

Region-based analysis of rare genomic variants in whole-genome sequencing datasets reveal two novel Alzheimer's disease-associated genes: DTNB and DLG2

— [Source link](#) 

Dmitry Prokopenko, Sanghun Lee, Julian Hecker, Kristina Mullin ...+9 more authors

Institutions: Harvard University, Dankook University, Brigham and Women's Hospital, Beth Israel Deaconess Medical Center ...+2 more institutions

Published on: 12 Jun 2021 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: Minor allele frequency, Whole genome sequencing and Genetic association

Related papers:

- [Functional mapping of the human genome by cDNA localization versus sequencing](#)
- [Sequences of human repetitive DNA, non-alpha-globin genes, and major histocompatibility locus genes.](#)
- [Preferential integration of human immunodeficiency virus type 1 into genes, cytogenetic R bands and GC-rich DNA regions: insight from the human genome sequence.](#)
- [Diversity of human genome](#)
- [VarCards: an integrated genetic and clinical database for coding variants in the human genome](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/region-based-analysis-of-rare-genomic-variants-in-whole-15vxafqdz>

Region-based analysis of rare genomic variants in whole-genome sequencing datasets reveal two novel Alzheimer's disease-associated genes: *DTNB* and *DLG2*

Dmitry Prokopenko^{1,2}, Sanghun Lee^{3,4}, Julian Hecker^{2,5}, Kristina Mullin¹, Sarah Morgan^{2,6}, Yuriko Katsumata^{7,8}, Alzheimer's Disease Neuroimaging Initiative (ADNI)*, Michael W. Weiner⁹, David W. Fardo^{7,8}, Nan Laird⁴, Lars Bertram^{10,11}, Winston Hide^{2,6}, Christoph Lange^{4#}, Rudolph E. Tanzi^{1,2#}

(1) Genetics and Aging Research Unit and The Henry and Allison McCance Center for Brain Health, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

(2) Harvard Medical School, Boston, MA, USA

(3) Department of medical consilience, Graduate school, Dankook University, South Korea

(4) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

(5) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA

(6) Department of Pathology, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA, USA

(7) Department of Biostatistics, University of Kentucky, Lexington, KY, USA

(8) Sanders-Brown Center on Aging, University of Kentucky, Lexington, KY, USA

(9) Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA, USA

(10) Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and Cardiogenetics, University of Lübeck, Germany

(11) Department of Psychology, University of Oslo, Oslo, Norway

#These authors jointly supervised this work.

* Data used in preparation of this article were in part obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Correspondence:

Christoph Lange, clange@hsph.harvard.edu;

Rudolph E. Tanzi, tanzi@helix.mgh.harvard.edu

Abstract

Alzheimer's disease (AD) is a genetically complex disease for which roughly 30 genes have been identified via genome-wide association studies. We attempted to identify rare variants (minor allele frequency <0.01) associated with AD in a region-based, whole genome sequencing (WGS) association study (GSAS) of two independent AD family datasets (NIMH/NIA; 2247 individuals; 605 families). Employing a sliding window approach across the genome, we identified several regions that achieved p-values $< 10^{-6}$, using the burden test or the SKAT statistic. The genomic region around the dystrobrevin beta (*DTNB*) gene was identified with the burden test and replicated in case/control samples from the ADSP study ($p_{\text{meta}} = 4.74 \times 10^{-8}$). SKAT analysis revealed region-based association around the discs large homolog 2 (*DLG2*) gene and replicated in case/control samples from the ADSP study ($p_{\text{meta}} = 1 \times 10^{-6}$). Here, in a region-based GSAS of AD we identified two novel AD genes, *DLG2* and *DTNB*, based on association with rare variants.

Introduction

Alzheimer's disease (AD) is a heterogeneous, genetically complex neurodegenerative disorder¹. Genome-wide association studies (GWAS) have identified common variants in roughly 30 genes associated with AD^{2,3}. GWAS heritability is estimated to be 24-33%^{4,5} - less than a half of the heritability calculated from twin studies¹. Identification of rare variants associated with AD may help explain the missing heritability, and lead to new biological insights⁶. Several rare variant loci previously associated with AD⁷, including *TREM2*⁸, have been identified with whole-exome sequencing (WES) studies⁹.

Identification of association signals that are driven by rare variants remains cumbersome due to low power and relatively small sample sizes. Hence, aggregation methods, such as burden tests^{10,11} and variance component tests (SKAT)^{12,13}, have been developed to jointly test regions of rare variants for association. Combining variant data increases the association signal and reduces the number of statistical tests. While burden tests are most powerful for signals with consistent effect directions, SKAT is more powerful for signals with different effect directions or when the fraction of causal variants is small. Previously, aggregated gene-based association analyses have been successful in identifying novel exome-wide significant associations with sporadic AD¹⁴.

Recently, a general framework for exact region-based association testing in family-based designs has been developed¹⁵. Using the proposed region-based testing framework, we performed a GSAS combining two AD family-based cohorts (605 families; 1509 affecteds; 738 unaffecteds) focusing on rare variants. For replication, we used case/control subjects from NIA ADSP, which included WGS data from a Non-Hispanic White (NHW) subcohort (983 cases; 686 controls), an African American (AA) subcohort (450 cases; 501 controls), and a Hispanic (HISP) subcohort (486 cases; 613 controls).

Using a p-value cutoff of 5×10^{-6} , the burden test and SKAT identified several genomic regions showing association with AD risk. A region identified by the burden test in the *DTNB* gene ($p = 7 \times 10^{-8}$) was replicated in the NHW samples. SKAT analysis revealed an association with variants encompassing a region around *DLG2* ($p = 4 \times 10^{-6}$), which replicated in the NHW and the AA samples.

Results

In a region-based whole-genome sequencing association study (GSAS) focusing on rare genomic variants, we combined two AD family-based cohorts, the NIMH Alzheimer's disease genetics initiative

study (NIMH) and the family component of the NIA ADSP sample. The combined sample consisted of 1509 affected and 738 unaffected siblings in families of predominantly European ancestry (Supplementary Table 1, Methods). 8,011,126 variants passed strict quality control and allele frequency (AF) filter of $\leq 1\%$ (gnomAD¹⁶). We grouped rare variants into consecutive regions/windows of ten variants and performed a sliding-window rare variant WGS scan over the whole genome (801,124 windows). We employed a recently developed framework for exact regional-based analysis within FBAT¹⁵ to analyze these sets of rare variants using both the Burden test and SKAT. These tests are able to detect different configurations of disease regions - dense regions with the same effect directions (Burden test) or less dense signals with varying effect directions (SKAT).

Since we restricted our analysis to rare variants (i.e. AF < 0.01) and given our modest sample size in the family-discovery cohort, we have used a relatively liberal p-value threshold $p < 5 \times 10^{-6}$ to identify “suggestive associations” by burden test or SKAT. A stricter Bonferroni-corrected significance threshold would be $p = 6.24 \times 10^{-8}$. Seven loci exhibited suggestive evidence for association with AD risk (Figure 1, Supplementary Figure 1, Table 1). For replication analysis, we selected the unrelated, multiethnic WGS AD subset from the NIA ADSP dataset (Methods). This dataset consists of three subpopulations: NHW (n=1669), AA (n=951), HISP (n=1099) (Sample sizes after quality control; Supplementary Table 1). A region located downstream to *DTNB*, with a Burden p-value of 7×10^{-8} in the discovery dataset, showed a burden p-value of 0.0324 in ADSP NHW (Table 1 and Supplementary Table 2). Another region, located in an intron of *DLG2* with a SKAT p-value of 4×10^{-6} in the discovery family-based dataset, showed replication with a significant SKAT p-value of 0.0143 in the ADSP NHW dataset and a SKAT p-value of 0.053 in the ADSP AA dataset (Table 1 and Supplementary Table 3).

Both *DLG2* and *DTNB* are highly expressed in the brain based on RNA-data from three different sources: Internally generated Human Protein Atlas (HPA) RNA-seq data, RNA-seq data from the Genotype-Tissue Expression (GTEx) project, and CAGE data from the FANTOM5 project, as well as the consensus dataset for each gene derived from the Human Protein Atlas¹⁷ (Supplementary figures 2 and 3). In the Alzheimer's Disease Dataset analysis¹⁸ (GSE48350) from the GEO database¹⁹ expression of *DLG2* and *DTNB* is significantly decreased in AD compared to control subjects (Supplementary table 4).

Network analysis revealed a network of 33 proteins interacting with *DLG2* and *DTNB* that were enriched for neuronal synaptic functions (Supplementary Figure 4). Functional enrichment of the subnetwork of proteins directly interacting with *DLG2* and *DTNB* revealed 694 enriched GO process/ pathway terms (Supplementary table 5). The most enriched part of the network was for proteins interacting with *DLG2* that are connected to neurexins and neuroligins, as well as trafficking of AMPA receptors. *DLG2* also interacted with 4 proteins (*NOS1*, *ERBB4*, *DLGAP2*, *NRXN3*) previously associated with AD risk²⁰, and 4 proteins (*GRIN1*, *GRIN2A*, *GRIN2B*, *GAPDH*) associated with AD in the KEGG Alzheimer's pathway. *DLG2* and *DTNB* also share protein-protein or co-expression interactions through *KIF1B*, *MLC1*, and *SH3D19*.

Discussion

Here, we describe a comprehensive region-based analysis of Alzheimer's disease using WGS datasets. We specifically searched for novel AD association signals driven by regions of rare variants in a large family-based cohort. To account for different disease region specifications, we employed both the burden test and SKAT. This yielded seven regions of suggestive evidence ($p < 5 \times 10^{-6}$) for association with AD risk in the family datasets. These results were followed up with replication analysis in independent

case-control samples of different ethnicities. Two loci, *DTNB* and *DLG2*, showed consistent evidence of replication in at least one of the subpopulations (i.e. NHW). The *DLG2* region was also confirmed in the African American sample.

DLG2 encodes a member of the membrane-associated guanylate kinase family, also known as post-synaptic density protein, PSD-93. Down-regulation of synaptic scaffolding proteins, including *DLG2*, has been described as an early event in AD²¹. *DLG2* has been proposed as a potential target for AD based on an integrated metabolomics-genetics-imaging systems approach in Agora (URLs); agonism of *DLG2* is predicted to reduce disease progression. An expression dataset of AD in the GEO database revealed reduced expression of *DLG2* in AD versus controls. A common variant in *DLG2*, rs683250, was previously associated with increases of shape asymmetry in controls as compared demented populations²². This same variant is in linkage disequilibrium (LD, $D'=1$) with all rare variants of the *DLG2* region found to be associated with AD here. *DLG2* variant, rs286043 (AF=0.03), which exhibited suggestive evidence for association with AD risk in IGAP ($p=5e-06$), is in LD with 4 out of 10 variants from our *DLG2* AD-associated region, suggesting possible allelic heterogeneity. *DLG2* has previously been associated with schizophrenia²³ and autism^{24,25}. Along these lines, *DLG2* deficiency in mice has been reported to lead to reduced sociability and increased repetitive behavior along with aberrant synaptic transmission in the dorsal striatum²⁶.

β -Dystrobrevin (*DTNB*) is associated with neurons in the cortex, hippocampus, and cerebellum, and has also been reported to be enriched in the post-synaptic density²⁷. Kinesin superfamily motor proteins (KIF) are responsible for anterograde protein transport within the axon of various cellular cargoes, including synaptic and structural proteins²⁸. Dysregulated KIF expression has also been associated with early AD pathology²⁹, and β -Dystrobrevin interacts directly with kinesin heavy chain in the brain³⁰. Dystrobrevin-binding protein 1, also known as dysbindin, has been reported to be associated with schizophrenia^{31,32}. Thus, both novel AD gene candidates identified in this study have been associated with post-synaptic function and schizophrenia.

Our approach utilized two region-based tests (burden and SKAT) in a family-based design, in which the joint distribution of rare variants is not estimated, but rather obtained by the haplotype algorithm for FBAT, which is robust against population structure and admixture, and allows for construction of exact or simulation-based p-values. Previously, we performed region-based rare variant testing, but with different region definitions, and using only burden tests with empirical estimation of the variant correlations and asymptotic p-values²⁰. While this is the largest combined family-based AD-specific WGS dataset available, larger datasets will be needed to confirm our findings in future studies. We also note that by utilizing a window size of 10 consecutive variants, we could have missed sparsely distributed signals. Since the number of possible haplotypes increases exponentially with the number of variants tested, larger window sizes were computationally infeasible.

In summary, we identified two novel loci associated with AD, based on association with rare variants in *DLG2* and *DTNB* in a family-based AD WGS sample using methods that are robust to population structure. We further showed replication in an independent multi-ethnic AD WGS dataset with unrelated cases and controls. These findings demonstrate the usefulness of WGS in capturing non-exonic, rare variant signals. Both novel AD-associated genes identified here encode post-synaptic density proteins and have been implicated for roles in schizophrenia.

Acknowledgements

This work was supported by Cure Alzheimer's Fund and NIH R56AG057191 (D.W.F. and Y.K.). The computations in this paper were run in part on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University with support from John Morrissey and in part on compute provided by Dell HPC Research Computing Solutions with support by Glen Otero. The funding body has no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Please refer to the Supplementary Note for full acknowledgements.

Author Contributions

D.P., C.L., R.E.T. contributed to the study concept and design. D.P., S.L., K.M., S.M., Y.K., D.W.F., L.B., W.H., C.L., R.E.T. contributed to data analysis and/or interpretation. J.H., N.L., C.L. contributed to statistical support. D.P., S.L., W.H., C.L. and R.E.T. wrote the original draft of the paper, and all authors critically reviewed the manuscript.

Competing Interests statement

The authors declare no competing interests.

Methods

Study populations

Discovery family-based dataset

Our discovery dataset consisted of two WGS family-based cohorts: the National Institute of Mental Health (NIMH) family AD cohort³³ and families from the National Institute of Aging Alzheimer's Disease Sequencing Project³⁴ (NIA ADSP). Whole-genome sequencing and variant calling in NIMH are described elsewhere³⁵. Variant calls for the families from the NIA ADSP cohort were obtained from the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) under accession number: NG00067. Both cohorts consisted of multiplex AD families with affected and unaffected siblings (Supplementary table 1). A subject was considered to be affected if he/she was included in one of the following categories: "Definite AD", "Probable AD" or "Possible AD". Unaffected subjects had either no dementia, suspected dementia (46 subjects), or non-AD dementia (10 subjects). It is important to note that NIA ADSP families by design did not include individuals with two APOE- ϵ 4 alleles. After standard quality control, both cohorts were merged together.

NIA ADSP case-control dataset

WGS variant calls for the NIA ADSP replication case-control dataset were obtained from the NIAGADS under accession number: NG00067 and consisted of the ADSP Discovery-Extension Case-Control WGS dataset³⁴ and the ADNI Case-Control WGS dataset. Samples were remapped to GRCh38 and jointly called with the families from the NIA ADSP cohort. Full details can be found on NIAGADS (<https://dss.niagads.org/datasets/ng00067/>) and elsewhere³⁶. Briefly, a subject was considered affected, if he/she met the NINCDS-ADRDA criteria for possible, probable, or definite AD, had documented age at onset or age at death (for pathologically verified cases), and APOE genotyping. All controls were 60 or more years old and were free of dementia.

Quality control

Briefly, we have excluded individuals based on genotyping rate, inbreeding coefficient, and family mismatches using identity by descent (IBD) sharing coefficients. After sample-based quality control, we have combined two WGS family-based cohorts NIMH (1,393 individuals in 446 families) and 854 individuals (families from NIA ADSP; 159 families). In the merged dataset we excluded multiallelic variants, monomorphic variants, singletons (i.e. variants with only one alternative allele across the dataset and variants seen only in one family), indels, and variants which had one missing allele among 2 alleles in an individual. The remaining variants were filtered based on Mendel errors, genotyping rate (95%), Hardy-Weinberg equilibrium ($p < 1e-08$), calling quality in TOPMed, which is a large WGS database with >100,000 individuals sequenced jointly, and allele frequency in gnomAD (AF ≤ 1% in either whole gnomAD or nonFinnish European sample).

WGS regional-based analysis

We have performed a whole genome scan for our combined family-based AD dataset using a newly developed exact framework in FBAT for region-based association testing¹⁵. We grouped rare variants in consecutive sets of ten. For each set of rare variants, we considered the burden test and the SKAT test using Affection Status minus offset as phenotype. We selected an offset of 0.15 which approximately corresponds to the population prevalence of AD. We have used FBAT³⁷, R³⁸, snakemake³⁹ and bash commands to implement and run the described analyses.

Replication

Replication significance level was set to 0.0143 (0.1 divided by 7 independent loci). Regions/windows with $P < 0.05$ were also reported as replicated. We have used the SKAT package to perform Burden and SKAT-O tests on the same sets of rare variants in the case-control replication cohorts. As covariates, we used sequencing center, age, sex, and principal components (to account for population structure). Principal components were calculated based on rare variants using the Jaccard index⁴⁰. We have also performed meta-analysis among datasets with similar ethnical background using the Fisher's combined probability test.

RNA-Seq and microarray analysis

We explored *DLG2* and *DTNB* genes' expression based on the Human Protein Atlas (HPA) RNA-seq data (<https://www.proteinatlas.org>) and tested for differential expression of synaptic and immune related genes including *DLG2* and *DTNB* genes between normal controls (N=173, aged 20-99 yrs) and AD cases (N=80) in the brain regions including hippocampus, entorhinal cortex, superior frontal cortex, and post-central gyrus using microarray dataset GSE48350, which is available from the Gene Expression Omnibus Web site (<http://www.ncbi.nlm.nih.gov/geo/>). Differential expression was tested using the "GEO2R" tool.

Network construction

We used Cytoscape 3.8.0 and the StringDB protein-protein interaction resource⁴¹ using only identified protein-protein interactions. Using a background that agglomerates protein-protein interaction datasets, we seeded the network with *DLG2* and *DTNB* and identified direct associations between proteins and *DLG2* and *DTNB* in a global network (supplementary table 5). Results were combined using the Genemania server (Utilizing significantly co-expressed genes across several experimental datasets)⁴²

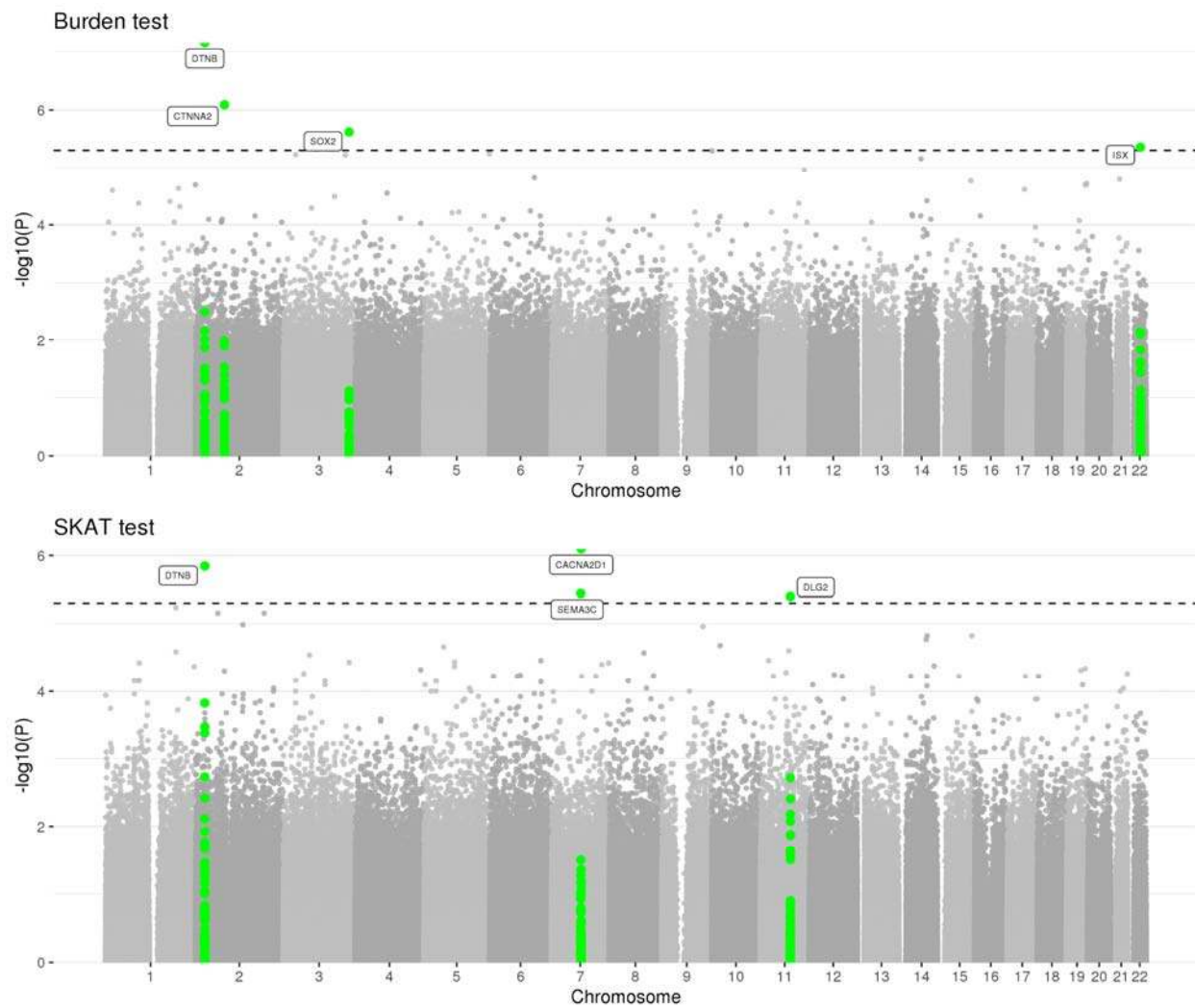
to further capture functional relationships and to build a combined protein-protein/gene co-expression network.

Functional enrichment

Functional enrichment within the network was performed using the remote StringDB server linked to Cytoscape “String App Enrichment function”⁴³, producing enrichments using the hypergeometric test, with P -values corrected for multiple testing using the method of Benjamini and Hochberg in known molecular pathways and GO terms as described in Franceschini *et al.*⁴⁴

Figures

Figure 1: Manhattan plots of sets of rare variants in the whole genome scan of the family-based discovery dataset using the burden and SKAT test. Dashed line corresponds to suggestive threshold of 5×10^{-6} .



Tables

Table 1: Top regions based on the burden or SKAT test with $p \leq 5e-06$ in the discovery family-based dataset using whole genome scan

First SNV in the region (chr:pos:ref:alt)	Last SNV in the region (chr:pos:ref:alt)	Nearest protein-coding gene	Discovery dataset (NIMH + NIA families)		Replication dataset NHW ADSP		Meta-analysis of family-based discovery and NHW ADSP replication datasets		Replication dataset AA ADSP		Replication dataset HISP ADSP	
			P-value	nsim	P-value	Number of SNVs in the region	Fisher chi-squared test statistic	P-value	P-value	Number of SNVs in the region	P-value	Number of SNVs in the region
Burden test												
2:25703040:G:A	2:25707419:T:G	DTNB	7.00E-08	1.00E+08	0.032	5	39.808	4.74E-08	0.390	9	0.353	8
2:79854141:T:G	2:79856252:C:T	CTNNA2	8.10E-07	1.00E+08	0.799	6	28.500	9.88E-06	0.774	8	0.785	7
3:181942653:A:G	3:181946475:G:A	SOX2	2.40E-06	1.00E+07	0.317	6	28.177	1.15E-05	0.619	9	0.163	8
22:35048628:G:C	22:35053269:C:T	ISX	4.40E-06	1.00E+07	0.766	8	25.201	4.58E-05	0.827	9	0.014	10
Variance component (SKAT) test												
11:83498255:A:G	11:83500398:T:G	DLG2	4E-06	1.00E+07	0.014	5	33.352	1E-06	0.053	8	0.893	9
2:25703040:G:A	2:25707419:T:G	DTNB	1.4E-06	1.00E+08	0.054	5	32.737	1.4E-06	0.409	9	0.455	8
7:82268137:T:C	7:82271095:A:T	CACNA2D1	8E-07	1.00E+07	0.591	5	29.131	7.4E-06	0.278	8	0.090	8
7:81141368:T:G	7:81143780:C:T	SEMA3C	3.6E-06	1.00E+07	0.251	5	27.832	1.4E-05	0.046	9	0.694	10

SNV - single nucleotide variant, NHW - non Hispanic white, AA - African-American, HISP - Hispanic

URLs

FBAT, <https://sites.google.com/view/fbatwebpage>; gnomAD, <https://gnomad.broadinstitute.org/>; Agora AMP-AD, <https://agora.ampadportal.org/genes>; TOPMED, <https://www.nhlbiwgs.org/>; Human Protein Atlas, <https://www.proteinatlas.org/>; GEO database, <https://www.ncbi.nlm.nih.gov/geo/>; NIAGADS, <https://www.niagads.org/>; StringDB, <https://string-db.org/>.

References

1. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).
2. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
3. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau, immunity and lipid processing. *Nat. Genet.* **51**, 414 (2019).
4. Lee, S. H. *et al.* Estimation and partitioning of polygenic variation captured by common snps for alzheimer's disease, multiple sclerosis and endometriosis. *Hum. Mol. Genet.* **22**, 832–841 (2013).
5. Ridge, P. G., Mukherjee, S., Crane, P. K. & Kauwe, J. S. K. Alzheimer's disease: Analyzing the missing heritability. *PLoS One* **8**, 1–10 (2013).
6. Manolio, T. a *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
7. Grozeva, D., Saad, S., Menzies, G. E. & Sims, R. Benefits and Challenges of Rare Genetic Variation in Alzheimer's Disease. *Curr. Genet. Med. Rep.* **7**, 53–62 (2019).
8. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116 (2013).
9. Lill, C. M. *et al.* The role of TREM2 R47H as a risk factor for Alzheimer's disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's disease. *Alzheimer's Dement.* **11**, 1407–1416 (2015).
10. Madsen, B. E. & Browning, S. R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5**, e1000384 (2009).
11. Li, B. & Leal, S. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* 311–321 (2008). doi:10.1016/j.ajhg.2008.06.024.
12. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
13. Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* **91**, 224–37 (2012).
14. Bis, J. C. *et al.* Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Mol. Psychiatry* (2018). doi:10.1038/s41380-018-0112-7
15. Hecker, J. *et al.* A unifying framework for rare variant association testing in family-based designs, including higher criticism approaches, SKATs, and burden tests. *Bioinformatics* 1–7 (2020). doi:10.1093/bioinformatics/btaa1055
16. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

17. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* (80-.). **347**, (2015).
18. Berchtold, N. C. *et al.* Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol. Aging* **34**, 1653–1661 (2013).
19. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Res.* **41**, 991–995 (2013).
20. Prokopenko, D., Morgan, S. L., Mullin, K., Hofmann, O. & Chapman, B. Whole- genome sequencing reveals new Alzheimer ' s disease -associated rare variants in loci related to synaptic function and neuronal development. *medRxiv* (2020).
21. Xu, J. *et al.* Regional protein expression in human Alzheimer's brain correlates with disease severity. *Commun. Biol.* doi:10.1038/s42003-018-0254-9
22. Wachinger, C., Nho, K., Saykin, A. J., Reuter, M. & Rieckmann, A. A Longitudinal Imaging Genetics Study of Neuroanatomical Asymmetry in Alzheimer's Disease. *Biol. Psychiatry* **84**, 522–530 (2018).
23. Ingason, A. *et al.* Expression analysis in a rat psychosis model identifies novel candidate genes validated in a large case-control sample of schizophrenia. *Transl. Psychiatry* **5**, e656 (2015).
24. Egger, G. *et al.* Identification of risk genes for autism spectrum disorder through copy number variation analysis in Austrian families. *Neurogenetics* **15**, 117–127 (2014).
25. Ruzzo, E. K. *et al.* Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850-866.e26 (2019).
26. Yoo, T. *et al.* A DLG2 deficiency in mice leads to reduced sociability and increased repetitive behavior accompanied by aberrant synaptic transmission in the dorsal striatum. *Mol. Autism* **11**, 1–14 (2020).
27. Blake, D. J., Hawkes, R., Benson, M. A. & Beesley, P. W. Different dystrophin-like complexes are expressed in neurons and glia. *J. Cell Biol.* **147**, 645–657 (1999).
28. Hirokawa, N. & Noda, Y. Intracellular transport and kinesin superfamily proteins, KIFs: Structure, function, and dynamics. *Physiol. Rev.* **88**, 1089–1118 (2008).
29. Andersson, M. E. *et al.* Kinesin gene variability may affect tau phosphorylation in early Alzheimer's disease. *Int. J. Mol. Med.* **20**, 233–239 (2007).
30. Macioce, P. *et al.* β -Dystrobrevin interacts directly with kinesin heavy chain in brain. *J. Cell Sci.* **116**, 4847–4856 (2003).
31. Baek, J. H. *et al.* Association of genetic variations in DTNBP1 with cognitive function in schizophrenia patients and healthy subjects. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **159 B**, 841–849 (2012).
32. Yang, Y. *et al.* Association of DTNBP1 With Schizophrenia: Findings From Two Independent Samples of Han Chinese Population. *Front. Psychiatry* **11**, 1–9 (2020).
33. Blacker, D. *et al.* ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative. *Neurology* **48**, 139–147 (1997).
34. Beecham, G. W. *et al.* The Alzheimer's Disease Sequencing Project: Study design and sample

- selection. *Neurol. Genet.* **3**, e194 (2017).
35. Prokopenko, D. *et al.* Identification of Novel Alzheimer's Disease Loci Using Sex-Specific Family-Based Association Analysis of Whole-Genome Sequence Data. *Sci. Rep.* **10**, 1–9 (2020).
 36. Leung, Y. Y. *et al.* VCPA: Genomic variant calling pipeline and data management tool for Alzheimer's Disease Sequencing Project. *Bioinformatics* **35**, 1768–1770 (2019).
 37. Laird, N., Horvath, S. & Xu, X. Implementing a unified approach to family-based tests of association. *Genet. Epidemiol.* **19**, (2000).
 38. Team, R. C. R: A Language and Environment for Statistical Computing. Available at: <https://www.r-project.org>.
 39. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
 40. Prokopenko, D. *et al.* Utilizing the Jaccard index to reveal population stratification in sequencing data: A simulation study and an application to the 1000 Genomes Project. *Bioinformatics* **32**, 1366–1372 (2016).
 41. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
 42. Warde-Farley, D. *et al.* The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, 214–220 (2010).
 43. Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* **18**, 623–632 (2019).
 44. Franceschini, A. *et al.* STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, 808–815 (2013).