

Synonymous variants associated with Alzheimer disease in multiplex families

Min Tang, PhD, Maria Eugenia Alaniz, PhD, Daniel Felsky, PhD, Badri Vardarajan, PhD, Dolly Reyes-Dumeyer, BA, Rafael Lantigua, MD, Martin Medrano, MD, David A. Bennett, MD, Philip L. de Jager, MD, PhD, Richard Mayeux, MD, MSc, Ismael Santa-Maria, PhD, and Christiane Reitz, MD, PhD

Correspondence
Dr. Reitz
cr2101@cumc.columbia.edu

Neurol Genet 2020;6:e450. doi:10.1212/NXG.000000000000450

Abstract

Objective

Synonymous variants can lead to disease; nevertheless, the majority of sequencing studies conducted in Alzheimer disease (AD) only assessed coding variation.

Methods

To detect synonymous variants modulating AD risk, we conducted a whole-genome sequencing study on 67 Caribbean Hispanic (CH) families multiply affected by AD. Identified disease-associated variants were further assessed in an independent cohort of CHs, expression quantitative trait locus (eQTL) data, brain autopsy data, and functional experiments.

Results

Rare synonymous variants in 4 genes (*CDH23*, *SLC9A3R1*, *RHBDD2*, and *ITIH2*) segregated with AD status in multiplex families and had a significantly higher frequency in these families compared with reference populations of similar ancestry. In comparison to subjects without dementia, expression of *CDH23* ($\beta = 0.53$, $p = 0.006$) and *SLC9A3R1* ($\beta = 0.50$, $p = 0.02$) was increased, and expression of *RHBDD2* ($\beta = -0.70$, $p = 0.02$) decreased in individuals with AD at death. In line with this finding, increased expression of *CDH23* ($\beta = 0.26 \pm 0.08$, $p = 4.9E-4$) and decreased expression of *RHBDD2* ($\beta = -0.60 \pm 0.12$, $p = 5.5E-7$) were related to brain amyloid load ($p = 0.0025$). *SLC9A3R1* expression was associated with burden of TDP43 pathology ($\beta = 0.58 \pm 0.17$, $p = 5.9E-4$). Using eQTL data, the *CDH23* variant was in linkage disequilibrium with variants modulating *CDH23* expression levels (top single nucleotide polymorphism: rs11000035, $p = 4.85E-6$, $D' = 1.0$). Using minigene splicing assays, the *CDH23* and *SLC9A3R1* variants affected splicing efficiency.

Conclusions

These findings suggest that *CDH23*, *SLC9A3R1*, *RHBDD2*, and possibly *ITIH2*, which are involved in synaptic function, the glutamatergic system, and innate immunity, contribute to AD etiology. In addition, this study supports the notion that synonymous variants contribute to AD risk and that comprehensive scrutinization of this type of genetic variation is warranted and critical.

From the The Taub Institute for Research on Alzheimer's Disease and the Aging Brain (M.E.A., B.V., R.L., P.L.J., R.M., I.S.-M., C.R.); The Gertrude H. Sergievsky Center (M.T., D.R.-D., R.L., R.M., C.R.); Department of Neurology (P.L.J., R.M., C.R.); Department of Epidemiology (R.M., C.R.); Department of Psychiatry (R.M.), Columbia University, New York; Department of Pathology and Cell Biology (M.E.A., I.S.-M.), Columbia University, New York; Rush Alzheimer's Disease Center (D.A.B.); Department of Neurological Sciences (D.A.B.); Department of Pathology (D.A.B.), Rush University Medical Center, Chicago, IL; Center for Innovation in Brain Science, Departments of Pharmacology and Neurology, University of Arizona College of Medicine (M.T.), Tucson; Department of Medicine (R.L.), College of Physicians and Surgeons, Columbia University, New York, NY; School of Medicine (M.M.), Mother and Teacher Pontifical Catholic University, Santiago, Dominican Republic; and The Krembil Centre for Neuroinformatics (D.F.), Centre for Addiction and Mental Health, Toronto, Ontario, Canada.

Go to [Neurology.org/NG](https://www.neurology.org/NG) for full disclosures. Funding information is provided at the end of the article.

The Article Processing Charge was funded by the NIH.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Glossary

ABI = Applied Biosystems; **AD** = Alzheimer disease; **CH** = Caribbean Hispanic; **DLPFC** = dorsolateral prefrontal cortex; **EFIGA** = Estudio Familiar de Influencia Genética en Alzheimer; **ExAC** = Exome Aggregation Consortium; **GWAS** = genome-wide association study; **IAIP** = inter-alpha inhibitor protein; **MAF** = minor allele frequency; **NHERF1** = Na⁺/H⁺ exchanger regulatory factor 1; **RADC** = Rush Alzheimer's Disease Center.

Alzheimer disease (AD), the most frequent cause of dementia, is a major public health burden, affecting currently 5.1 million persons in the United States and posing a projected burden of over 13.8 million affected individuals by the year 2050.¹ Genome-wide association studies (GWASs) conducted over the past 15 years have identified over 25 common variants with low effect sizes (odds ratios ranging from 1.1 to 1.25) at over 25 loci containing genes predominantly clustering in inflammation/immune response, lipid metabolism, and endocytosis/trafficking pathways.² Recent studies using next-generation sequencing technologies identified in addition rare risk or protective variants in a set of genes, with odds ratios more similar to that of 1 APOEε4 allele, which also largely cluster in these pathways.^{3–6}

Following the traditional dogma of molecular biology postulating that only nonsynonymous mutations, i.e., variants that alter the encoded amino acid, have an effect on the protein sequence, and thereby cellular function, most of the sequencing studies conducted have only assessed coding variants resulting in amino acid substitutions, usually identified via an ensemble of commonly applied annotations and scores predicting effects based on amino acid substitutions and conservation. However, although previously called silent mutations, it is now recognized that synonymous variants can have deleterious effects and cause disease by altering splicing, translation, messenger RNA structure, protein folding, protein expression, and enzymatic activity.⁷ Identification of these disease-associated synonymous variants is not only critical to explain the missing heritability of disease but can also significantly advance the understanding of the underlying molecular disease etiology and help identify preventive and therapeutic targets.

To identify rare, synonymous variants associated with AD, we used whole-genome sequencing on 67 Caribbean Hispanic (CH) families highly loaded for AD. The frequency of AD in these families is approximately 5 times more than in non-Hispanic white individuals of similar age, significantly increasing statistical power for rare variant identification.⁸ Putative disease-associated variants were validated in an independent data set of CH subjects, expression quantitative trait locus (eQTL) and brain autopsy data from the ROS/MAP study, and functional analyses determining the effect of select identified variants on splicing.

Methods

Samples selected for whole-genome sequencing

The 67 families included in the whole genome sequencing (WGS) are part of the Estudio Familiar de Influencia Genética

en Alzheimer (EFIGA) cohort that has been described elsewhere in detail.⁹ EFIGA study participants were recruited since January 1998 from the Alzheimer Disease Research Center Memory Disorders Clinic at Columbia University in New York City and hospitals in the Dominican Republic and Puerto Rico. Diagnosis of individuals with AD was validated by standardized neurologic and neuropsychological evaluations. Assessment of family history was conducted to determine existence of additional living relatives with AD; if existence of affected siblings was confirmed, all other living relatives were assessed via standardized medical, neurologic, and neuropsychological examinations. All study participants were followed up in 18-month intervals completing at each visit a standardized assessment of medical history, physical and neurologic examination, and an extensive neuropsychological battery¹⁰ assessing cognitive function in key domains affected by aging and dementia, including memory, visuospatial function, executive function, and psychomotor speed, applying the Selective Reminding Test, the Benton Visual Retention Test, the Rosen Drawing Test, the Boston Naming Test, the Controlled Oral Word Association Test, the Category Fluency Test, the Color Trails Test, the Similarities subtest from the Wechsler Adult Intelligence Scale, and the orientation items from the Mini-Mental State Examination. Functional status was evaluated using the Disability and Functional Limitation Instrument. Severity of cognitive impairment was assessed using The Clinical Dementia Rating Scale. Diagnosis of AD was made based on National Institute of Neurological and Communicative Disorders and Stroke—the Alzheimer Disease and Related Disorders Association guidelines in a consensus conference consisting of physicians and neuropsychologists. APOE genotypes were derived by a modification of the protocol described by Hixson and Vernier.¹¹ To optimize ability to identify novel sequence variants in the WGS analyses, priority was given to families most heavily affected by AD (≥4 affected members with DNA available) and lowest frequency of the APOEε4 allele. On average, 5 individuals per family underwent WGS.

Preparation of samples for whole-genome sequencing

The Qiagen Genra Puregene salting out method was used for isolation of high-molecular-weight genomic DNA from blood. Genomic DNA from saliva was isolated using prepIT.L2P (DNA Genotek Inc., Ottawa, Canada). DNA concentrations were determined by fluorescence base determination (pico green).

Whole-genome sequence data in the 67 families (302 affecteds and 49 unaffecteds) were generated at Baylor University, the

Broad Institute, and Washington University (i.e., 3 National Human Genome Research Institute Large Scale Sequencing and Analysis Centers). Using the Burrows-Wheeler Aligner (v0.6.2), sequence reads were aligned to the GRCh37 reference genome, and variant calling was performed using the Atlas V2 and Genome Analysis Tool Kit–HaplotypeCaller pipelines. Discrepancies between calls from both pipelines were reconciled creating a consensus data set. Variants not labeled “Pass” or with a low mapping score (QUAL score <22, Prob <0.95), low read depth (<6), an out of range variant read to total read depth (<3 reads or <10% alternate read for heterozygotes), or variants in regions represented by a single strand (>0.99) were removed, similar to monomorphic variants, variants with high missingness (>20%), variants with excessive heterozygosity ($|z| > 1.22$ for minor allele frequency [MAF] <0.2, $|z| > 5$ SD for $MAF \geq 0.2$), and variants with very high read depth (>500 reads). Annotation of variants passing quality control was based on ANNOVAR and included functional prediction by SIFT, PolyPhen2, and GERP. Allele frequencies were derived from the Exome Sequencing Project, 1000 Genomes, and the Exome Aggregation Consortium (ExAC).

Postmortem brain data from the ROS/MAP cohort

All postmortem brain data were generated as part of the Religious Orders Study¹² and Rush Memory and Aging Project¹³ (ROS/MAP): 2 ongoing, longitudinal cohort studies of aging coordinated out of the Rush Alzheimer’s Disease Center (RADC) in Chicago, IL. All subjects are older and recruited free of dementia (mean age at entry 78 ± 9 years), are administered annual cognitive and clinical assessments, and sign an Anatomical Gift Act.

RNA sequencing and gene expression quantification

Gene expression data were generated using RNA-seq from the dorsolateral prefrontal cortex (DLPFC) of 638 ROS/MAP subjects, according to previously published methods.¹⁴ Briefly, RNA was extracted using the Qiagen miRNeasy mini kit and RNase-Free DNase Set. Libraries were prepared by the Broad Institute’s Genomic Platform (strand-specific 2’-deoxyuridine 5’-triphosphate method and poly-A selection). Sequencing was performed using the Illumina HiSeq platform (50 million paired-end reads, 101bp each). Alignment was performed using Trinity to the GENCODE v14 transcriptome (GRCh37; gencodegenes.org/releases/). Quantification of gene counts was performed using RNA-Seq by Expectation Maximization, batch normalization with Combat, and mean-variance correction weights calculated using the edgeR and voom Bioconductor packages in R (v3.4.1).

Postmortem neuropathologic assessment

All ROS/MAP subjects were administered detailed neuropathologic evaluations at autopsy by a board-certified neuropathologist who was blinded to clinical data. Five Alzheimer disease–related pathologies were evaluated: neuritic and diffuse plaque counts, neurofibrillary tangle counts, immunohistochemical quantifications of total β -amyloid by image

analysis (continuous), and TDP43 proteinopathy (4 ordinal stages).¹⁵ Detailed descriptions of each pathologic measure are readily available at the RADC website (radc.rush.edu/).

Standard protocol approvals, registrations, and patient consents

The EFIGA study was approved by the Institutional Review Board of Columbia University. All participants or their guardians provided written informed consent. All study protocols of the ROS/MAP cohort were approved by the Institutional Review Board of Rush University Medical Center, and all study participants have provided informed, written consent and signed a repository consent that allows their data to be shared.

Data availability

Anonymized data for the WGS data set can be obtained by qualified investigators through the database of Genotypes and Phenotypes (phs000572.v1.p1) and through the National Institute on Aging Genetics of Alzheimer’s Disease Data Storage Site (niagads.org). Data generated from ROS/MAP can be obtained via the RADC Resource Sharing Hub (radc.rush.edu/).

Statistical methods

Single-marker association and gene-based tests for variant validation

Variants that were found to segregate with AD status in the sequenced families and were therefore included in the follow-up genotyping in all family members and the 444 unrelated controls were checked for Hardy-Weinberg equilibrium and tested for association with AD using a generalized linear mixed model accounting for population structure and relatedness. Models were first adjusted for kinship coefficient, age, and sex and subsequently in addition for APOE genotype. For a second, independent validation, we leveraged GWAS data available in an independent set CHs of individuals from the Washington Inwood Columbia Aging Project¹⁶ and conducted gene-based tests using Sequence Kernel Association Test optimal test adjusting for principal components for population substructure, age, sex, and subsequently in addition for APOE genotype.

To assess whether the variants associated with AD were enriched in these families, we used data from the Genome Aggregation Database (gnomAD) (gnomad.broadinstitute.org/), which includes 123,136 whole exomes and 15,496 whole genomes from unrelated individuals from various ethnic groups that were derived by a variety of large-scale disease-specific and population genetic sequencing projects. To determine enrichment of putative disease-associated synonymous variants in the family set, we compared their allele frequencies with the gnomAD Latino population using a Fisher exact test first including all family members and then selecting only 1 affected from each family.

Expression quantitative trait locus analyses

Expression QTL analyses were performed in DLPFC tissue from 494 ROS/MAP subjects for 13,484 expressed genes (average sequencing depth = 90 million reads). Genotype data

for this analysis were imputed using BEAGLE (v.3.3.2) and the 1000 Genomes Project Consortium interim phase 1 haplotype reference data set. Details of the methodology and results of eQTL analyses have been recently published,¹⁷ and full summary statistics have been made publicly available for query and download (mostafavilab.stat.ubc.ca/xqtl/).

Statistical correction for multiple testing of the marginal cis-eQTL results was reassessed for this study according to the estimated number of effective tests, which was calculated using the Genetic Type I Error Calculator¹⁸ and takes into account linkage disequilibrium in each region of interest. Using this method, *p* values were first corrected within each selected locus (including a buffer of 100 kb around each target gene) and then underwent Bonferroni correction for the number of independent loci tested (*n* = 4). For example, at CDH23, this yielded a significance threshold of *p* < 2.9 × 10⁻⁵ (0.05/437.55 effective tests/4 loci).

Association of gene expression with clinical diagnosis and neuropathologic endophenotypes of AD in the ROS/MAP cohort

To detect associations between the expression of target genes and clinical diagnosis at death or levels of postmortem neuropathology, gene counts from voom were analyzed in robust linear models using iterative reweighted least squares regression (implemented in the MASS R package), covarying for effects of sex, age at death (age at last visit for clinical AD diagnosis), postmortem interval, RNA integrity, APOEε4 status, first 3 genomic principal components calculated using EIGENSTRAT, and composition of cell types. Significance was based on a conservative Bonferroni threshold of *p* < 0.0025 (0.05/4 genes/5 tested pathologies).

Analysis of effect of identified variants on splicing

Minigene constructs

Reference and variant allele sequences were obtained by gene synthesis from GenScript (Piscataway Township, NJ). Subsequent cloning of the synthesized sequences (table 5) was performed following methods previously described.¹⁹ PUC19 plasmid vector and Gateway cloning vectors pENTR/D-TOPO (Invitrogen) and pDESTsplice minigene splicing vectors (Addgene; ref#32484) were used to obtain the final minigene splicing vectors. Final constructs for CDH23 were verified by restriction digestion with BstEII and HindIII, and final constructs for SLCA3R1 were verified by restriction digestion with SbfI and HindII followed by DNA sequencing (Genewiz, South Plainfield, NJ).

In vitro transfection

HEK 293 cells (2.5 × 10⁵) were transfected with 0.75 μg of plasmid DNA carrying either the reference allele or variant allele of interest using CalFectinTM Mammalian DNA Transfection Reagent (SigmaGen Laboratories, Ijamsville, MD). We performed 3 transfection replicates for each plasmid. After 24 hours, RNA isolation from the 3 replicates was performed using

the RNeasy[®] Plus mini kit (Qiagen). The First Strand complementary DNA (cDNA) Synthesis Kit (Origene, Rockville, MD) was used to generate cDNA from 500 ng of the RNA following the manufacturer's instructions.

Analysis of minigene expression

The cDNA containing the spliced products of the minigene vector was amplified using the Phusion High-Fidelity DNA Polymerase (New England Biolabs) using previously reported PCR primers specific for the rat insulin 2 exons (5'-CCTGCTCATCCTCTGGGAGC-3' and 5'-AGGTCTGAAGGTCACGGGCC-3').¹⁹ PCR conditions were set at 98°C denaturation for 30 seconds, followed by 35 cycles of 98°C for 7 seconds, 62°C for 30 seconds, and 72°C for 1 minute, with a final 72°C hold for 5 minutes. Products were visualized by gel electrophoresis, and subsequent DNA gel extraction/sequencing was performed to confirm that the spliced sequences were flanked by rat insulin 2 exon sequences and were mapped to the CDH23/SC9A3R1 genomic sequence using BLAT. Following previously described methods,¹⁹ quantitation of the cDNA products was performed in parallel using the same forward PCR primer, with an additional tag at the 5' end with 6-FAMTM to enable measurement of differences in the size and quantity of the reference and variant minigene products. Quantitation samples were run on the Applied Biosystems (ABI) ABI 3730xl DNA Analyzer (Genewiz, South Plainfield, NJ). The ABI internal lane size standard GeneScan 500 LIZ was run with all the samples, which allows to accurately sizing 35–500 bp fragments. Quantitation data were analyzed using GeneMapperTM and Peak ScannerTM (ThermoFisher Scientific) software. In addition, PCR products of the cDNA were sequenced (Genewiz, DNA Sanger sequencing) to determine the precise boundaries generated during splicing. The sequences were verified by confirming that the spliced sequences were flanked by rat insulin 2 exon sequences following previously validated methods¹⁹ and were mapped to the CDH23 and SLCA3R1 genomic sequences using BLAT.

Results

Analysis of sequencing data

Characteristics of the discovery data sets are shown in table 1. Variant calling, recalibration, and application of quality control filters resulted in 97,866 synonymous variants, out of which 46,897 were rare. Eleven variants (located in *RHBDD2*, *ITIH2*, *ANK3*, *CDH23*, *OTUD7A*, *EDC4*, *FUK*, *SLC9A3R1*, *MBP*, *CC2D1A*, and *TGM6*) segregated with late-onset Alzheimer disease (LOAD) status in the whole exome sequencing or WGS data set and were genotyped for validation in all family members of the discovery families, 48 additional multiplex CH families (total 115 families with 1,024 subjects genotyped), and 444 independent controls of similar age and ancestry. Genotyping for all variants was successful in all these subjects. In linear mixed models adjusted for kinship, age, sex, and APOE genotype, the variants located in *SLC9A3R1* (rs41282067),

Table 1 Characteristics of the discovery sample set

Families, n	67
Total number of subjects sequenced, n	351
Number of subjects sequenced per pedigree, mean (SD)	5.24 ± 1.67
Affected subjects sequenced, n	302
Unaffected subjects sequenced, n	49
Age (sequenced individuals)	73.02 ± 10.0
Women (%)	58.72
APOEε4 frequency	17.03%

ITIH2 (rs143731868), and *RHBDD2* (rs190871206) remained significantly associated with LOAD ($p \leq 0.005$ adjusted for multiple testing, table 2), and the variant in *CDH23* (rs56013867) was close to significance. Twenty-two of the 539 affected individuals in the 115 genotyped families carried the *SLC9A3R1* variant, 16 carried the *RHBDD2* variant, 10 the *ITIH2* variant, and 20 the *CDH23* variant. For all variants, these frequencies were significantly higher than the expected frequencies based on the gnomAD database (gnomad.broadinstitute.org/; Fisher exact test p value <0.0001). When restricting the analyses to 1 affected person per family, and using the overall number of alleles genotyped or sequenced for each variant as denominator in the Fisher exact test, the results remained significant for all 4 variants (*RHBDD2* ($p < 0.0001$), *ITIH2* ($p < 0.0001$), *CDH23* (0.0004), and *SLC9A3R1* ($p < 0.0001$)). In gene-based tests including synonymous and functional variants or synonymous variants only in these genes

in an independent cohort of 1,511 subjects of the same ancestry (779 affecteds and 732 unaffecteds), *CDH23*, *SLC9A3R1*, and *ITIH2* were replicated (*ITIH2*: $p = 0.03$, *CDH23*: $p = 0.04$, and *SLC9A3R1*: $p = 0.03$). All 4 genes have a positive constraint metric for synonymous variants based on ExAC data suggesting intolerance to variation (exac.broadinstitute.org/) and are expressed in the brain.

Association between gene expression, clinical diagnosis, and pathologic AD phenotypes in the ROS/MAP cohort

To further validate the association of these 4 genes with AD, we explored the association of their expression with clinical Alzheimer dementia diagnosis proximate to death and AD brain pathology in the ROS/MAP cohort ($n = 400$). Compared with individuals without cognitive impairment, expression of *CDH23* ($\beta = 0.53$, $p = 0.006$) and *SLC9A3R1* ($\beta = 0.50$, $p = 0.02$) was increased, and expression of *RHBDD2* ($\beta = -0.70$, $p = 0.02$) decreased in persons with a clinical diagnosis of AD (table 3). In line with this finding, higher expression of *CDH23* (beta = 0.26 ± 0.08 , $p = 4.9E-4$) and lower expression of *RHBDD2* (beta = -0.60 ± 0.12 , $p = 5.5E-7$) were related to brain β -amyloid load (Bonferroni-corrected p value for multiple testing = 0.0025; table 4). *SLC9A3R1* expression was associated with burden of TDP43 pathology (beta = 0.58 ± 0.17 , $p = 5.9E-4$). Using the eQTL data available in the ROS/MAP data set, the identified *CDH23* variant—while not present itself on the microarray chip—was shown to be in linkage disequilibrium with a set of single nucleotide polymorphisms modulating *CDH23* expression levels (top single nucleotide polymorphism: rs11000035 at 73625754bp, $p = 4.85E-6$, $D' = 1.0$).

Table 2 General linear model relating candidate variants with AD diagnosis in 115 families and 444 independent controls

Gene	Chr	Marker	Bp (hg37)	Minor allele	Reference allele	MAF	Model 1			Model 2		
							OR	SE	p	OR	SE	p
<i>RHBDD2</i>	7	rs190871206	75,508,445	T	C	0.006	1.37	0.1	0.002	1.37	0.1	0.003
<i>ITIH2</i>	10	rs143731868	7,791,240	T	C	0.004	1.49	0.13	0.002	1.5	0.13	0.001
<i>ANK3</i>	10	rs55680239	61,868,656	C	G	0.008	1.02	0.09	0.75	1.03	0.09	0.72
<i>CDH23</i>	10	rs56013867	73,558,133	C	G	0.012	1.22	0.08	0.01	1.22	0.08	0.01
<i>OTUD7A</i>	15	rs140728111	31,779,430	A	G	0.005	1.19	0.12	0.15	1.19	0.12	0.14
<i>EDC4</i>	16	rs117562635	67,913,509	T	C	0.008	1.02	0.09	0.77	1.02	0.09	0.78
<i>FUK</i>	16	rs747173917	70,509,389	C	T	0.008	1.07	0.1	0.46	1.07	0.1	0.47
<i>SLC9A3R1</i>	17	rs41282067	72,759,529	A	G	0.014	1.25	0.07	0.002	1.25	0.07	0.002
<i>MBP</i>	18	rs564110678	74,729,037	A	G	0.007	1.09	0.1	0.37	1.09	0.1	0.36
<i>CC2D1A</i>	19	rs368632032	14,037,873	T	C	0.006	1.2	0.11	0.08	1.19	0.11	0.11
<i>TGM6</i>	20	rs765423180	2,384,303	T	C	0.002	1.35	0.18	0.1	1.37	0.18	0.08

Abbreviations: AD = Alzheimer disease; MAF = minor allele frequency; OR = odds ratio; SE = standard error.

Model 1 is adjusted for kinship coefficient, age, and sex, and model 2 is adjusted for kinship, age, sex, and APOE genotype.

Bold values indicate statistical significance.

Table 3 Association between gene expression and clinical diagnosis of AD in the ROS/MAP data set

Gene	p Value	Beta	SE
<i>RHBDD2</i>	0.02	-0.70	0.30
<i>ITIH2</i>	0.76	0.02	0.08
<i>CDH23</i>	0.006	0.53	0.19
<i>SLC9A3R1</i>	0.02	0.50	0.22

Abbreviations: AD = Alzheimer disease; beta = beta coefficient; SE = standard error.
Adjusted for sex, age at last visit, postmortem interval, RNA integrity, APOEε4, and PCAs for population stratification.
Bold value indicates statistical significance.

Analysis of effect of identified variants on splicing

For accurate intron excision, splicing requires sequence components from both the intron and the exon. Synonymous variants can influence splicing accuracy or efficiency. To determine whether identified synonymous variants affect AD risk through an effect on splicing accuracy or efficiency, we performed minigene splicing assays. Although functional analysis by minigene assay resulted in unchanged splicing patterns between reference and variant alleles for rs56013867 (*CDH23*) and rs41282067 (*SLC9A3R1*), there was evidence of altered splicing efficiency for these variants (table 5; figure, A). Although the rs56013867 variant in *CDH23* exon 50 showed a significant decrease, the rs41282067 variant in *SLC9A3R1* exon 3 showed a significant increase in splicing efficiency compared with their reference sequences (figure, A). Fluorescent quantitation also showed lower splicing efficiency for the rs56013867 variant (*CDH23*) compared with the reference allele (relative fluorescence units 4,295 and 13,374, respectively) (figure, B) and a significantly higher splicing efficiency rate for the rs41282067 variant (*SLC9A3R1*) compared with the reference allele (relative fluorescence units 25,825 and 21,778, respectively) (figure, C).

Discussion

Capitalizing on this data set of heavily loaded CH families, we identified rare synonymous variants in 4 genes not previously implicated in AD by genome-wide association or sequencing

studies (*CDH23*, *SLC9A3R1*, *RHBDD2*, and *ITIH2*) and demonstrated by functional splicing assays²⁰ an effect of 2 of these variants on splicing efficiency.

All 4 genes identified are acting in pathways established in AD. There is strong evidence from previous epigenomic and functional studies for a role of *CDH23* in AD. *CDH23* encodes a member of the cadherin superfamily a group of calcium-dependent cell adhesion glycoproteins essential for cell-cell adhesion, morphogenesis, neuronal connectivity, and tissue integrity, highly expressed in the brain. In addition to a recent large-scale gene-gene interaction study on AD, which identified *CDH23* as a risk gene,²¹ 2 independent epigenome-wide association studies of AD identified overlapping methylation signals in *CDH23*^{22,23} and suggested that expression of *CDH23* is increased in astrocytes near neuritic plaques,²⁴ in line with our finding of an association of *CDH23* with brain amyloid pathology. Astrocytes have a vital function in the brain circuitry and are critical for a variety of functions such as formation and maturation of synapses, modulation of synaptic activity and plasticity, neurotransmitter clearance, and homeostasis.²⁵ Reactive astrocytes are induced by injury and disease and activated by neuroinflammatory microglia; both activated microglia and astrocytes are prominent features in AD.²⁵

Further consistent with our findings of an association of *CDH23* with amyloid pathology, epithelial cadherin (encoded by *CDH1*) binds to presenilin-1,²⁶ and neural (N-) cadherin has been implicated in amyloid-β release via interaction with presenilin-1. In addition, N-cadherin plays an essential role in synapse function, and inhibition of N-cadherin function accelerates amyloid-β-induced synapse impairment.²⁷

The *SLC9A3R1* gene (solute carrier family 9 [sodium/hydrogen exchanger], isoform 3 regulatory factor 1) encodes a scaffold protein (Na⁺/H⁺ exchanger regulatory factor 1 [NHERF1]) that connects plasma membrane proteins to the actin cytoskeleton, thereby regulating their surface expression. Notably, NHERF1 modulates the cell surface expression of the glutamate transporter GLAST.²⁸ A series of key pathologic changes of AD including deposition of Aβ in plaques, soluble Aβ oligomers, hyperphosphorylated tau protein, oxidative stress, and neuronal inflammation have been linked to increased activation of the glutamatergic system.^{29–32} Notably,

Table 4 Association between gene expression and AD pathologic outcomes in the ROS/MAP data set

Gene	Pathology	Beta	SE	p Value	n
<i>CDH23</i>	Amyloid load	0.264	0.075	4.90 × 10 ⁻⁴	594
<i>CDH23</i>	Neuronal loss	0.457	0.136	7.67 × 10 ⁻⁴	598
<i>RHBDD2</i>	Amyloid load	-0.600	0.119	5.47 × 10 ⁻⁴	594
<i>SLC9A3R1</i>	TDP43	0.576	0.168	5.93 × 10 ⁻⁴	499

Abbreviations: AD = Alzheimer disease; beta = beta coefficient; SE = standard error; n = sample size.
Adjusted for sex, age at death, postmortem interval, RNA integrity, APOEε4, and principal components for population stratification.

Table 5 Summary information for the variants tested by the minigene splicing assay

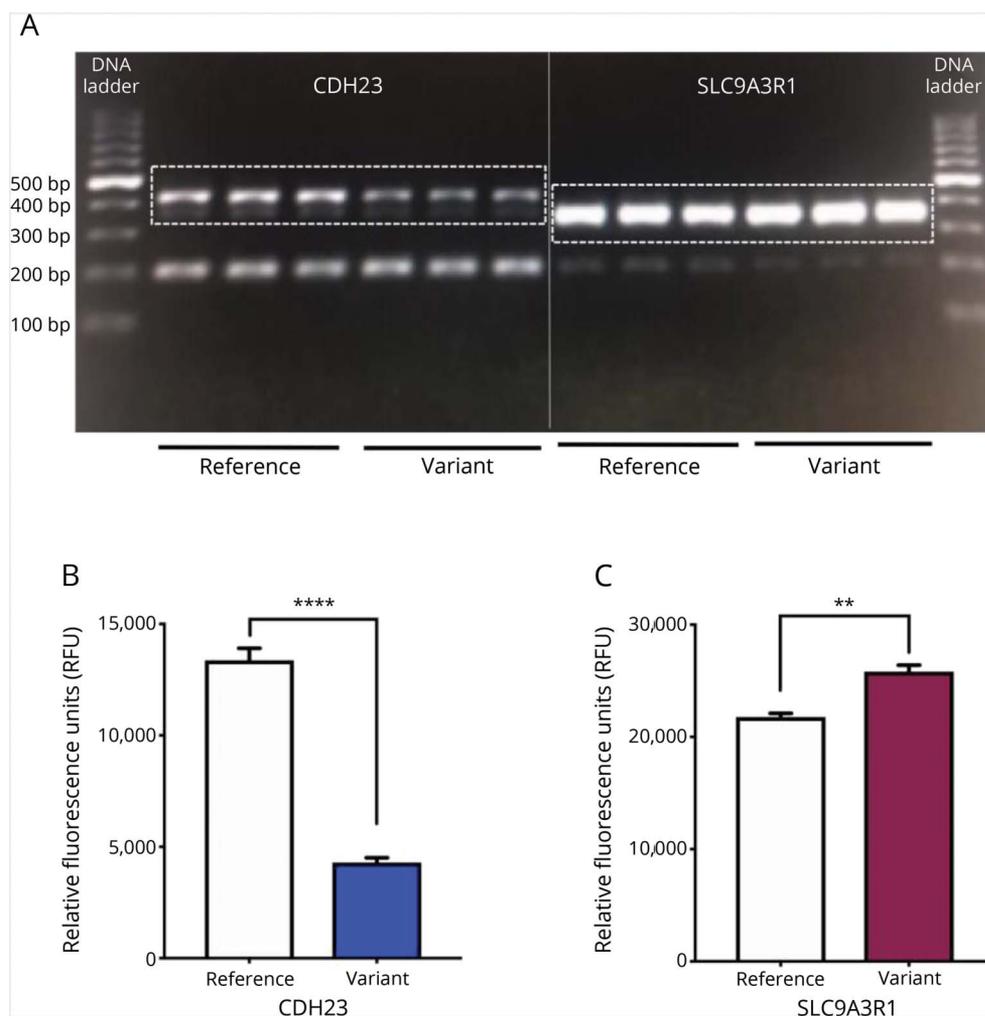
Transcript	rsSNP ID	Molecular consequence	Type of variant	Chromosome position	Exon no.	Exon size (bps)	Position within exon
CDH23	rs56013867	c.6852G>C	Synonymous	Chr10:73558133 (GRCh37.p13)	50	226	24
SLC9A3R1	rs41282067	c.627G>A	Synonymous	Chr17:72759529 (GRCh37.p13)	3	153	24

NHERF1 encoded by SLC9A3R1 also binds to β -catenin and stabilizes the interaction with E-cadherin at cell-cell junctions,³³ indicating that our observations on the SLC9A3R1 and CDH23 genes might converge on a common mechanism.

The rhomboid domain containing 2 (RHBDD2) gene, also named RHBDL7, encodes a member of the highly conserved rhomboid family of membrane-bound proteases. Rhomboids bind membrane proteins and direct them with high precision into various different cellular pathways.³⁴ As a result of this vital

function, members of the rhomboid family have been implicated in a variety of diseases including AD and Parkinson disease. Rhomboid protease *RHBDL4* (alias: *RHBDD1*), cleaves amyloid precursor protein inside the cell, causing it to bypass amyloidogenic processing, leading to reduced A β levels.³⁵ In line with this notion, in our analyses, higher expression of RHBDD2 was associated with lower brain amyloid load. In addition, RHBDD2—similar to its homologs RHBDD1 and RHBDD3—has also been associated with stress response and apoptosis.^{36–38}

Figure Experimental analysis of splicing interference by AD-associated synonymous variants



(A) Gel electrophoresis with splicing products collected following transient transfection into HEK 293 cells. Constructs with the reference and variant alleles were analyzed separately. White dotted rectangles indicate the bands of spliced products for CDH23 and SLC9A3R1; numbers correspond to band sizes. Results of quantitation with fluorescent labeled PCR primers (conducted in a separate experiment) are shown for variants rs56013867 (CDH23) (B) and rs41282067 (SLC9A3R1) (C), revealing different splicing efficiencies between reference and variant alleles. ** $p \leq 0.01$; **** $p \leq 0.0001$ by 2-tailed Student *t* test.

Although *ITIH2* could not be validated by gene expression analyses in this study, it should not be readily discarded as a potential candidate gene. Inter-alpha inhibitor proteins (IAIPs) including *ITIH2* are anti-inflammatory molecules, and recent studies in rodents have demonstrated that IAIPs play a critical role in regulating inflammatory response and cell survival in the CNS after brain injury.^{39,40} Over the past 10 years, large-scale genomic studies have firmly established immune response as an etiologic pathway in AD. Administration of human plasma-derived IAIPs following neonatal HI brain injury in rodents decreases neuronal cell death and improves cognitive and behavioral function.⁴¹

There is limited overlap of our results with the findings of the previous large GWAS. In part, this is likely due to our focus on rare variants ($MAF < 0.05$), which have not been well covered by most GWASs, our focus on CHs, which have been significantly underrepresented in previous genomic studies, and our focus on synonymous variants, which have been widely neglected by the first-pass analyses conducted in previous large-scale sequencing studies in AD.

A significant strength of this study lies in the ascertainment scheme. Prioritization of multigenerational families highly loaded for AD allows to detect disease-associated variants by segregation with disease status. Prioritization of families with minimal clustering for APOE ϵ 4 further enriches the analytic data set for disease-associated variants other than APOE ϵ 4, yielding an additional increase in power to detect novel variants. The focus on a minority population allows to detect novel loci not identified in studies of non-Hispanic whites. Related to the latter notion, a limitation of the study is the lack of an additional independent data set of CH families available for replication. However, validation of the identified loci in the larger set of families and independent controls, demonstration of a significantly higher frequency of the identified variants in our data set compared with reference populations, validation of identified variants in brain expression and eQTL data with both clinical and neuropathologic key outcomes of AD, and demonstration of an effect of the *CDH23* and *SLC9A3R1* mutant variants on splicing efficiency significantly reduce the likelihood of false-positive findings.

In summary, this study identified 4 rare synonymous variants in *CDH23*, *SLC9A3R1*, *RHBDD2*, and *ITIH2* that modulate risk of AD in CH families and are associated with LOAD pathology. Although replication efforts in subjects of other ancestral background are critical to generalize these findings, and studies are needed to characterize in detail the impact of the identified variants in different cell types in the brain in vitro and in vivo, these genes cluster with amyloid precursor protein processing, synaptic function and plasticity, glutamatergic neurotransmitter signaling, and immune system/inflammation in pathways involved in AD etiology. In addition, this study strongly suggests that synonymous variants contribute to AD risk and that comprehensive scrutinization of this type of variant in AD research is warranted and critical. A comprehensive functional follow-up of the identified variants is needed to fully

understand the exact mechanisms through which they exert their effect on AD.

Study funding

Data collection for the EFIGA Study was supported by the National Institute on Aging (NIA) and by the NIH (RF1AG015473). Data collection and sharing for Whicap was supported by the Washington Heights-Inwood Columbia Aging Project (WHICAP, RF1AG054023) funded by the National Institute on Aging (NIA). Dr. Reitz was further supported by NIH grants RF1AG054080, R01AG064614, U01AG052410, and P50AG008702. Dr. Santa-Maria was supported by NIH grants P50AG008702 and R01NS095922. The Religious Orders Study and Rush Memory and Aging Project are supported by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG36836, and U01AG61366. The whole-genome sequencing was conducted as part of the Alzheimer's Disease Sequencing Project (ADSP). The ADSP is comprised of 2 Alzheimer disease (AD) genetics consortia and 3 National Human Genome Research Institute (NHGRI)-funded Large Scale Sequencing and Analysis Centers (LSAC). The 2 AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other NIH institutes and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Lindsay Farrer, Jonathan Haines, Richard Mayeux, Margaret Pericak-Vance, and Gerard Schellenberg); U01AG049505 to Dr. Sudha Seshadri; U01AG049506 to Dr. Eric Boerwinkle; U01AG049507 to Dr. Ellen Wijsman; and U01AG049508 to Dr. Alison Goate, and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr. Goate, U01AG052410 to Dr. Pericak-Vance, and U01 AG052409 to Drs. Seshadri and Fornage. Data generation and harmonization in the Follow-up Phases is supported by U54AG052427 to Drs. Schellenberg and Li-San Wang.

Disclosure

None of the authors have a conflict of interest. Go to Neurology.org/NG for full disclosures.

Publication history

Received by *Neurology: Genetics* November 22, 2019. Accepted in final form May 5, 2020.

Appendix Authors

Name	Location	Contribution
Min Tang, PhD	University of Arizona	Designed and conceptualized the study; analyzed the data; and drafted the manuscript for intellectual content

Appendix (continued)

Name	Location	Contribution
Maria Eugenia Alaniz, PhD	Columbia University, NY	Major role in the acquisition of data
Daniel Felsky, PhD	The Krembil Centre for Neuroinformatics, Toronto, Ontario, Canada	Data analysis; interpreted the data; and revised the manuscript for intellectual content
Badri Vardarajan, PhD	Columbia University, NY	Data analysis; interpreted the data; and revised the manuscript for intellectual content
Dolly Reyes-Dumeyer, BA	Columbia University, NY	Major role in the acquisition of data
Rafael Lantigua, MD	Columbia University, NY	Major role in the acquisition of data
Martin Medrano, MD	Mother and Teacher Pontifical Catholic University, Santiago, Dominican Republic	Major role in the acquisition of data
David A. Bennett, MD	Rush University, IL	Major role in the acquisition of data; interpreted the data; and revised the manuscript for intellectual content
Philip L. de Jager, MD	Columbia University, NY	Major role in the acquisition of data; interpreted the data; and revised the manuscript for intellectual content
Richard Mayeux, MD	Columbia University, NY	Major role in the acquisition of data; interpreted the data; and revised the manuscript for intellectual content
Ismael Santa-Maria, PhD	Columbia University, NY	Acquisition of data; interpreted the data; and revised the manuscript for intellectual content
Christiane Reitz, MD, PhD	Columbia University, NY	Designed and conceptualized the study; interpreted the data; and revised the manuscript for intellectual content

References

- Kowalski TJ, Pawelczyk M, Rajkowska E, Dudarewicz A, Sliwiska-Kowalska M. Genetic variants of CDH23 associated with noise-induced hearing loss. *Otol Neurotol* 2014;35:358–365.
- Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* 2019;51:414–430.
- Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012;488:96–99.
- Benitez BA, Cooper B, Pastor P, et al. TREM2 is associated with the risk of Alzheimer's disease in Spanish population. *Neurobiol Aging* 2013;34:1711.e15–1717.
- Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013;45:1452–1458.
- Cruchaga C, Haller G, Chakraverty S, et al. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One* 2012;7:e31039.
- Nackley AG, Shabalina SA, Tchivileva IE, et al. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 2006;314:1930–1933.
- Vardarajan BN, Faber KM, Bird TD, et al. Age-specific incidence rates for dementia and Alzheimer disease in NIA-LOAD/NCRAD and EFIGA families: National Institute on Aging Genetics Initiative for Late-Onset Alzheimer Disease/National Cell Repository for Alzheimer Disease (NIA-LOAD/NCRAD) and Estudio Familiar de Influencia Genética en Alzheimer (EFIGA). *JAMA Neurol* 2014;71:315–323.
- Romas SN, Santana V, Williamson J, et al. Familial Alzheimer disease among Caribbean Hispanics: a reexamination of its association with APOE. *Arch Neurol* 2002;59:87–91.
- Stern Y, Andrews H, Pittman J, et al. Diagnosis of dementia in a heterogeneous population. Development of a neuropsychological paradigm-based diagnosis of dementia and quantified correction for the effects of education. *Arch Neurol* 1992;49:453–460.
- Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J Lipid Res* 1990;31:545–548.
- Bennett DA, Schneider JA, Arvanitakis Z, Wilson RS. Overview and findings from the religious orders study. *Curr Alzheimer Res* 2012;9:628–645.
- Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and findings from the rush memory and aging Project. *Curr Alzheimer Res* 2012;9:646–663.
- Yu L, Chibnik LB, Srivastava GP, et al. Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol* 2015;72:15–24.
- Buchman AS, Leurgans SE, Nag S, Bennett DA, Schneider JA. Cerebrovascular disease pathology and parkinsonian signs in old age. *Stroke* 2011;42:3183–3189.
- Tang MX, Stern Y, Marder K, et al. The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *JAMA* 1998;279:751–755.
- Ng B, White CC, Klein HU, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci* 2017;20:1418–1426.
- Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet* 2012;131:747–756.
- Scott A, Petrykowska HM, Hefferon T, Gotea V, Elnitski L. Functional analysis of synonymous substitutions predicted to affect splicing of the CFTR gene. *J Cyst Fibros* 2012;11:S11–S17.
- Gaidrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol* 2010;653:249–257.
- Hohman TJ, Bush WS, Jiang L, et al. Discovery of gene-gene interactions across multiple independent data sets of late onset Alzheimer disease from the Alzheimer Disease Genetics Consortium. *Neurobiol Aging* 2016;38:141–150.
- Lord J, Cruchaga C. The epigenetic landscape of Alzheimer's disease. *Nat Neurosci* 2014;17:1138–1140.
- Lunnon K, Smith R, Hannon E, et al. Methylomic profiling implicates cortical de-regulation of ANK1 in Alzheimer's disease. *Nat Neurosci* 2014;17:1164–1170.
- De Jager PL, Srivastava G, Lunnon K, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci* 2014;17:1156–1163.
- Vasile F, Dossi E, Rouach N. Human astrocytes: structure and functions in the healthy brain. *Brain Struct Funct* 2017;222:2017–2029.
- Baki L, Marambaud P, Efthimiopoulos S, et al. Presenilin-1 binds cytoplasmic epithelial cadherin, inhibits cadherin/p120 association, and regulates stability and function of the cadherin/catenin adhesion complex. *Proc Natl Acad Sci U S A* 2001;98:2381–2386.
- Andreyeva A, Nieweg K, Horstmann K, et al. C-terminal fragment of N-cadherin accelerates synapse destabilization by amyloid-beta. *Brain* 2012;135:2140–2154.
- Sato K, Otsu W, Otsuka Y, Inaba M. Modulatory roles of NHERF1 and NHERF2 in cell surface expression of the glutamate transporter GLAST. *Biochem Biophys Res Commun* 2013;430:839–845.
- Gray CW, Patel AJ. Neurodegeneration mediated by glutamate and beta-amyloid peptide: a comparison and possible interaction. *Brain Res* 1995;691:169–179.
- Mattson MP, Pedersen WA, Duan W, Culmsee C, Camandola S. Cellular and molecular mechanisms underlying perturbed energy metabolism and neuronal degeneration in Alzheimer's and Parkinson's diseases. *Ann N Y Acad Sci* 1999;893:154–175.
- De Felice FG, Wu D, Lambert MP, et al. Alzheimer's disease-type neuronal tau hyperphosphorylation induced by A beta oligomers. *Neurobiol Aging* 2008;29:1334–1347.
- Gasparini L, Dityatev A. Beta-amyloid and glutamate receptors. *Exp Neurol* 2008;212:1–4.
- Kreimann EL, Morales FC, de Orbeta-Cruz J, et al. Cortical stabilization of beta-catenin contributes to NHERF1/EBP50 tumor suppressor function. *Oncogene* 2007;26:5290–5299.
- Adrain C, Zettl M, Christova Y, Taylor N, Freeman M. Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. *Science* 2012;335:225–228.
- Paschkowski S, Hamze M, Oestereich F, Munter LM. Alternative processing of the amyloid precursor protein family by rhomboid protease RHBDD1. *J Biol Chem* 2016;291:21903–21912.
- Wang Y, Guan X, Fok KL, et al. A novel member of the rhomboid family, RHBDD1, regulates BIK-mediated apoptosis. *Cell Mol Life Sci* 2008;65:3822–3829.
- Ren X, Song W, Liu W, et al. Rhomboid domain containing 1 inhibits cell apoptosis by upregulating AP-1 activity and its downstream target Bcl-3. *FEBS Lett* 2013;587:1793–1798.
- Lacunza E, Rabassa ME, Canzoneri R, et al. Identification of signaling pathways modulated by RHBDD2 in breast cancer cells: a link to the unfolded protein response. *Cell Stress Chaperones* 2014;19:379–388.
- Garantzios S, Hollingsworth JW, Ghanayem RB, et al. Inter-alpha-trypsin inhibitor attenuates complement activation and complement-induced lung injury. *J Immunol* 2007;179:4187–4192.
- Threlkeld SW, Gaudet CM, La Rue ME, et al. Effects of inter-alpha inhibitor proteins on neonatal brain injury: age, task and treatment dependent neurobehavioral outcomes. *Exp Neurol* 2014;261:424–433.
- Chen X, Rivard L, Naqvi S, et al. Expression and localization of inter-alpha inhibitors in rodent brain. *Neuroscience* 2016;324:69–81.