173168087	4q34.1	LOC101930370	25188341
rs11158198	chr14	58109602	14q23.1	C14orf37	25188341
rs3003214	chr1	244441734	1q44	ADSS	25188341
rs41526548	chr9	92129496	9q22.31	SPTLC1, LOC100128076	25188341
rs8105265	chr19	2920707	19p13.3	LOC101928631	25188341
rs7589728	chr2	88218921	2p11.2	THNSL2, RNY4P15	26545630
rs4545046	chr8	27699009	8p21.1	SCARA3	26545630
rs12470837	chr2	131824938	2q21.2	C2orf27B, LOC647996	26545630
rs10102274	chr8	90639859	8q21.3	TMEM64	26545630
rs1693575	chr8	100670549	8q22.3	LOC105375672, PABPC1	26545630
rs1662046	chr4	99350902	4q23	ADH1C	26545630
rs741668	chr13	45942333	13q14.13	LOC105370191, ZC3H13	26545630
rs1006064	chr13	51272469	13q14.3	FAM124A	26545630
rs2442825	chr3	9437458	3p25.3	SETD5	26545630
rs131814	chr22	50521672	22q13.33	NCAPH2	26545630
rs144495862	chr7	48193794	7p12.3	ABCA13	26545630
rs316341	chr6	2838014	6p25.2	SERPINB1	28247064
rs184539343	chr2	158437237	2q24.1	CCDC148	28247064
rs115141604	chr3	47209901	3p21.31	KIF9-AS1	28247064
rs13255475	chr8	120455836	8q24.12	MTBP	28247064
rs60871478	chr7	787688	7p22.3	DNAAF5, SUN1	28247064
rs41157	chr22	30009162	22q12.2	MTMR3, HORMAD2-AS1	28247064
rs4267554	chr2	46673906	2p21	LOC105374585	28247064
rs28825742	chr16	70624299	16q22.1	IL34	28247064
rs656900	chr15	79809690	15q25.1	RPS12P25, RNU6-667P	28247064
rs13012722	chr2	169920011	2q31.1	UBR3	26268530
rs8129913	chr21	25692027	21q21.3	JAM2	26268530
rs927675	chr10	28094004	10p12.1	MPP7	21116278
rs6686643	chr1	165647351	1q24.1	MGST3	21116278
rs1569476	chr1	169639679	1q24.2	SELP, LOC107985745	21116278
rs7294478	chr12	7114209	12p13.31	C1RL-AS1	25188341
rs9938198	chr16	20542161	16p12.3	ACSM2B	25188341
rs17879437	chr19	36151982	19q13.12	COX7A1	25188341
rs7631605	chr3	37193098	3p22.2	LOC105377642	20932310
rs12643654	chr4	95238666	4q22.3	UNC5C	20932310
rs2290720	chr12	101293265	12q23.2	UTP20	20197096
rs242557	chr17	45942346	17q21.31	MAPT	28100725
rs7072793	chr10	6064303	10p15.1	IL2RA, RPL32P23	28100725
rs1539581	chr1	111404336	1p13.2	PGCP1, OVGP1	28641921
rs11910985	chr21	46622854	21q22.3	S100B, PRMT2	28641921
rs1981331	chr21	46575530	21q22.3	DIP2A, S100B	28641921
rs17027633	chr1	111419634	1p13.2	OVGP1	28641921
rs1727638	chr6	71429869	6q13	LOC102724000	20932310
rs5998432	chr22	32349929	22q12.3	SLC5A4	20932310
rs76137255	chr19	40277925	19q13.2	AKT2	26252872
rs79811809	chr7	140933681	7q34	LOC105375536	26252872

(*Table continues.*)

Table 2. Continued

rs ID	Chromosome	Position	SNP loci	Mapped genes	PMID
rs138451097	chr19	2873631	19p13.3	ZNF556	26252872
rs2301659	chr19	18924545	19p13.11	DDX49	26252872
rs28671666	chr7	12144804	7p21.3	THSD7A, TMEM106B	25188341
rs1116547	chr5	113344640	5q22.2	MCC	25188341
rs9309711	chr2	3468367	2p25.3	TRAPPC12	25188341
rs11675119	chr2	3472651	2p25.3	TRAPPC12	25188341
rs10166461	chr2	127101837	2q14.3	BIN1	25188341
rs12595161	chr15	52630693	15q21.3	FAM214A	25188341
rs7818345	chr8	19449107	8p21.3	CSGALNACT1	28577822
rs2018293	chr22	26308595	22q12.1	SEZ6L	25188341
rs62174474	chr2	187315682	2q32.1	LOC105373786	25188341
rs2124379	chr11	18276016	11p15.1	SAA1, HPS5	25188341
rs28523990	chr7	124796085	7q31.33	LOC154872, POT1	25188341
rs2452591	chr4	94555568	4q22.3	PDLIM5	25188341
rs11148252	chr13	52434913	13q14.3	VPS36	25188341
rs727505	chr7	124822027	7q31.33	LOC154872, POT1	25188341
rs11024598	chr11	18270189	11p15.1	SAA1, HPS5	25188341
rs4845552	chr1	153507522	1q21.3	RN7SL44P, S100A6	19668339
rs682748	chr5	17148802	5p15.1	LOC285696	19668339
rs1364705	chr8	119212566	8q24.12	MAL2	19668339
rs10781380	chr9	76793228	9q21.2	PRUNE2	19668339
rs1082714	chr12	67235051	12q14.3	RAB11AP2, GGTA2P	19668339
rs8115854	chr20	37137934	20q11.23	MROH8	19668339
rs6031882	chr20	37181380	20q11.23	RPN2	19668339
rs1795240	chr1	171122735	1q24.3	LOC105371611	22903471
rs7414227	chr1	153876520	1q21.3	GATAD2B	22903471
rs11264736	chr1	153966654	1q21.3	SLC39A1	22903471
rs6941712	chr6	130961070	6q23.2	EPB41L2	22903471
rs9426935	chr1	153796924	1q21.3	LOC105371448	22903471
rs2252508	chr1	153941294	1q21.3	DENND4B	22903471
rs16928809	chr11	2915722	11p15.4	SLC22A18	19414484
rs12714207	chr2	88016274	2p11.2	LOC100419917, RNU2-63P	19414484
rs869244	chr10	111149347	10q25.2	LOC724065, BTBD7P2	20526338
rs2893923	chr10	63501424	10q21.3	JMJD1C	20526338
rs4947339	chr6	28948475	6p22.1	TRM-CAT3-1, TRK-TTT3-5	20526338
rs12367822	chr12	56810376	12q13.3	HSD17B6, YWHAQP3	20526338
rs1260326	chr2	27508073	2p23.3	GCKR	27094239
rs6471717	chr8	58464798	8q12.1	UBXN2B, CYP7A1	27094239
rs579459	chr9	133278724	9q34.2	ABO, LCN1P2	19729612
rs1671152	chr19	55014977	19q13.42	LOC107985325, GP6	20526338
rs12922317	chr16	11983775	16p13.13	SNX29	23358160
rs17496332	chr1	107003753	1p13.3	LOC105378889, PRMT6	22829776
rs780093	chr2	27519736	2p23.3	GCKR	22829776
rs7910927	chr10	63379150	10q21.3	JMJD1C	22829776
rs12150660	chr17	7618597	17p13.1	SHBG	22829776
rs1573036	chrX	110576840	Xq23	TDGF1P3, LOC100131200	22829776
rs10454142	chr2	48419260	2p16.3	FOXN2, PPP1R21	22829776
rs3779195	chr7	98364050	7q21.3	BAIAP2L1	22829776
rs1641537	chr17	7642403	17p13.1	SHBG, ATP1B2	22829776
rs11983798	chr7	105640741	7q22.3	ATXN7L1	22881374
rs472926	chr11	126035363	11q24.2	CDON	22881374
rs4937314	chr11	128319206	11q24.3	LOC107984408, LOC105369566	22881374
rs16830122	chr1	154713065	1q21.3	KCNN3	25778476
rs11761441	chr7	82377	7p22.3	LOC101929756	25778476
rs10498633	chr14	92460608	14q32.12	SLC24A4	25778476; 24162737
rs6733839	chr2	127135234	2q14.3	LOC105373605	25778476; 24162737; 25188341
rs2876189	chr6	10101984	6p24.3	LOC107983965	25778476
rs721146	chr21	24540925	21q21.2	LOC105372751	25778476
rs6927354	chr6	6316080	6p25.1	F13A1	26830138
rs2445130	chr1	21911229	1p36.12	HSPG2	26830138
rs11588387	chr1	109995242	1p13.3	AHCYL1	26830138
rs75009721	chr2	24254185	2p23.3	ITSN2	26830138
rs183562580	chr2	26471784	2p23.3	OTOF	26830138
rs182928794	chr2	65140292	2p14	RAB1A, LOC729317	26830138
rs140661185	chr2	65382753	2p14	SPRED2	26830138

(*Table continues.*)

Table 2. Continued

rs ID	Chromosome	Position	SNP loci	Mapped genes	PMID
rs140009341	chr2	136433314	2q22.1	RN7SKP141, SMC4P1	26830138
rs58920042	chr3	71931938	3p13	RN7SL271P, LOC105377157	26830138
rs141568462	chr5	89179048	5q14.3	MEF2C-AS1	26830138
rs188951355	chr5	143553786	5q31.3	LOC105378208	26830138
rs543844	chr6	44457063	6p21.1	CDC5L, LOC105375074	26830138
rs17062407	chr6	133383011	6q23.2	LOC107984121, EYA4	26830138
rs75290158	chr7	97023147	7q21.3	DLX5	26830138
rs188420713	chr8	37276671	8p11.23	SMARCE1P4, LOC100507403	26830138
rs7047415	chr9	95778950	9q22.32	LOC105376161, LINC00476	26830138
rs61860854	chr10	59580135	10q21.2	LOC105378318	26830138
rs117250828	chr11	12171710	11p15.3	MICAL2	26830138
rs141218484	chr11	62481191	11q12.3	AHNAK	26830138
rs3911569	chr11	95362871	11q21	LOC100129203, LOC105369439	26830138
rs116938548	chr12	4699085	12p13.32	NDUFA9, LOC101929549	26830138
rs189465671	chr13	80962230	13q31.1	LOC102724139, LINC00377	26830138
rs117969561	chr13	100558935	13q32.3	GGACT	26830138
rs150511909	chr14	97442349	14q32.2	LOC105370648, LOC101929241	26830138
rs8033755	chr15	28136400	15q13.1	HERC2	26830138
rs137967137	chr16	80433296	16q23.2	LOC102724084	26830138
rs142176337	chr18	59914255	18q21.32	PMAIP1, LOC105372151	26830138
rs115786578	chr20	34149795	20q11.22	RPS2P1, ASIP	26830138
rs147775533	chr20	50382958	20q13.13	LOC105372657, RN7SL636P	26830138
rs189677472	chr21	42641102	21q22.3	LOC101928233, LOC101928255	26830138
rs141503849	chr22	19704370	22q11.21	LOC100420103, SEPT5	26830138
rs11610206	chr12	47245743	12q13.11	LOC105369746	19118814
rs17006206	chr2	27684606	2p23.3	SLC4A1AP	22159054
rs1923775	chr4	2101369	4p16.3	POLN	22159054
rs956225	chr8	121897448	8q24.13	LOC105375732, MRPS36P3	22159054
rs157580	chr19	44892009	19q13.32	TOMM40	21123754; 19125160
rs17798800	chr13	34376390	13q13.2	LOC105370158	23374588
rs514716	chr9	3929424	9p24.2	GLIS3	23562540; 28247064
rs6922617	chr6	41368363	6p21.1	NCR2, LOC100505711	23562540
rs9832461	chr3	39708102	3p22.1	NFU1P1, LOC105377039	20197096
rs58370486	chr7	16668236	7p21.1	BZW2	23535033
rs2392492	chr7	37325592	7p14.1	ELMO1	23535033
rs17172199	chr7	43337677	7p13	HECW1	23535033
rs11023139	chr11	14202800	11p15.2	SPON1	23535033
rs17301739	chr15	58438440	15q21.3	LOC101928694, LIPC	23535033
rs75617873	chr22	44130225	22q13.31	PARVB	23535033
rs59007384	chr19	44893408	19q13.32	TOMM40	23419831
rs9384488	chr6	156688247	6q25.3	NMTRV-TAC1-1, ARID1B	23419831
rs10219670	chr12	105714941	12q23.3	CASC18	23419831
rs61812598	chr1	154447611	1q21.3	IL6R	25340798
rs17429217	chr12	116857528	12q24.22	HRK	22005931
rs2104362	chr6	33857581	6p21.31	LOC105375027	22005931
rs1037757	chr18	59084822	18q21.32	LOC107987259, SEC11C	22005931
rs10792830	chr11	86127766	11q14.2	PICALM, RNU6-560P	22005930
rs157582	chr19	44892962	19q13.32	TOMM40	22005930; 26421299; 28641921
rs3764640	chr19	1207239	19p13.3	STK11	22005930
rs4670766	chr2	37713399	2p22.2	LOC107985870, LOC105374465	21116278
rs7805803	chr7	50185795	7p12.2	C7orf72, IKZF1	21116278
rs4318070	chr13	98308508	13q32.2	FARP1	21116278
rs3784609	chr15	60618351	15q22.2	RORA-AS1, RORA	21116278
rs3905000	chr9	104894789	9q31.1	ABCA1	21116278
rs2243170	chr1	206836565	1q32.1	IL19	25649651
rs2400749	chr14	99570681	14q32.2	CCDC85C	25649651
rs17090219	chr18	56523802	18q21.31	LOC105372132, LOC105372135	28560309
rs56378310	chr13	110537326	13q34	RAB20	28560309
rs11121365	chr1	9297665	1p36.22	SPSB1	28560309
rs2484	chr3	197541698	3q29	BDH1	28560309
rs6016505	chr20	41049649	20q12	TOP1	28560309
rs12525341	chr6	155173590	6q25.2	TIAM2	28560309
rs283811	chr19	44885243	19q13.32	PVRL2	28183528
rs727153	chr4	154733269	4q32.1	NDUFB2P1, LRAT	18823527
rs7081208	chr10	13949865	10p13	FRMD4A, FRMD4A	22430674

(*Table continues.*)

Table 2. Continued

rs ID	Chromosome	Position	SNP loci	Mapped genes	PMID
rs17314229	chr10	13974159	10p13	FRMD4A, FRMD4A	22430674
rs1532278	chr8	27608798	8p21.1	CLU	21460841
rs7274581	chr20	56443204	20q13.31	CASS4	24162737
rs28834970	chr8	27337604	8p21.2	PTK2B	24162737
rs6448799	chr4	11628425	4p15.33	LOC107986178	24162737
rs115798104	chr3	169273046	3q26.2	MECOM	27770636
rs181696879	chr8	6680850	8p23.1	LOC100507530	27770636
rs141846544	chr8	80432793	8q21.13	RNU6-1213P, LOC105375922	27770636
rs117022236	chr1	19288434	1p36.13	AKR7A3	27770636
rs181299481	chr9	111149099	9q31.3	LOC105376219	27770636
rs146322114	chr15	101501820	15q26.3	LOC105371028, SNRPCP18	27770636
rs76816469	chr16	68759780	16q22.1	CDH1	27770636
rs2244526	chr1	169617708	1q24.2	SELP	27770636
rs1034435	chr22	48492443	22q13.32	FAM19A5	27770636
rs7313581	chr12	25269783	12p12.1	KRAS, LOC105369701	27770636
rs147985478	chr12	69997949	12q15	MYRFL, LOC100125409	27770636
rs150269952	chr1	206526352	1q32.1	RASSF5	27770636
rs138543081	chr9	107043881	9q31.2	LOC340512	27770636
rs117756856	chr7	18212452	7p21.1	HDAC9	27770636
rs188392327	chr2	233651485	2q37.1	UGT1A10, UGT1A8, UGT1A	27770636
rs4667682	chr2	171271410	2q31.1	LOC105373737	22745009
rs10932886	chr2	220855368	2q36.1	LOC107985990, LOC107985988	20100581
rs7610017	chr3	189625635	3q28	TP63	20100581
rs4846835	chr1	230145409	1q42.13	GALNT2	25188341
rs3850579	chr5	142484808	5q31.3	LOC101926941, RPS12P10	25188341
rs55643152	chr14	50020073	14q21.3	C14orf183	25188341
rs34487851	chr2	106026098	2q12.2	LOC105373531, C2orf40	25188341
rs12492269	chr3	178442780	3q26.32	LINC01014	26545630
rs2029773	chr3	107825799	3q13.12	BBX, LINC00635	26545630
rs2581305	chr16	86340602	16q24.1	LINC00917	26545630
rs57375391	chr7	22665007	7p15.3	LOC401312	26545630
rs417387	chr3	42530054	3p22.1	VIPR1	26545630
rs9972327	chr15	90087660	15q26.1	IDH2	26545630
rs9305339	chr21	27903308	21q21.3	LINC00113, LINC00314	26545630
rs17068510	chr8	4005589	8p23.2	CSMD1	26545630
rs6758001	chr2	215750203	2q35	LINC00607	26545630
rs12279261	chr11	113235733	11q23.2	NCAM1	21116278
rs76881547	chr14	96166655	14q32.2	C14orf132, BDKRB2	28247064
rs149151450	chr19	45094525	19q13.32	PPP1R37	28247064
rs17725296	chr2	21837068	2p24.1	LOC645949, RN7SL117P	28247064
rs6770219	chr3	186476609	3q27.3	LOC107986165, LOC253573	28247064
rs2198044	chr8	110161577	8q23.2	RPSAP48, LOC100132280	28247064
rs142199880	chr9	8401021	9p24.1	PTPRD	28247064
rs10225144	chr7	17462966	7p21.1	LOC102659288, LOC105375172	28247064
rs34871495	chr20	56735227	20q13.31	PTMAP6, RNU6-929P	28247064
rs10470013	chr20	50993140	20q13.13	MOCS3, KCNG1	26268530
rs11744848	chr5	57841672	5q11.2	LOC101928505, LOC101928539	26268530
rs903027	chr8	61496869	8q12.3	CLVS1	21116278
rs9471576	chr6	41336067	6p21.1	NCR2	21116278
rs62341097	chr4	173173789	4q34.1	GALNT7	25188341
rs28479400	chr15	99455679	15q26.3	LOC105371017, LOC107984790	25188341
rs61041336	chr16	58699258	16q21	SLC38A7, GOT2	25188341
rs2280302	chr9	94587238	9q22.32	FBP2	28641921
rs12265790	chr10	15831289	10p13	FAM188A	28641921
rs9806191	chr15	63945957	15q22.31	DAPK2	28641921
rs6506440	chr18	6781017	18p11.31	ARHGAP28	28641921
rs2899472	chr15	51223858	15q21.2	CYP19A1	20932310
rs12534221	chr7	131603231	7q32.3	PODXL, EEF1B2P6	20932310
rs36056951	chr8	138953555	8q24.3	COL22A1, KCNK9	26252872
rs76478271	chr19	40819294	19q13.2	CYP2F2P	26252872
rs55704525	chr7	43528967	7p13	HECW1	26252872
rs8190569	chr9	96235779	9q22.32	HSD17B3	26252872
rs509477	chr18	34979331	18q12.1	MAPRE2	26252872
rs113027826	chr2	206684788	2q33.3	DYTN	26252872
rs12316703	chr12	118402652	12q24.23	SUDS3	25188341

(*Table continues.*)

Table 2. Continued

rs ID	Chromosome	Position	SNP loci	Mapped genes	PMID
rs8074980	chr17	57933599	17q22	CUEDC1	25188341
rs12446940	chr16	3912619	16p13.3	CREBBP, LOC102724927	25188341
rs12084151	chr1	238246145	1q43	YWHAQP9, LOC105373220	25188341
rs11118993	chr1	206542522	1q32.1	RASSF5	25188341
rs7626019	chr3	42229189	3p22.1	TRAK1, LOC105377048	25188341
rs13053731	chr22	36286661	22q12.3	MYH9	28577822
rs897148	chr8	125568924	8q24.13	LOC105375746	28577822
rs11769293	chr7	28872190	7p14.3	CREB5, TRIL	25188341
rs7048146	chr9	109537042	9q31.3	PTPN3	25188341
rs34660913	chr11	13136463	11p15.3	LOC105376558, ARNTL	25188341
rs6127813	chr20	56679998	20q13.31	RNU6-1146P, RN7SL170P	25188341
rs262741	chr5	165705969	5q34	RN7SKP60, LOC574080	25188341
rs133911	chr22	44127282	22q13.31	PARVB	25188341
rs6887317	chr5	11370935	5p15.2	CTNND2	25188341
rs6881634	chr5	78335030	5q14.1	RNU6-183P, SCAMP1-AS1	19668339
rs10074258	chr5	107646859	5q21.3	EFNA5	19668339
rs10276619	chr7	50273756	7p12.2	C7orf72, IKZF1	19668339
rs6590322	chr11	128336515	11q24.3	LOC107984408, LOC105369566	19668339
rs3026968	chr1	159177662	1q23.2	CADM3	22291609
rs1919922	chr2	122379314	2q14.3	LOC105373592	22903471
rs11083866	chr19	29245435	19q12	RN7SL340P, LOC284395	22903471
rs17140547	chr11	80666008	11q14.1	ARL6IP1P3, LOC105369409	22903471
rs6742078	chr2	233763993	2q37.1	UGT1A10, UGT1A4, UGT1A8, UGT1A9, UGT1A5, UGT1A3, UGT1A, UGT1A6, UGT1A7, UGT1A1	19414484
rs4773330	chr13	111166485	13q34	ARHGEF7	19414484
rs7940646	chr11	10647681	11p15.4	MRVI1	20526338
rs179429	chr11	2529500	11p15.5	KCNQ1	20526338
rs9843304	chr3	149493600	3q25.1	TM4SF4	27094239
rs4757144	chr11	13309679	11p15.3	ARNTL	23358160
rs8057927	chr16	82659207	16q23.3	CDH13	23358160
rs2411984	chr17	49368389	17q21.33	LOC102724596	22829776

Bold and underlined entries represent the set of 30 SNPs selected for follow up screening in this study.

Data Fig. 1-1). To facilitate observation of differential binding interactions with AD-associated SNPs and TF proteins and to avoid potential bias caused by the fluorophores used to end-label each allele probe, we employed a "dye-swap" approach to simultaneously survey a pair of risk and nonrisk alleles on the TF arrays as depicted in Figure 2*A*. If the TF protein array-based assay is sensitive enough to distinguish single base-pair changes, we would expect to observe three possible modes of allele-specific TF interactions, as described in Figure 2*B*: (1) loss-of-function, when the introduction of a risk allele weakens or prevents binding with a canonical TF; (2) enhanced function, in which the presence of the risk allele increases the binding affinity of the canonical TF and preserves its function; and (3) gain-of-function, meaning the risk allele introduces an alternative canonical binding site for a new TF.

To identify allele-specific SNP–TF interactions, each allele of a given SNP, accompanied by 15 nt flanking contextual sequences on both sides, was synthesized as 31 nt DNA oligo attached with a common reverse T7 primer sequence at the 3′-end (Table 3). This primer sequence enables addition of either Cy3 or Cy5 fluorophore to the end of the T7 sequences used to convert them into double-stranded allele probes and allowed for detection of binding events with transcription factors on the arrays. After the probes are generated, each pair of the risk and nonrisk alleles with different colors were mixed at equal molar ratios and probed to the TF arrays in pairs as a competition assay. For example, on one TF array an equimolar mixture of the Cy3-labeled risk and Cy5-labeled nonrisk sequences was probed, while the opposite pairing was probed to a separate array. This approach serves the purpose of eliminating potential inherent bias from the dyes, as well as acting as an experimental replicate. It also enables a more accurate and sensitive quantitative measurement of binding ratios between a nonrisk and risk allele. A preferred allele is expected to show a stronger signal on a protein spot in both labels than its counterpart. For example, SMAD4 showed a strong preference for the nonrisk allele of rs17366218, as a strong Cy3 signal was observed on the first array and an equally strong Cy5 signal was observed on the second array, with little to no signal for the corresponding risk allele (Fig. 2*C*, left). Nonpreferential binding manifests as detectable signals in both colors on both arrays (Fig. 2*C*, right). As we expected, many of the TFs (1,407) did not show any detectable binding signals, presumably due to the small number of SNPs tested. Of those that did produce binding signals (205), 154 did not show differential binding activity between the risk and nonrisk alleles of a SNP. However, 51 TF proteins did show a very notable distinction. To analyze the binding patterns for each protein spot on the arrays, ratiometric binding analysis is conducted using the formula shown in Equation 1 (i.e., the R^* value; see Materials and Methods; Dudoit et al., 2002). Proteins with highly positive or strongly negative binding ratios demonstrate a clear preference for either the nonrisk or risk alleles, respectively.

Our results clearly showed that the TF protein array-based allele-binding assay was sensitive enough to distinguish single base-pair changes, as exemplified in Figure 2*C*. Furthermore, we observed all three modes of action as predicted (Fig. 2*B*). For example, the interaction of the TF SMAD4 with rs17366218 illustrates the loss-of-function binding modality,

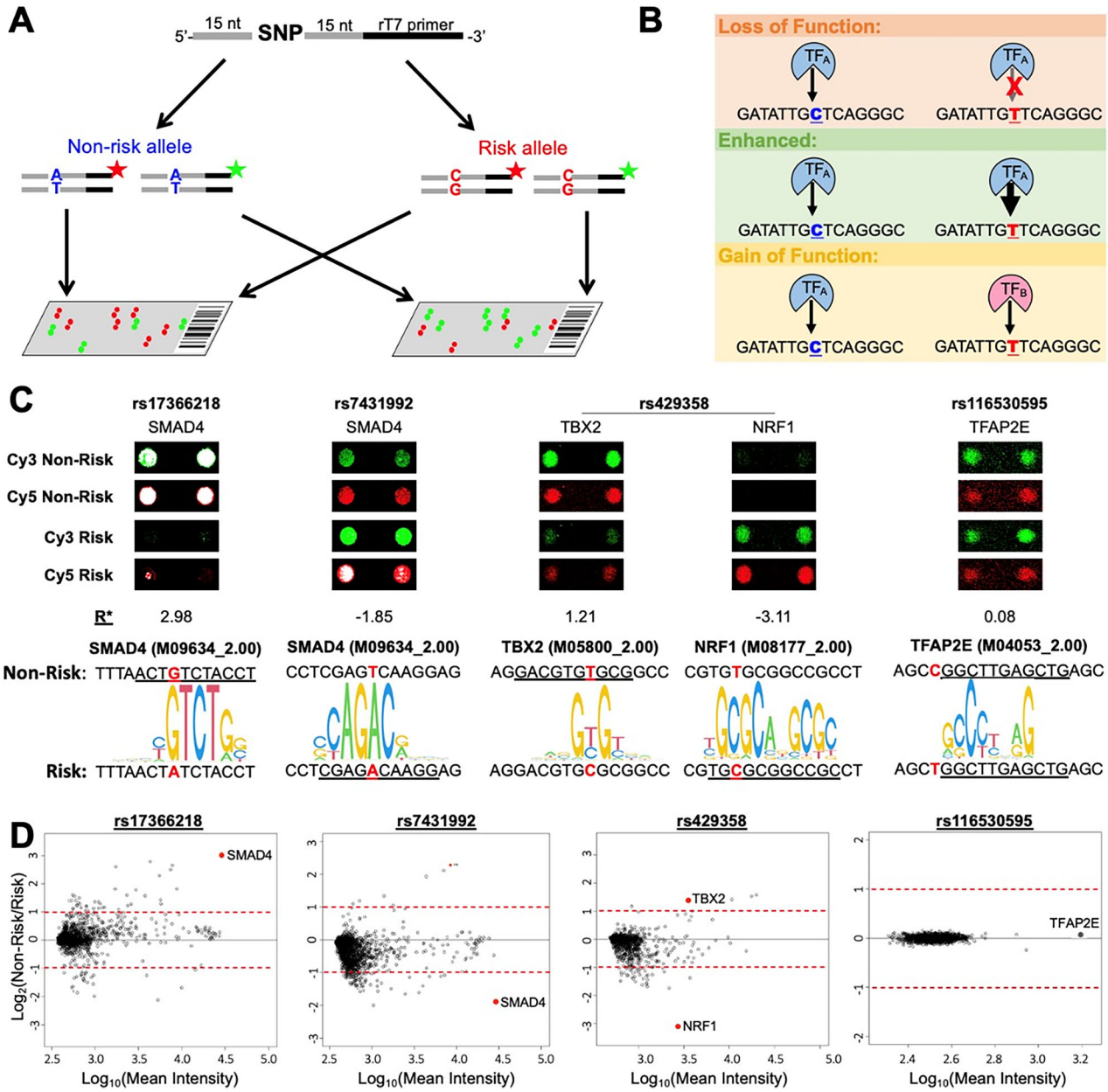

Figure 2. Dye-swap analysis of GWAS SNPs on the TF arrays. ***A***, Illustration of the dye-swap approach. SNP oligos are synthesized (IDT) with 15 bp flanking contextual sequences on either side, and a common T7 primer sequence is included on the 3′-end. Using this common primer sequence, the oligos can be labeled with both Cy3 and Cy5 fluorescent dyes. Labeled oligos are probed to the arrays in oppositely labeled pairs. Ratiometric binding analysis can be conducted on the resulting array signals to determine if there are any allelic preferences with certain transcription factors on the array (R^*, Eq. 1 in Materials and Methods). ***B***, Potential modes of action that may be identified through this workflow. Modes of action include the following: loss of function, in which the risk allele abolishes a canonical transcription factor binding site; enhanced function, in which the risk allele enhances transcription factor binding at a particular motif; gain of function, in which the risk allele introduces a binding site for a new transcription factor. ***C***, Array images for a selection of interactions demonstrating each of the differential binding modes described, as well as one interaction with no differential binding. R^* values representing the ratiometric binding calculation for each of the interactions are shown. Consensus binding motifs (CISBP) are shown in comparison with the SNP sequences. ***D***, MA plots showing the differential binding interactions observed for rs17366218 (left), rs7431992 (second), rs429358 (third). A threshold value of 1 was utilized to distinguish significant differential binding events, as shown by the red dashed lines on each plot. There were no significant differential binding events observed with rs116530595. Binding events were filtered to only those with proteins that are expressed in the brain and nervous system. See Extended Data Figure 2-1 for more details.

evidenced by a robust interaction with the nonrisk allele that is effectively abolished upon the introduction of the risk allele (Fig. 2*C*). This G-to-A mutation disrupts the canonical binding motif for SMAD4, providing plausible explanation for this loss of binding (Fig. 2*C*). In the case of the SNP rs7431992, SMAD4 exhibits an enhanced binding mode of interaction, binding with both alleles but displaying increased affinity for the risk allele. The T-to-A mutation in this sequence context improves the alignment with the consensus binding motif of SMAD4, likely contributing to the increased affinity for the risk allele (Fig. 2*C*). A gain-of-function interaction can be observed with the SNP rs429358. The TF TBX2 showed preference for the nonrisk allele of SNP rs429358, with binding reduced with the introduction of the risk allele (Fig. 2*C*). This matches with the consensus binding motif of TBX2, as *T* is slightly preferred to C at the SNP position. At the same time, this T-to-C mutation in the risk allele introduces the binding motif for another TF NRF1, resulting in minimal to no binding with the nonrisk allele

Table 3. Synthesized oligos for TF array, EMSA, and OCTET experiments

SNP	Allele	Sequence
rs75635567 (BAALC)	NonRisk	5′ GGACCTTAGCCCTTTCGCAAAGTCTCCCAAAACCCTATAGTGAGTGCTATTA 3′
rs75635567 (BAALC)	Risk	5′ GGACCTTAGCCCTTTTGCAAAGTCTCCCAAAACCCTATAGTGAGTGCTATTA 3′
rs17366218 (PDE1A)	NonRisk	5′ ATATTGATTTTAACTGTCTACCTAAGTGAGTACCCTATAGTGAGTGCTATTA 3′
rs17366218 (PDE1A)	Risk	5′ ATATTGATTTTAACTATCTACCTAAGTGAGTACCCTATAGTGAGTGCTATTA 3′
rs10273775 (CNTNAP2)	NonRisk	5′ CTTGCTAACTCCTGCAACAGCTCCACGGATGACCCTATAGTGAGTGCTATTA 3′
rs10273775 (CNTNAP2)	Risk	5′ CTTGCTAACTCCTGCGACAGCTCCACGGATGACCCTATAGTGAGTGCTATTA 3′
rs7431992 (CACNA2D3)	NonRisk	5′ ACACCCTTCCTCGAGTCAAGGAGAAACTGAGACCCTATAGTGAGTGCTATTA 3′
rs7431992 (CACNA2D3)	Risk	5′ ACACCCTTCCTCGAGACAAGGAGAAACTGAGACCCTATAGTGAGTGCTATTA 3′
rs429358 (APOE)	NonRisk	5′ GACATGGAGGACGTGTGCGGCCGCCTGGTGCACCCTATAGTGAGTGCTATTA 3′
rs429358 (APOE)	Risk	5′ GACATGGAGGACGTGCGCGGCCGCCTGGTGCACCCTATAGTGAGTGCTATTA 3′
rs769449 (APOE)	NonRisk	5′ CCTGGCCCCATTCAGGCAGACCCTGGGCCCCACCCTATAGTGAGTGCTATTA 3′
rs769449 (APOE)	Risk	5′ CCTGGCCCCATTCAGACAGACCCTGGGCCCCACCCTATAGTGAGTGCTATTA 3′
rs519113 (PVRL2)	NonRisk	5′ CCTATACTCACACCTCGTAATGTTACCCAGAACCCTATAGTGAGTGCTATTA 3′
rs519113 (PVRL2)	Risk	5′ CCTATACTCACACCTGGTAATGTTACCCAGAACCCTATAGTGAGTGCTATTA 3′
rs2279590 (CLU)	NonRisk	5′ GGAAGTCCTCCTGCTTCTCCAAGGAAACCTAACCCTATAGTGAGTGCTATTA 3′
rs2279590 (CLU)	Risk	5′ GGAAGTCCTCCTGCTCCTCCAAGGAAACCTAACCCTATAGTGAGTGCTATTA 3′
rs9331896 (CLU)	NonRisk	5′ GTCCAGACACAGCTTCGTGGAGGAGGCCTGGACCCTATAGTGAGTGCTATTA 3′
rs9331896 (CLU)	Risk	5′ GTCCAGACACAGCTTTGTGGAGGAGGCCTGGACCCTATAGTGAGTGCTATTA 3′
rs17767225 (PCNX1 - FOXN3)	NonRisk	5′ CTCCAATGGGAATGACGTCTCACAGTGTGAGACCCTATAGTGAGTGCTATTA 3′
rs17767225 (PCNX1 - FOXN3)	Risk	5′ CTCCAATGGGAATGATGTCTCACAGTGTGAGACCCTATAGTGAGTGCTATTA 3′
rs6738181 (DSTNP5)	NonRisk	5′ AAAATTCTAGAGAAGGCAAAATCATAATGACACCCTATAGTGAGTGCTATTA 3′
rs6738181 (DSTNP5)	Risk	5′ AAAATTCTAGAGAAGACAAAATCATAATGACACCCTATAGTGAGTGCTATTA 3′
rs6665019 (RUNX3 - MIR4425)	NonRisk	5′ CGCAGACTACACACTGGTCAGCTGTTCCGGAACCCTATAGTGAGTGCTATTA 3′
rs6665019 (RUNX3 - MIR4425)	Risk	5′ CGCAGACTACACACTAGTCAGCTGTTCCGGAACCCTATAGTGAGTGCTATTA 3′
rs4676049 (EDAR)	NonRisk	5′ TCCCTGCTGAGAGCACGTACAGCAACACTTGACCCTATAGTGAGTGCTATTA 3′
rs4676049 (EDAR)	Risk	5′ TCCCTGCTGAGAGCATGTACAGCAACACTTGACCCTATAGTGAGTGCTATTA 3′
rs73239797 (RNU7-188P - SEM1)	NonRisk	5′ AGGAGGGTTTAGAGGTCAATAGCTCCTGTGAACCCTATAGTGAGTGCTATTA 3′
rs73239797 (RNU7-188P - SEM1)	Risk	5′ AGGAGGGTTTAGAGGACAATAGCTCCTGTGAACCCTATAGTGAGTGCTATTA 3′
rs186588455 (NPEPPSP1 - MRPL45)	NonRisk	5′ GGATCACCTGAGGTCAGAAGTTCGAGACCAGACCCTATAGTGAGTGCTATTA 3′
rs186588455 (NPEPPSP1 - MRPL45)	Risk	5′ GGATCACCTGAGGTCCGAAGTTCGAGACCAGACCCTATAGTGAGTGCTATTA 3′
rs116530595 (C9orf152 - TXN)	NonRisk	5′ CCAAAGAGAGGGAGCCGGCTTGAGCTGAGCAACCCTATAGTGAGTGCTATTA 3′
rs116530595 (C9orf152 - TXN)	Risk	5′ CCAAAGAGAGGGAGCTGGCTTGAGCTGAGCAACCCTATAGTGAGTGCTATTA 3′
rs143083071 (PLPPR4)	NonRisk	5′ TATAAACGTGTGTGCGTGTGTCTTTGTCATAACCCTATAGTGAGTGCTATTA 3′
rs143083071 (PLPPR4)	Risk	5′ TATAAACGTGTGTGCATGTGTCTTTGTCATAACCCTATAGTGAGTGCTATTA 3′
rs145049847 (SDR42E2)	NonRisk	5′ GGGCACGCTCCTGCTCCGCCCCCTGAATCCTACCCTATAGTGAGTGCTATTA 3′
rs145049847 (SDR42E2)	Risk	5′ GGGCACGCTCCTGCTGCGCCCCCTGAATCCTACCCTATAGTGAGTGCTATTA 3′
rs143638193 (RN7SL152P - TBC1D9)	Risk	5′ GCCTCTATCACCTGCCGGGCAGGTGGGAGAGACCCTATAGTGAGTGCTATTA 3′
rs143638193 (RN7SL152P - TBC1D9)	NonRisk	5′ GCCTCTATCACCTGCTGGGCAGGTGGGAGAGACCCTATAGTGAGTGCTATTA 3′
rs10792832 (RNU6-560P - LINC02695)	NonRisk	5′ GTGGGAAAAATGTAGAAGCAAAACATACACAACCCTATAGTGAGTGCTATTA 3′
rs10792832 (RNU6-560P - LINC02695)	Risk	5′ GTGGGAAAAATGTAGGAGCAAAACATACACAACCCTATAGTGAGTGCTATTA 3′
rs11168036 (PFDN1 - HBEGF)	NonRisk	5′ GAAGTGATATTTTTGTACAGAGTTGCTGTTCACCCTATAGTGAGTGCTATTA 3′
rs11168036 (PFDN1 - HBEGF)	Risk	5′ GAAGTGATATTTTTGGACAGAGTTGCTGTTCACCCTATAGTGAGTGCTATTA 3′
rs857551 (LINC01679 - SIK1)	NonRisk	5′ AATCACATTCAAATACGTGAAATAATAATAAACCCTATAGTGAGTGCTATTA 3′
rs857551 (LINC01679 - SIK1)	Risk	5′ AATCACATTCAAATAAGTGAAATAATAATAAACCCTATAGTGAGTGCTATTA 3′
rs56131196 (APOC1 - APOC1P1)	NonRisk	5′ GCATTGAGGCCCAGAGAGGTGAAGTTACTTGACCCTATAGTGAGTGCTATTA 3′
rs56131196 (APOC1 - APOC1P1)	Risk	5′ GCATTGAGGCCCAGAAAGGTGAAGTTACTTGACCCTATAGTGAGTGCTATTA 3′
rs4420638 (APOC1 - APOC1P1)	NonRisk	5′ TGCTACACTTTTCCTAGTGTGGTCTACCCGAACCCTATAGTGAGTGCTATTA 3′
rs4420638 (APOC1 - APOC1P1)	Risk	5′ TGCTACACTTTTCCTGGTGTGGTCTACCCGAACCCTATAGTGAGTGCTATTA 3′
rs35862341 (LINC00520)	NonRisk	5′ GATGGGGTTTCACCATGTTGGCCAGGATGGTACCCTATAGTGAGTGCTATTA 3′
rs35862341 (LINC00520)	Risk	5′ GATGGGGTTTCACCACGTTGGCCAGGATGGTACCCTATAGTGAGTGCTATTA 3′
rs9271192 (HLA-DRB1 - HLA-DQA1)	NonRisk	5′ AATACCCCTCTCATAAAAAGTCATATTTTACACCCTATAGTGAGTGCTATTA 3′
rs9271192 (HLA-DRB1 - HLA-DQA1)	Risk	5′ AATACCCCTCTCATACAAAGTCATATTTTACACCCTATAGTGAGTGCTATTA 3′
rs7039300 (RPL3P11 - ATP5PDP3)	NonRisk	5′ CTTAAAGGGCAGAAGTTACTAAAGCTCCTTAACCCTATAGTGAGTGCTATTA 3′
rs7039300 (RPL3P11 - ATP5PDP3)	Risk	5′ CTTAAAGGGCAGAAGGTACTAAAGCTCCTTAACCCTATAGTGAGTGCTATTA 3′
rs7920721 (USP6NL-AS1 - ECHDC3)	NonRisk	5′ CTCAGCTGTTCACATATTGTCTGTGGCTGCTACCCTATAGTGAGTGCTATTA 3′
rs7920721 (USP6NL-AS1 - ECHDC3)	Risk	5′ CTCAGCTGTTCACATGTTGTCTGTGGCTGCTACCCTATAGTGAGTGCTATTA 3′
rs9304861 (ZNF599 - LINC01801)	Risk	5′ ACACGATGAAACCCCATCTCTACTAAAAATAACCCTATAGTGAGTGCTATTA 3′
rs9304861 (ZNF599 - LINC01801)	NonRisk	5′ ACACGATGAAACCCCGTCTCTACTAAAAATAACCCTATAGTGAGTGCTATTA 3′
Primer Sequences		5′ Cy5 – TAATAGCACTCACTATAGGGT 3′
		5′ Cy3 – TAATAGCACTCACTATAGGGT 3′

and strong binding with the risk allele (Fig. 2*C*). Finally, the interaction between the TF TFAP2E and SNP rs116530595 serves as an example of nondifferential binding, as there is similarly strong signal in both channels for both alleles, resulting in an R^* value close to zero (0.08). This nonpreferential binding event is explained by the fact that the known binding consensus sequence for TFAP2E lies outside of the SNP site. This kind of sensitivity agrees with our previous observation that methylation-dependent

DNA-TF interactions could also be readily detected on the TF arrays (Hu et al., 2013). We visualized the results for each SNP using an MA plot, which displays the $\log_2$ ratio of intensities (M) versus the average log intensity (A), allowing for the identification of differential interactions between the SNP alleles and TFs (Dudoit et al., 2002). In this study, differential binding hits were characterized as R^* values above or below the threshold of ±1, representing two-fold difference in binding (Fig. 2*D*). Identified differential binding events were filtered to consist of interactions with proteins expressed in the brain and nervous system (GTEx Portal; Extended Data Fig. 2-1).

In vitro validation of identified allele-specific SNP–TF interactions

Through the described analysis of array assays conducted on the selected 30 AD-associated SNPs, a total of 90 differential interactions were identified across 51 unique TF proteins (Fig. 3*A*, Table 4). In addition to these differential interactions, we also observed a total of 794 nondifferential binding interactions across each of the SNPs tested, encompassing 154 distinct TF proteins (Extended Data Fig. 3-1). Such nondifferential interactions by the 154 TFs could be mostly explained by finding their corresponding consensus sequences in the probe sequences, suggesting that they might interact with the TFs showing preferential binding activity. Given the length of the DNA probes that we utilized in these experiments (31 bp), it is possible to observe sequence-specific binding of 2–3 TF proteins with the same sequence, as the average TF consensus motif is between 5 and 20 bases long (Pachkov et al., 2007; Pratt et al., 2022). Identification of heterodimers between differentially bound and nondifferentially bound proteins can provide further evidence of pathways involved in the molecular mechanisms underlying these variants. Therefore, we decided to determine whether there are any known interaction networks formed out of the proteins shown to bind each of the different SNPs, including differential and nondifferential binders. Using the STRING analysis, we mapped the known protein–protein interactions among the identified TFs and found several SNPs with significantly interconnected protein–protein interactions networks (Szklarczyk et al., 2023). We pinpointed three SNPs, rs17366218 (*PDE1A*), rs7431992 (*CACNA2D3*), and rs769449 (*APOE*) exhibiting a range of binding preferences with the TF SMAD4, and notably,

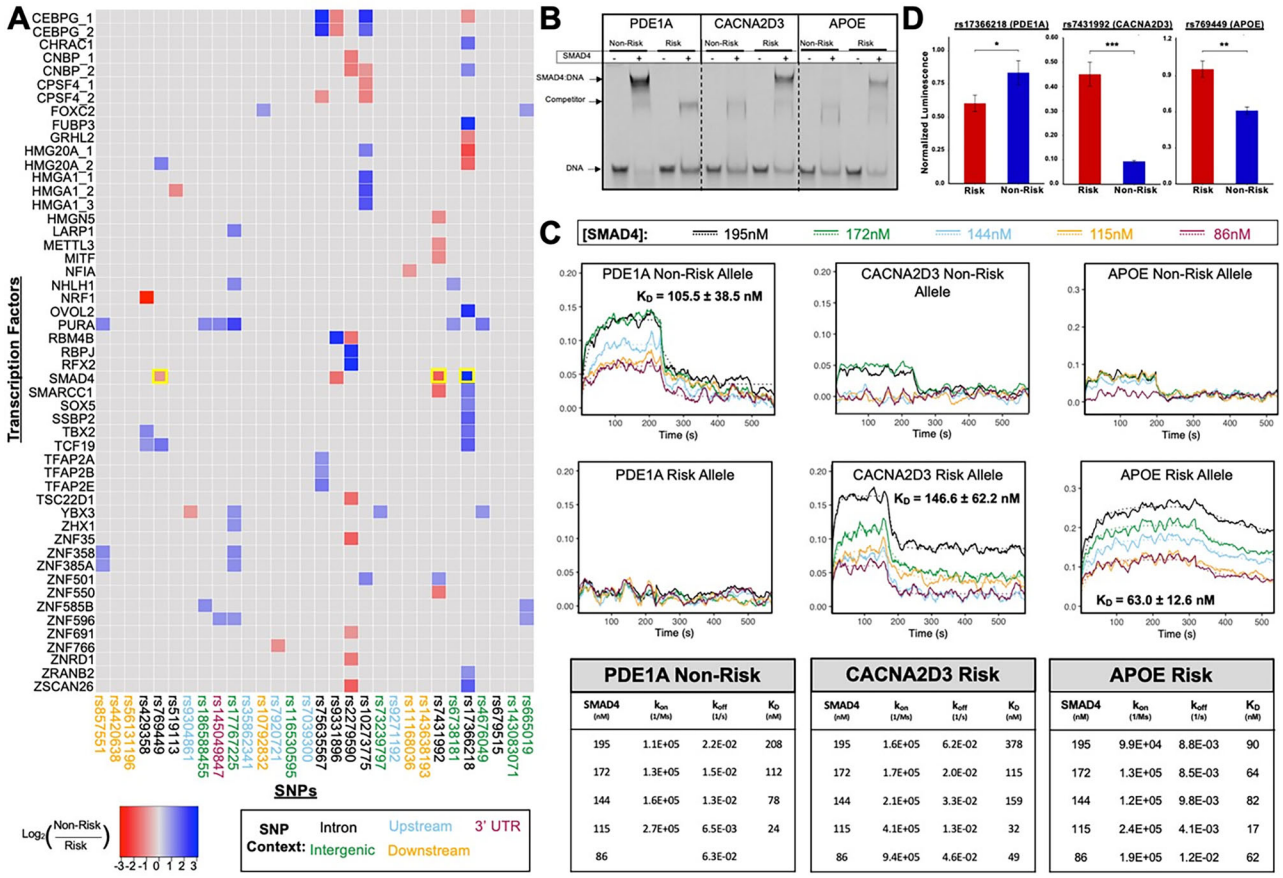

Figure 3. In vitro validation of hits identified through orthogonal assays. ***A***, An allele-specific heatmap showing differential binding events across all SNPs tested on the TF arrays based on calculated R^* values (Table 4). On the *x*-axis are the SNPs tested, and on the *y*-axis are all the interacting transcription factors. Boxes colored in shades of blue represent a range of preference for the nonrisk allele of a SNP, while boxes shaded in red represent a range of preference for the risk allele. SNPs boxed in yellow were chosen for follow-up studies. In addition to these allele-specific interactions, nondifferential interactions were also observed, as shown in Extended Data Figure 3-1. STRING and GO analysis for the three selected SNPs can be found in Extended Data Figure 3-2. ***B***, Validation of selected interactions with SMAD4 using EMSA. Oligos for both alleles of each SNP were end-labeled with Cy5 fluorescent dye, converted to dsDNA, and incubated with purified SMAD4 protein. Transcription factor binding is observed through a shift in the bands on the PAGE gel from the no-protein control samples. Preferences demonstrated on the TF arrays were all replicated through this assay. ***C***, Biolayer interferometry (OCTET) was utilized to determine the kinetic parameters for each of these interactions. SMAD4-*PDE1A* nonrisk $K_D = 105.5 \pm 38.5$ nM; SMAD4-*CACNA2D3* risk $K_D = 146.6 \pm 62.2$ nM; SMAD4-*APOE* risk $K_D = 63.0 \pm 12.6$ nM; K_D values for SMAD4-*PDE1A* risk, SMAD4-*CACNA2D3* nonrisk and SMAD4-*APOE* nonrisk were not able to be determined. ***D***, Results of luciferase assays to determine the transcriptional impact of these interactions in cells. Sequences utilized for these experiments can be found in Table 5. In each case, previously identified preference of SMAD4 persisted. $^*p < 0.05$; $^{**}p < 0.00005$; ***0.0000005 (Student's *t* test).

Table 4. Identified allele-specific TF–DNA interactions; *R** calculated using Equation 1

SNP	Protein name	Protein ID	*R**
rs75635567	CEBPG	I0H6858	2.7
rs75635567	CEBPG	I0H6858	2.46
rs75635567	CPSF4	I0H4475	−1.06
rs75635567	TFAP2A	I0H22136	1.06
rs75635567	TFAP2C	I0H28749	1.02
rs75635567	TFAP2E	I0H26264	1.45
rs17366218	BOD1L	I0H11787	1.28
rs17366218	CEBPG	I0H6858	−1.05
rs17366218	CHRAC1	I0H10035	1.6
rs17366218	CNBP	I0H59002	1.22
rs17366218	FUBP3	I0H6696	2.76
rs17366218	GRHL2	I0H38073	−1.25
rs17366218	HMG20A	I0H10020	−2.09
rs17366218	HMG20A	I0H10020	−1.68
rs17366218	OVOL2	I0H6650	2.58
rs17366218	SMAD4	I0H3638	2.99
rs17366218	SMARCC1	BC113465	1.43
rs17366218	SOX5	I0H43581	1.28
rs17366218	SSBP2	I0H10629	1.79
rs17366218	TBX2	I0H29393	1.57
rs17366218	TCF19	I0H3967	1.81
rs17366218	ZSCAN26	I0H14153	1.82
rs17366218	ZRANB2	I0H26283	1.07
rs10273775	CEBPG	I0H6858	2.49
rs10273775	CEBPG	I0H6858	1.73
rs10273775	CNBP	I0H59002	−1.01
rs10273775	CPSF4	I0H27075	−1.17
rs10273775	CPSF4	I0H4475	−1.29
rs10273775	HMG20A	I0H10020	1.36
rs10273775	HMGA1	I0H6516	2.02
rs10273775	HMGA1	I0H5224	1.91
rs10273775	HMGA1	I0H6516	2.06
rs10273775	ZNF501	I0H10819	1.25
rs7431992	BOD1L	I0H11787	−1.51
rs7431992	HMGN5	I0H7341	−1.12
rs7431992	METTL3	I0H3728	−1.08
rs7431992	MITF	I0H22837	−1.14
rs7431992	SMAD4	I0H3638	−1.85
rs7431992	SMARCC1	BC113465	−1.6
rs7431992	ZNF501	I0H10819	1.04
rs7431992	ZNF550	I0H22049	−1.42
rs429358	NRF1	I0H11918	−3.11
rs429358	TBX2	I0H29393	1.22
rs429358	TCF19	I0H3967	1.01
rs769449	HMG20A	I0H10020	1.34
rs769449	SMAD4	I0H3638	−0.95
rs769449	TCF19	I0H3967	1.48
rs519113	HMGA1	I0H6516	−1.18
rs2279590	CNBP	I0H3404	−1.35
rs2279590	CNBP	I0H59002	−1.39
rs2279590	RBM4B	I0H5506	−1.46
rs2279590	RBPJ	I0H52060	2.64
rs2279590	RFX2	I0H11384	2.74
rs2279590	TSC22D1	I0H3511	−1.56
rs2279590	ZSCAN26	I0H14153	−1.77
rs2279590	ZNF35	I0H9658	−1.77
rs2279590	ZNF691	I0H2943	−1.01
rs2279590	ZNRD1	I0H12855	−1.36
rs9331896	CEBPG	I0H6858	−1.26
rs9331896	CEBPG	I0H6858	−1.24
rs9331896	RBM4B	I0H5506	2.67
rs9331896	SMAD4	I0H3638	−1.54
rs17767225	PURA	BC036087	2.18
rs17767225	LIN28	I0H13570	1.42
rs17767225	LARP1	I0H21797	1.33

(*Table continues.*)

Table 4. Continued

SNP	Protein name	Protein ID	*R**
rs17767225	ZNF358	I0H22979	1.22
rs17767225	NHLH1	I0H9734	1.2
rs17767225	ZNF385A	I0H22141	1.12
rs17767225	YBX3	I0H13867	1.06
rs17767225	ZHX1	I0H55729	1.05
rs17767225	ZNF596	I0H11010	1.01
rs6738181	PURA	BC036087	1.06
rs6738181	NHLH1	I0H9734	0.95
rs6665019	ZNF585B	I0H61959	1.04
rs6665019	FOXC2	BC111589	0.96
rs6665019	ZNF596	I0H11010	0.96
rs4676049	PURA	BC036087	1.46
rs4676049	YBX3	I0H13867	1.01
rs73239797	YBX3	I0H13867	1.07
rs186588455	PURA	BC036087	1.2
rs186588455	ZNF585B	I0H61959	1.16
rs145049847	PURA	BC036087	1.25
rs145049847	ZNF596	I0H11010	1.06
rs10792832	FOXC2	BC111589	0.96
rs11168036	NFIA	I0H12791	−0.96
rs857551	PURA	BC036087	1.25
rs857551	ZNF358	I0H22979	1.19
rs857551	ZNF385A	I0H22141	1.05
rs7920721	ZNF766	I0H55318	−1.1
rs9304861	YBX3	I0H13867	−0.97

SMAD4 emerges as a central node within the interaction networks of the proteins associated with each of these SNPs. Associated GO terminology with these networks also shows relevancy to processes in the brain, suggesting that a role in AD is possible (Extended Data Fig. 4; Thomas et al., 2022). SMAD4 is a signal transduction protein that plays a role in the TGFβ signaling pathway that is critical toward proper neural development and function (Meyers and Kessler 2017). Modulation of the activities of SMAD family proteins in the TGFβ pathway have been shown to impact neurogenesis (Hiew et al., 2021). Additionally, SMAD4 plays a significant role in the BMP signaling pathway, which is vital for neurogenesis in the adult hippocampus—a region notably affected in AD (Zhou et al., 2022). Considering SMAD4's apparent centrality within potential interaction networks for these SNPs, its pivotal role in neural processes, and its potential relevance to AD pathology, we were motivated to pursue further investigation into these interactions.

To begin to validate the findings we observed on the arrays, we focused on the interaction of these three SNPs with SMAD4 (Fig. 3*A*, highlighted in yellow). Our first goal was to replicate the binding interactions we observed on the arrays using an orthogonal assay. To achieve this, we employed an electrophoretic mobility shift assay (EMSA). For each of the selected SNPs, we generated Cy5 fluorescent probes for each allele and subsequently incubated each of these probes separately with purified SMAD4 protein. The resulting reactions were then analyzed using gel electrophoresis to assess any mobility shifts of the DNA probes in comparison with the protein-free control reactions. Ultimately, the EMSA results precisely reproduced each of the binding preferences observed on the TF arrays. Specifically, SMAD4 exhibited a preference for the risk allele of both rs7431992 and rs769449, while favoring the nonrisk allele of rs17366218 (Fig. 3*B*).

Next, we determined the affinity values for each of these interactions, by utilizing the OCTET system, a real-time, label-free kinetics instrument. This instrument utilizes biolayer interferometry to determine the k_{on} and k_{off} values so that the K_D values can be

deduced for a given interaction (Barrows and Van Dyke 2022; Pluta et al., 2022). For this experiment, each of the alleles for all three SNPs were end-labeled with biotin and loaded onto OCTET biosensors coated with streptavidin. After a blocking step, each of the biosensors was immersed into purified SMAD4 protein at various concentrations, allowing us to observe binding with the immobilized probe sequences in real time. From this experiment, the fitted on- and off-curves were generated based on the raw data (Fig. 3*C*). The binding activities were consistent with our previous results, with SMAD4 exclusively binding the expected alleles and K_D values ranging from 63 to 146 nM, aligning closely with values previously reported in the literature for TF–DNA interactions.

Examining the impacts of identified allele-specific SNP–TF interactions

It has been well established that many TF proteins can act both as transcription activators and repressors, depending on its binding partners and/or the surrounding chromatin context (Bylino et al., 2020; Weidemüller et al., 2021). It is also possible that a TF-binding event does not translate to changes in downstream gene transcription when it sits on a poised enhancer (Spivakov 2014; Banks et al., 2016). Therefore, we next sought to determine if these binding interactions could have a direct impact on transcriptional regulation with a cell-based luciferase assay. First, SNP allele sequences were cloned into luciferase reporter vector pGL4.32 to replace the NF-kB response element present to drive the luciferase reporter gene (Table 5). Additionally, an overexpression vector for SMAD4 was generated, driven by a CMV promoter. Cells were cotransfected with the luciferase reporter carrying either the risk or nonrisk allele, the SMAD4 overexpression vector, and a Renilla control vector. Following 2 d of incubation, cell lysates were processed through a dual luciferase reporter assay system, measuring the luminescence generated after successive treatment with Firefly luciferase and Renilla substrates. To account for well-to-well variability in cell count, Firefly luciferase signals were normalized to Renilla control signals. Signals were further normalized to control wells containing SNP luciferase vector with no overexpression vector added to account for background interaction with endogenous proteins. After quantification of this assay, we observed significantly increased induction of luciferase expression with each of the alleles with which SMAD4 had previously shown preference, suggesting that SMAD4 is likely to act as a transcription activator via preferential binding activity to the identified alleles (Fig. 3*D*).

Cell-based validation of SMAD4 differential interactions

Finally, we wanted to see if the differential binding interactions observed on the arrays could persist on the expected genetic backgrounds in cells. To achieve this, the homozygous and heterozygous HEK293t cell lines carrying the risk and nonrisk alleles of *CACNA2D3* and *PDE1A* were generated using a CRISPR-based genetic engineering method as illustrated in Figure 4*A–C* and as described previously (Ran et al., 2013). Due to challenges with cell line generation, we were unable to produce cells with the corresponding nonrisk and risk alleles at the *APOE* locus. SMAD4 expression construct was transiently transfected to these cell lines under the control of the CMV promoter (Fig. 4*C*). After 48 h of incubation, cells were harvested and anti-SMAD4 antibodies were used to perform chromatin immunoprecipitation (ChIP). The SMAD4 occupancy at different loci was determined via qPCR using primer pairs specific for each SNP containing region (Table 6). For the *PDE1A* and *CACNA2D3* experiments, only these loci were altered, with all others tested being wild type. mRNA expression of *PDE1A* and *CACNA2D3* in the same cell lines after SMAD4 overexpression were also determined in parallel (Table 7).

As a result of these experiments, SMAD4 showed a significantly higher occupancy to the homozygous nonrisk allele than the risk allele at the PDE1A locus, with attenuated binding observed in the heterozygous line compared with the nonrisk (Fig. 4*D*, left panel). No significant changes in occupancy were observed in the other wild-type loci tested. Following SMAD4 overexpression in the same cell lines, the expression level of *PDE1A* was observed to be elevated in the nonrisk line compared with the risk and heterozygous lines (Fig. 4*E*, left panel). This supports the notion that the SNP located at rs17366218 interferes with SMAD4's engagement with the DNA at this locus, potentially resulting in reduced *PDE1A* expression in the context of AD. In the same experiment using the *CACNA2D3* cell lines, SMAD4 was found to occupy the homozygous risk allele to a higher degree than the nonrisk or heterozygous lines, with the heterozygous line showing intermediate occupation (Fig. 4*D*, right panel). Again, no significant impact in allele occupation could be observed in the other wild-type loci tested. However, no allele-dependent effects can be observed in the expression of *CACNA2D3* after SMAD4 overexpression (Fig. 4*E*, right panel). This contrasts with the result we observed in the luciferase assay, in which the binding interaction between SMAD4 and the SNP in *CACNA2D3* impacted the transcription of the luciferase gene. One potential explanation for this discrepancy is that the luciferase expression vector utilized in the assay contained three copies of the SNP, potentially amplifying the effect of TF binding compared with the single copy present in the native genomic context. Additionally, in a real cellular context, compensation by other endogenous factors might mitigate the impact of decreased SMAD4 binding at that locus. The absence of expression changes could also reflect regulatory redundancy, where multiple regulatory elements can compensate for each other to maintain proper

Table 5. Synthesized alleles for luciferase experiments

rs17366218 (PDE1A)	NonRisk	5′CTAGCTTTAACT**G**TCTACCTTTTAACT**G**TCTACCTTTTAACT**G**TCTACCTTTTAACT**G**TCTACCTA 3′
		5′GAAATTGA**C**AGATGGAAAATTGA**C**AGATGGAAAATTGA**C**AGATGGAAAATTGA**C**AGATGGATTCGA 3′
	Risk	5′CTAGCTTTAACT**A**TCTACCTTTTAACT**A**TCTACCTTTTAACT**A**TCTACCTTTTAACT**A**TCTACCTA 3′
		5′GAAATTGA**T**AGATGGAAAATTGA**T**AGATGGAAAATTGA**T**AGATGGAAAATTGA**T**AGATGGATTCGA 3′
rs7431992 (CACNA2D3)	NonRisk	5′CTAGCCCTCGAG**T**CAAGGAGCCTCGAG**T**CAAGGAGCCTCGAG**T**CAAGGAGCCTCGAG**T**CAAGGAGA 3′
		5′GGGAGCTC**A**GTTCCTCGGAGCTC**A**GTTCCTCGGAGCTC**A**GTTCCTCGGAGCTC**A**GTTCCTCTTCGA 3′
	Risk	5′CTAGCCCTCGAG**A**CAAGGAGCCTCGAG**A**CAAGGAGCCTCGAG**A**CAAGGAGCCTCGAG**A**CAAGGAGA 3′
		5′GGGAGCTC**T**GTTCCTCGGAGCTC**T**GTTCCTCGGAGCTC**T**GTTCCTCGGAGCTC**T**GTTCCTCTTCGA 3′
rs769449 (APOE)	NonRisk	5′CTAGCCATTCAG**G**CAGACCCCATTCAG**G**CAGACCCCATTCAG**G**CAGACCCCATTCAG**G**CAGACCCA 3′
		5′GGTAAGTC**C**GTCTGGGGTAAGTC**C**GTCTGGGGTAAGTC**C**GTCTGGGGTAAGTC**C**GTCTGGGTTCGA 3′
	Risk	5′CTAGCCATTCAG**A**CAGACCCCATTCAG**A**CAGACCCCATTCAG**A**CAGACCCCATTCAG**A**CAGACCCA 3′
		5′GGTAAGTC**T**GTCTGGGGTAAGTC**T**GTCTGGGGTAAGTC**T**GTCTGGGGTAAGTC**T**GTCTGGGTTCGA 3′

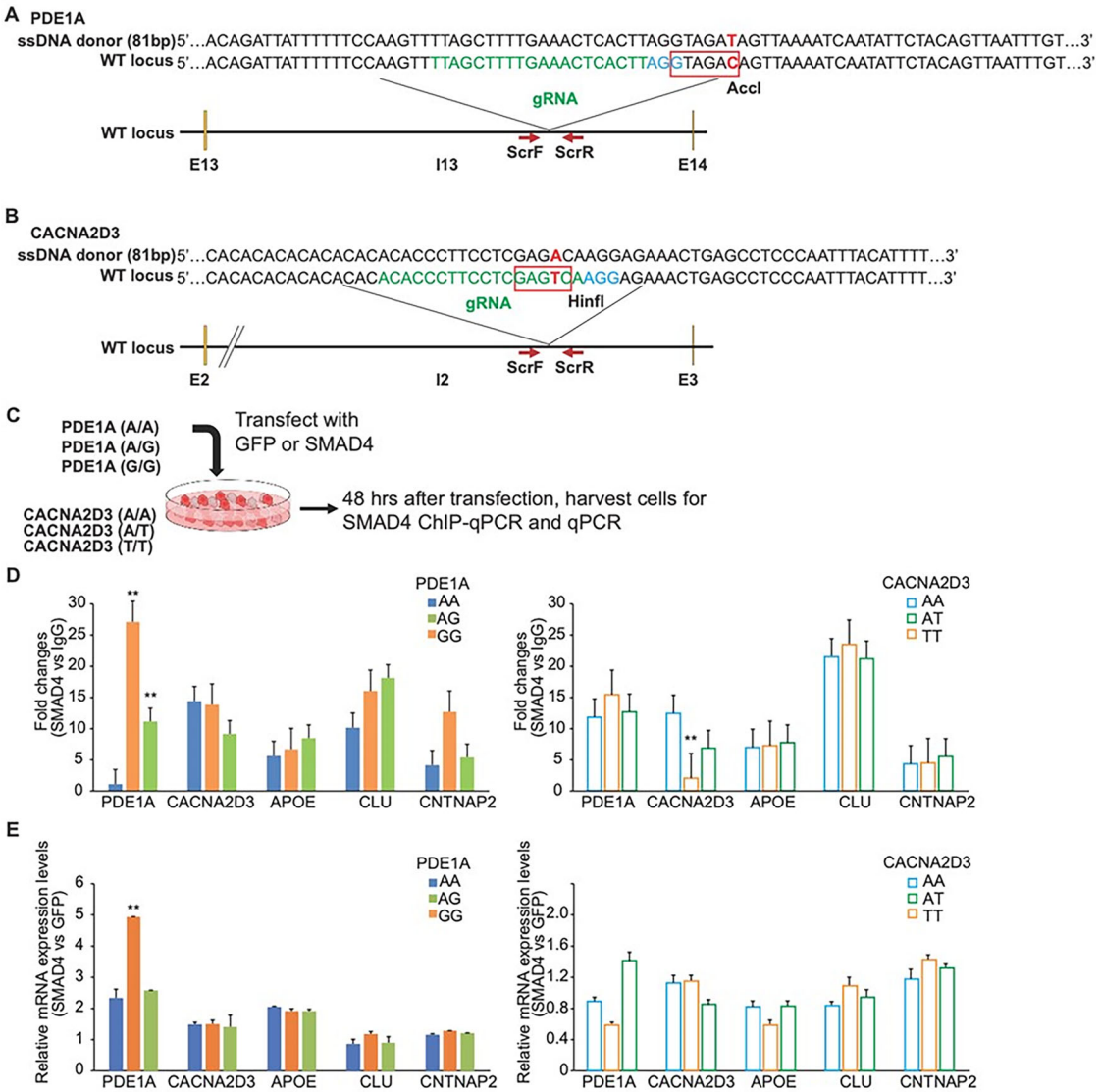

Figure 4. ChIP-qPCR validation in CRISPR-generated cell lines ***A***, Generation of *PDE1A* locus point mutations in HEK293t cells. ***B***, Generation of *CACNA2D3* locus point mutations in HEK293t cells. ***C***, Illustration of the transfection protocol utilized for SMAD4 or GFP control overexpression. ***D***, Relative occupation by SMAD4 at the *PDE1A* (left panel) and *CACNA2D3* (right panel) loci as determined by qPCR compared with an IgG control. Primers utilized can be found in Table 6. ***E***, Relative mRNA expression levels of *PDE1A* (left panel) and *CACNA2D3* (right panel) following SMAD4 overexpression compared with a GFP control. Primer sequences utilized can be found in Table 7.

Table 6. ChIP-qPCR primers utilized for experiments in Figure 4

Gene target	Forward primer	Reverse primer
rs17366218 (PDE1A)	5′ AGAGAAATGCCAAAAATTCACA 3′	5′ CCAAGTTTTAGCTTTTGAAACTCAC 3′
rs7431992 (CACNA2D3)	5′ GGAAGGGAAGCATTTTGTGA 3′	5′ GGGAGGCTCAGTTTCTCCTT 3′
rs769449 (APOE)	5′ CCAATCACAGGCAGGAAGAT 3′	5′ AGGAGGTTGAGGTGAGGATG 3′

Table 7. mRNA primers used for experiments in Figure 4

Gene target	Forward primer	Reverse primer
rs17366218 (PDE1A)	5′ AATGTGGCAGCGCCTGAAAGGA 3′	5′ CTTCCAGCACAGATGCCGCATA 3′
rs7431992 (CACNA2D3)	5′ GAACATCCCGATGTGTCCTTGG 3′	5′ ACTGGAGCAGAGGTTCTTTGCC 3′
rs769449 (APOE)	5′ GTGGATGTGCTCAAAGACAGCG 3′	5′ GCTTGCTGAAGGTGGAGGTCAC 3′

gene expression levels when individual elements are disrupted. Another important consideration to note is the choice of cells for this experiment. We used HEK293t cells, which differ from brain cells where *CACNA2D3* is primarily active. Therefore, the results we obtained might not accurately represent what happens in the brain. *CACNA2D3* is known to have its highest expression in brain tissue, and it is possible that a different cellular environment could lead to changes in the gene's expression. This suggests the importance of considering the specific cellular context when interpreting these results.

Discussion

Despite its widespread prevalence, the precise molecular mechanisms underlying AD remain poorly understood. GWAS have begun to uncover variations in the human genome that are associated with AD (Han et al., 2010; Shen et al., 2010; Schott et al., 2016; Jun et al., 2017); however, the causal nature of many of these variants has yet to be determined in a systemic way. Most variations identified via GWAS are SNPs; however, the challenge is that many of the SNPs are located within linkage disequilibrium and/or noncoding regions of the genome, making it

challenging to establish potential functionality. In this study, we aimed to generate a high-throughput TF-WAS pipeline to facilitate the identification of functional GWAS SNPs and associated TFs that may play a role in the etiology of AD. Using multimodality bioinformatics analysis, we were able to narrow down a list of AD SNPs that (1) are associated with AD through GWAS, (2) have demonstrated an impact on gene expression in the brain and nervous system via eQTL, and (3) are located within noncoding regions of the genome (as annotated by EnhancerAtlas and SEA 3.0). This allowed us to focus on SNPs that are likely to be functional in cell types relevant to disease pathology. By screening SNPs from this list on the human TF protein arrays, we were able to detect differential binding interactions with transcription factors, potentially providing molecular insights into how these genetic variants may convey dysregulation of transcription and therefore, influencing AD susceptibility. Using several orthogonal in vitro methods, we could validate all the allele-specific SNP–TF interactions observed on the TF array and thus paving the way to identify functional SNPs in other complex diseases.

Of the notable findings from our study were the allele-specific binding preferences of SMAD4 with several AD-associated SNPs (rs17366218, rs7431992, and rs769449), with SMAD4 showing a range of binding preferences with these SNPs. SMAD4 is highly conserved from yeast to humans and is also widely expressed in various human tissues and organs, including the human brain (Human Protein Atlas; De Caestecker et al., 1997; Uhlén et al., 2015). This protein acts as a signal transducer in the TGFβ and BMP signaling pathways and plays numerous roles in various human diseases, particularly in cancer, and has also been implicated in AD (Das and Golde 2006; Wan et al., 2021; Wang et al., 2021; Kapoor and Chinnathambi 2023). The role of SMAD4 in the BMP signaling pathway is especially interesting, as this pathway is critical to neurogenesis in the adult hippocampus, a region of the brain that is notably affected in AD (Zhou et al., 2022).

One target of interest for SMAD4 that we identified using the TF arrays is rs17366218, an intron variant mapped to the *PDE1A* gene. *PDE1A* is a part of a family of cyclic nucleotide phosphodiesterases (PDEs) and functions in the degradation of cyclic nucleotide second messengers, showing preference toward the degradation of cyclic guanidine monophosphate (cGMP; Azevedo et al., 2014). *PDE1A* is specifically expressed at high levels in the brain, kidney, and thyroid (Michibata et al., 2001; Fidock et al., 2002; Lefièvre et al., 2002). In one study, an association was identified between decreased cGMP levels in the cerebral spinal fluid of patients, but not cAMP, and the severity of dementia in AD patients (Hesse et al., 2017). *PDE1A* has also been shown to have a close connection to aging in rodent models (Kelly et al., 2014).

Another interesting candidate target we discovered is rs7431992, an intron in the gene *CACNA2D3*. *CACNA2D3* codes for a subunit of the voltage-gated calcium channel (VGCC) complex and is expressed most abundantly in the brain (Human Protein Atlas; Uhlén et al., 2015; Ablinger et al., 2020). Elevated levels of calcium have been shown to be associated with and exacerbate the symptoms of neurological disorders such as AD (Guan et al., 2021; Ge et al., 2022). More specifically, elevated levels of cellular calcium have been shown to accelerate beta-amyloid peptide aggregation, contributing to the onset of amyloid-associated pathology (Isaacs et al., 2006).

The third target we identified as a potential target for SMAD4 is rs769449, a SNP that falls between exons 2 and 3 of *APOE*. Variations in the *APOE* gene have been determined to be significant genetic risk factors for late-onset AD, and *APOE* gene is one of the most widely studied genes associated with the disease (Yamazaki et al., 2019). APOE is abundantly expressed in the central nervous system, serving the primary function of mediating lipid transport in the brain (Raulin et al., 2022). A variety of studies have identified an association of rs769449 with cognitive decline, as well as levels of tau and Aβ42 (Cruchaga et al., 2013; Zhang and Pierce 2014).

Using EMSA and real-time kinetics measurements, we validated these interactions with SMAD4 and determined their binding affinity values. These experiments confirmed the results that we observed on the arrays. Furthermore, our cell-based luciferase assay provided additional evidence of the functional impact of these allele-specific interactions, suggesting that they may have a direct influence on transcription. While these traditional luciferase assays provided valuable insights, future studies could benefit from implementing massively parallel reporter assays (MPRAs), which offer increased statistical power through simultaneous testing of thousands of sequences and could provide more comprehensive characterization of allele-specific effects (McAfee et al., 2022).

In the final phase of our investigation, we extended our analysis to cellular contexts by generating homozygous and heterozygous cell lines for specific SNPs. We observed that SMAD4's binding preference in these cells was consistent with our in vitro findings. Moreover, the expression of *PDE1A* exhibited allele-dependent changes in expression, providing further evidence of the potential functional consequences of this SNP–TF interaction in a cellular context. Nonetheless, our study also revealed the importance of considering the specific cellular context when interpreting these results. The choice of HEK293t cells, which differ from brain cells in which certain genes are primarily active, underscores the need for further research in physiologically relevant cell types to better understand the consequences of these allele-specific interactions. Future directions for this study will include validation in more physiologically relevant cell types, such as neurons or microglia. Variant screening strategies can also be strengthened by incorporating knockdown studies, particularly through advanced methods like highly multiplexed CRISPRi screening, to evaluate the functional impact of candidate variants prior to cell line generation (Gasperini et al., 2019). Additionally, chromosome conformation capture techniques can reveal the three-dimensional interactions between regulatory elements, providing crucial spatial context. These complementary approaches can help differentiate between direct eQTL effects and those stemming from genetic linkage, improving our ability to identify causal variants.

Our study highlights the power of using TF-WAS to explore the functional implications of GWAS-identified SNPs in AD. By combining bioinformatics, in vitro validation, and cellular experiments, we were able to shed light on some of the intricate molecular mechanisms that may underlie AD susceptibility. Further research in this direction, including screening and validation of a more comprehensive list of SNPs, will undoubtedly contribute to the ongoing efforts to combat this devastating neurodegenerative condition. TF-WAS can further be expanded to proteome-wide association studies (PWAS) using human proteome microarrays (Jeong et al., 2012), allowing for systematic identification of protein–DNA interactions and potential regulatory networks beyond just transcription factors, thus providing a more comprehensive understanding of genetic regulatory mechanisms at the proteome level. This methodology holds great promise for advancing our understanding of not just AD, but many other complex diseases and traits, potentially leading to

the discovery of new therapeutic targets and strategies for the treatment of AD and beyond. A key strength of this methodology is its broad applicability to essentially any complex disease or trait with available GWAS datasets. Our future pursuits will include the profiling of diverse complex diseases, with the aim of generating data that can provide guidance toward clinical interventions across a spectrum of health conditions.

References

Ablinger C, Geisler SM, Stanika RI, Klein CT, Obermair GJ (2020) Neuronal A2δ proteins and brain disorders. Pflugers Arch 472:845–863.

Azevedo MF, Faucz FR, Bimpaki E, Horvath A, Levy I, De Alexandre RB, Ahmad F, Manganiello V, Stratakis CA (2014) Clinical and molecular genetics of the phosphodiesterases (Pdes). Endocr Rev 35:195–233.

Banks CJ, Joshi A, Michoel T (2016) Functional transcription factor target discovery via compendia of binding and expression profiles. Sci Rep 6:1–10.

Barrows JK, Van Dyke MW (2022) Biolayer interferometry for DNA-protein interactions. PLoS One 17:5–9.

Bekris LM, Yu CE, Bird TD, Tsuang DW (2010) Review article: genetics of Alzheimer disease. J Geriatr Psychiatry Neurol 23:213–227.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10:1213–1218.

Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, Mann M (2012) Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet 8:1–8.

Bylino OV, Ibragimov AN, Shidlovskii YV (2020) Evolution of regulated transcription. Cells 9:1–38.

Chen C, Zhou D, Gu Y, Wang C, Zhang M, Lin X, Xing J, Wang H, Zhang Y (2020) SEA version 3.0: a comprehensive extension and update of the super-enhancer archive. Nucleic Acids Res 48:D198–203.

Cruchaga C, et al. (2013) GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. Neuron 78:256–268.

Das P, Golde T (2006) Dysfunction of TGF-β signaling in Alzheimer's disease. J Clin Invest 116:2855–2857.

De Caestecker MP, Hemmati P, Larisch-Bloch S, Ajmera R, Roberts AB, Lechleider RJ (1997) Characterization of functional domains within Smad4/DPC4. J Biol Chem 272:13690–13696.

Dudoit S, Yang YH, Callow MJ, Speed TP (2002) Statistical methods for identifying differentially expressed genes in replicated CDNA microarray experiments. Stat Sin 12:111–139.

Fidock M, Miller M, Lanfear J (2002) Isolation and differential tissue distribution of Two human CDNAs encoding PDE1 splice variants. Cell Signal 14: 53–60.

Gao T, Qian J (2020) Enhanceratlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine Species. Nucleic Acids Res 48:D58–64.

Gasperini M, et al. (2019) A genome-wide framework for mapping gene regulation via cellular genetic screens. Cell 176:377–390.e19.

Ge M, Zhang J, Chen S, Huang Y, Chen W, He L, Zhang Y (2022) Role of calcium homeostasis in Alzheimer's disease. Neuropsychiatr Dis Treat 18: 487–498.

Guan PP, Cao LL, Wang P (2021) Elevating the levels of calcium ions exacerbate Alzheimer's disease via inducing the production and aggregation of β-amyloid protein and phosphorylated tau. Int J Mol Sci 22:1–27.

Han MR, Schellenberg GD, Wang LS (2010) Genome-wide association reveals genetic effects on human Aβ42and τ protein levels in cerebrospinal fluids: a case control study. BMC Neurol 10:90.

Hesse R, et al. (2017) Reduced CGMP levels in CSF of AD patients correlate with severity of dementia and current depression. Alzheimers Res Ther 9:1–11.

Hiew LF, Poon CH, You HZ, Lim LW (2021) Tgf-β/smad signalling in neurogenesis: implications for neuropsychiatric diseases. Cells 10:6.

Hu S, et al. (2009) Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. Cell 139: 610–622.

Hu S, et al. (2013) DNA methylation presents distinct binding sites for human transcription factors. Elife 2013:1–16.

Isaacs AM, Senn DB, Yuan M, Shine JP, Yankne BA (2006) Acceleration of amyloid β-peptide aggregation by physiological concentrations of calcium. J Biol Chem 281:27916–27923.

Jeong JS, et al. (2012) Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. Mol Cell Proteomics 11:1–10.

Jun G, Chung J, Mez J, Barber R, Beecham GW (2017) Transethnic genome-wide scan identifies novel Alzheimer disease loci. Alzheimers Dement 13: 727–738.

Kapoor M, Chinnathambi S (2023) TGF-B1 signalling in Alzheimer's pathology and cytoskeletal reorganization: a specialized tau perspective. J Neuroinflammation 20:1–20.

Kelly MP, Adamowicz W, Bove S, Hartman AJ, Mariga A, Pathak G, Reinhart V, Romegialli A, Kleiman RJ (2014) Select 3',5'-cyclic nucleotide phosphodiesterases exhibit altered expression in the aged rodent brain. Cell Signal 26:383–397.

Landt SG, et al. (2012) ChIP-seq guidelines and practices of the ENCODE and ModENCODE consortia. Genome Res 22:1813–1831.

Lefièvre L, De Lamirande E, Gagnon C (2002) Presence of cyclic nucleotide phosphodiesterases PDE1A, existing as a stable complex with calmodulin, and PDE3A in human spermatozoa. Biol Reprod 67:423–430.

Li S, Kvon EZ, Visel A, Pennacchio LA, Ovcharenko I (2019) Stable enhancers are active in development, and fragile enhancers are associated with evolutionary adaptation. Genome Biol 20:1–17.

Maurano MT, et al. (2012) Systematic localization of common disease-associated variation in regulatory DNA. Science 337:1190–1195.

McAfee JC, Bell JL, Krupa O, Matoba N, Stein JL, Won H (2022) Focus on your locus with a massively parallel reporter assay. J Neurodev Disord 14:50.

Meyers EA, Kessler JA (2017) TGF-β family signaling in neural and neuronal differentiation, development, and function. Cold Spring Harb Perspect Biol 9:1–25.

Michibata H, Yanaka N, Kanoh Y, Okumura K, Omori K (2001) Human Ca2+/calmodulin-dependent phosphodiesterase PDE1A: novel splice variants, their specific expression, genomic organization, and chromosomal localization. Biochim Biophys Acta 1517:278–287.

Myers AJ, et al. (2007) A survey of genetic human cortical gene expression. Nat Genet 39:1494–1499.

Neph S, et al. (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 489:83–90.

Pachkov M, Erb I, Molina N, van Nimwegen E (2007) Swissregulon: a database of genome-wide annotations of regulatory sites. Nucleic Acids Res 35:127–131.

Pluta R, et al. (2022) Molecular basis for DNA recognition by the maternal pioneer transcription factor FoxH1. Nat Commun 13:1–15.

Pratt HE, Andrews GR, Phalke N, Purcaro MJ, Van Der Velde A, Moore JE, Weng Z (2022) Factorbook: an updated catalog of transcription factor motifs and candidate regulatory motif sites. Nucleic Acids Res 50: D141–49.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8:2281–2308.

Raulin AC, Doss SV, Trottier ZA, Ikezu TC, Bu G, Liu CC (2022) Apoe in Alzheimer's disease: pathophysiology and therapeutic strategies. Mol Neurodegener 17:1–26.

Reddy TE, et al. (2012) Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. Genome Res 22: 860–869.

Schott JM, et al. (2016) Genetic risk factors for the posterior cortical atrophy variant of Alzheimer's disease. Alzheimers Dement 12:862–871.

Shen L, et al. (2010) Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. Neuroimage 53:1051–1063.

Sollis E, et al. (2023) The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. Nucleic Acids Res 51:D977–85.

Song M, et al. (2019) Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. Nat Genet 51:1252–1262.

Spivakov M (2014) Spurious transcription factor binding: non-functional or genetically redundant? Bioessays 36:798–806.

Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci U S A 90: 1977–1981.

Strober BJ, Wen X, Wucher V, Kwong A, Lappalainen T, Li X, Liang Y (2020) The GTEx consortium atlas of genetic regulatory effects across human tissues. Science 18:1318–1330.

Szklarczyk D, et al. (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 51:D638–46.

Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, Mi H (2022) PANTHER: making genome-scale phylogenetics accessible to all. Protein Sci 31:8–22.

Uhlén M, et al. (2015) Tissue-based map of the human proteome. Science 347:1–9.

Van Cauwenberghe C, Van Broeckhoven C, Sleegers K (2016) The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med 18:421–430.

Wan R, Feng J, Tang L (2021) Consequences of mutations and abnormal expression of Smad4 in tumors and t cells. OncoTargets Ther 14:2531–2540.

Wang Y, Xue Q, Zheng Q, Jin Y, Shen X, Yang M, Zhou X, Li Y (2021) SMAD4 mutation correlates with poor prognosis in non-small cell lung cancer. Lab Invest 101:463–476.

Webster JA, et al. (2009) Genetic control of human brain transcript expression in Alzheimer disease. Am J Hum Genet 84:445–458.

Weidemüller P, Kholmatov M, Petsalaki E, Zaugg JB (2021) Transcription factors: bridge between cell signaling and gene regulation. Proteomics 21:1–14.

Yamazaki Y, Zhao N, Caulfield TR, Liu CC, Bu G (2019) Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. Nat Rev Neurol 15:501–518.

Yan J, et al. (2021) Systematic analysis of binding of transcription factors to noncoding variants. Nature 591:147–151.

Zhang C, Pierce BL (2014) Genetic susceptibility to accelerated cognitive decline in the US health and retirement study. Neurobiol Aging 35: 1512.e11–1512.e18.

Zhou Y, et al. (2022) Molecular landscapes of human hippocampal immature neurons across lifespan. Nature 607:527–533.