Co-occurring Pathogenic Variants on 6q27 associate with Dementia spectrum disorders in a Peruvian Family

Karla Lucia F. Alvarez[1], Jorge A. Aguilar-Pineda[1], Michelle M. Ortiz-Manrique[1], Marluve F. Paredes-Calderon[2], Bryan C. Cardenas-Quispe[2], Karin J. Vera-Lopez[1], Luis D. Goyzueta-Mamani[1], Miguel A. Chavez-Fumagalli[1], Gonzalo Davila Del-Carpio[3], Antero Peralta-Mestas[2], Patricia L. Musolino[4,5], and Christian L. Lino Cardenas[6*]

[1]Laboratory of Genomics and Neurovascular Diseases, Universidad Católica de Santa María, Arequipa, Peru.

[2]Division of Neurology, Psychiatry and Radiology of the National Hospital ESSALUD-HNCASE, Arequipa, Peru.

[3]Vicerrectorado de Investigación, Universidad Católica de Santa María, Arequipa, Peru.

[4]Department of Neurology, Massachusetts General Hospital

[5]Center for Genomic Medicine, Massachusetts General Hospital

[6]Cardiovascular Research Center, Cardiology Division, Massachusetts General Hospital, Boston, MA, USA.

*Correspondence to: Christian L. Lino Cardenas

Massachusetts General Hospital & Harvard Medical School, The Richard B. Simches Research Center, 185 Cambridge Street, Room 3224, Boston, MA 02114
E-mail: clinocardenas@mgh.harvard.edu

Keywords: Unbiased gene discovery, whole genome sequencing, family-specific genetic factor, Amerindian ancestral background, Alzheimer's disease

Abstract

Evidence suggests that there may be racial differences in risk factors associated with the development of Alzheimer's disease and related dementias (ADRD). Using whole genome sequencing analysis, we identified the combination of 3 pathogenic variants at the heterozygous state (*UNC93A*: rs7739897 and *WDR27*: rs61740334; rs3800544) in a Peruvian family with a strong clinical history of ADRD. Remarkably, the combination of these variants was present in two generations of affected individuals but absent in healthy members within the family. *In silico* and *in vitro* studies have provided insights into the pathogenicity of these variants. These studies predict the loss of function of the mutant UNC93A and WDR27 proteins which induced dramatic changes on the global transcriptomic signature of brain cells including neurons, astrocytes and especially pericytes and vascular smooth muscle cells and thus indicating that the combination of these three variants may affect the neurovascular unit. In addition, key known molecular pathways associated with ADRD were enriched in brain cells with low levels of UNC93A and WDR27. Together our findings have identified a genetic risk factor for familial ADRD in a Peruvian family with Amerindian ancestral background.

Background

Neurological disorders are an important cause of disability and death around the world. Interestingly, Alzheimer's disease, dementia, Parkinson, epilepsy, schizophrenia, and autism spectrum disorder can share common anatomical alterations and cognitive defects [1]. Certainly, Alzheimer's disease is the main cause of dementia and contributes with 50 to 75% of cases [2] and can be presented in two forms as defined by age: early-onset Alzheimer's disease (EOAD) that occurs prior to 65 years of age [3] and a late-onset Alzheimer's disease (LOAD) that is mostly present later than 65 years of age [4]. Both, EOAD an LOAD can have a family origin and an inheritance mode of autosomal dominant transmission [5]. In this regard, genetic variations in 3 genes including the amyloid precursor protein (APP), presenilin 1 and 2 (PSEN1 and PSEN2) have nearly 100% penetrance and were identified as causative of EOAD [6–8]. On the other hand, expression of the apolipoprotein E (APOE) ε4 gene is the major risk factor for LOAD in Caucasian population [9]. However, there has been evidences that the risk to develop Alzheimer's disease in ε4 carriers differs among ethnic groups. For instance, ε4 carriers of African descent shows a low risk for Alzheimer's disease [10]. While, Amerindian genetic ancestry seemed to be protected from cognitive decline [11]. Similarly, variants in the *TREM2* (R47H, H157Y, and L211P) gene that are highly associated with Alzheimer's disease in Caucasian population [12] were not replicated in Japanese descendants [13]. These epidemiological observations indicate that certain genetic risk factors for neurological disorders have different impact from one ethnicity to another.

In recent years, genome wide association studies (GWAS) have permitted the identification and characterization of multiple genetic rick loci associated with ADRD (Alzheimer's Disease and Alzheimer's Disease Related Dementias) [14]. However, most of these genetic loci were not associated with the APP processing, but rather with the immune response (*TREM2, CLU, CR1, CD33, EPHA1, MS4A4A/MS4A6E*), endosomal trafficking (*PICALM, BIN1, CD2AP*) or lipid metabolism (*ABCA7*) [15]. Furthermore, it has been reported an overlapping between Alzheimer's disease pathogenic variants with other neurogenerative or neuropsychiatric disorders indicating a shared

genetics and molecular origin. For example, an Alzheimer's disease variant in the *TREM2* gene (rs75932628) was also correlated with amyotrophic lateral sclerosis [16], while a variant in the *MARK2* gene (rs10792421) was associated with Alzheimer's disease and bipolar disorder [17].

Identification of individuals at high risk for ADRD remains a global health need and a major challenge for minority populations. Here, we performed a whole genome sequencing analysis in a Peruvian family with strong clinical history of ADRD including Alzheimer's disease, schizophrenia, and cognitive deficit. Furthermore, we explored effect of these variants on the neurovascular unit of the brain through *in silico* and *in vitro* studies.

Methods

Patient sample collection

A family (n=14) originally from Peru with five members diagnosed with neurological and neuropsychiatry disorders were enrolled in this study. For the variant validation study, non-familial AD patients (n=8) and healthy individual (n=50) were recruited. Regarding the healthy individual the selection criteria was as following: age >60 years, without signs of dementia, and who have no familial history of AD. Probable AD was diagnosed according to the guidelines of the National Institute of Neurological and Communicative Disorders and the Stroke and Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) [18]. Whenever possible, cognitive status of each family member was diagnosed based on a neuropsychological test (MoCa blind test and the clock drawing test).

Genetic Analysis

Genomic DNA was extracted from saliva samples using the prepIT.L2P reagent (Genotek, Cat. No PT-L2P-5) according to the manufacturer's instructions. The qualified genomic DNA samples were randomly fragmented by Covaris Technology, obtaining a fragment of 350bp. The DNA Nanoballs (DNBs) were produced through rolling circle amplification (RCA) and the qualified DNBs were loaded into the patterned nanoarrays. The whole-genome sequencing (WGS)

was conducted on BGISEQ-500 platform (BGI Genomics, Shenzhen, China). Raw sequencing reads were aligned to the human reference genome (GRCh38/HG38) with The Burrows-Wheeler Aligner (BWA) software and variant calling were performed with the Genome Analysis Toolkit (GATK v3.5) according to best practices. On average, 88.10% mapped successfully and 93.23% mapped uniquely. The duplicate reads were removed from total mapped reads, resulting in about 2.48% duplicate rate and 30.72-fold mean sequencing depth on the whole genome excluding gap regions. On average per sequencing individual, 99.34% of the whole genome excluding gap regions were covered by at least 1X coverage, 98.78% had at least 4X coverage and 97.42% had at least 10X coverage. In addition, the distributions of per-base sequencing depth and cumulative sequencing depth.

Sanger sequencing

Based on the results of the WGS, the variants that were present on affected members of the family but absent in healthy individuals were validated by Sanger sequencing. All PCR products were sequenced using an ABI 3130 Genetic Analyzer. Sequence analysis was performed with the Chromas program in DNASTAR analysis package.

Gene	SNP ID	PCR primers		Sequencig primers
		Forward primer sequence (5´-3´)	Reverse primer sequence (5´-3´)	Forward primer sequence (5´-3´)
WDR27	rs3800544	GTTTGCGCTCCTAGTTTCATG	GCATTCCGTACTTCCTTCCATC	TGTCCTACCGACCTCTCCACTG
WDR27	rs61740334	ACTGTGAATGTCTCCCGATCAC	ACTTGAAGTTGCATGGCATGG	TTCCCTCAGGGAGGCATAC
UNC93A	rs7739897	TACGGCGTTCTGTTTGAGAAG	TACGGCGTTCTGTTTGAGAAG	GCTGCCTTGTCGCCAATTAC

Cell lines

Primary human vascular smooth muscle cells (VSMCs) from carotid of healthy donors were purchased from Cell Ap-plications Inc. (Cat. No 3514k-05a, neural crest origin). Human brain vascular pericytes were purchased from ScienCell (Cat. No 1200). Human neurons (SH-SY5Y) were purchased from ATCC (Cat. No CRL-2266). Human astrocytes were purchased from Cell Applications, Inc. (Cat. No 882A05f).

RNA sequencing and qPCR

Total RNA was extracted using a miRNeasy kit (Qiagen, Cat. No. / ID: 217084) following the manufacturer's protocol instructions. For RNA-seq, BGISEQ platform were used, averagely generating about 4.28G Gb bases per sample. The average mapping ratio with reference genome is 97.01%, the average mapping ratio with gene is 74.05%; 17029 genes were identified. We used HISAT to align the clean reads to the reference genome and Bowtie2 to align the clean reads to the reference genes. For qPCR, 100 ng of total RNA was used as the starting template for cDNA synthesis. The cDNA was prepared by reverse transcription (RT), and gene expression was analyzed by quantitative PCR (qPCR) on a SYBR green system (Applied Biosystems). Expression results were analyzed by the DDCT method, and GAPDH (encoding glyceraldehyde-3-phosphate dehydrogenase) was used as a house -keeping gene. Fold changes were calculated as the average relative to the control carotid as the baseline.

Computational Details

System building, structural refinements and molecular dynamic simulations (MDS)

The Q86WB7-1 (UNC93A, 457 aa), and A2RRH5-4 (WDR27, 827 aa) sequences [19,20] were used to build the 3D wild-type protein structures using the I-TASSER server [21]. The mutant variants (UNC93A V409I, and WDR27 R467H - T542S), were built based on these 3D models by performing site direct mutagenesis using UCSF Chimera software [22]. To avoid the residue overlapping in all protein systems, a structural refinement was carried out using ModRefiner server [23]. Classical MD simulations were performed using GROMACS 2020.4 package with the OPLS-AA force field parameters [24,25]. All protein systems were built in a triclinic simulation box considering periodic boundary conditions (PBC) in all directions (x, y, and z). Then, they were

solvated using the TIP4P water model [26], and Cl^- or Na^+ ions were used for neutralization of total charge in the simulation box. Mimicking of physiological conditions was performed by ionic strengthening with the addition of 150 mM NaCl. The distance of the protein surfaces to the edge of the periodic box was set at 1.5 nm. And 1 fs step was applied to calculate the motion equations using the Leap-Frog integrator [27]. The temperature for proteins and water-ions in all simulations was set at 309.65 K using the modified Berendsen thermostat (V-rescale algorithm) with a coupling constant of 0.1 ps [28]. The pressure was maintained at 1 bar using the Parrinello-Rahman barostat with compressibility of 4.5×10^{-5} bar^{-1} and coupling constant of 2.0 ps [29]. Particle mesh Ewald method was applied to long-range electrostatic interactions with a cutoff equal to 1.1 nm for nonbonded interactions with a tolerance of 1×10^5 for contribution in real space of the 3D structures. The Verlet neighbor searching cutoff scheme was applied with a neighbor-list update frequency of 10 steps (20 fs). Bonds involving hydrogen atoms were constrained using the Linear constraint solver (LINCS) algorithm [30]. Energy minimization in all simulations was performed with the steepest descent algorithm for a maximum of 100,000 steps. For the equilibration process, we performed two steps, one step of dynamics (1 ns) in the NVT (isothermal-isochoric) ensemble followed by 2 ns of dynamics in the NPT (isothermal-isobaric) ensemble. Then the final simulation was performed in the NPT ensemble for 500 ns followed by the analysis of the structures and their energy properties.

Structural and energetic analysis of 3D protein structures

All MD trajectories were corrected, and the 3D structures were recentering in the simulation boxes. RMSD, RMSF, radius of gyrations, H-bonds, residue distances, and solvent accessible surface area analyses were performed using the Gromacs tools and results were plotted using XMGrace software. For visualization of the structures, we used the UCSF Chimera, VMD software packages. Analysis of atomic interactions and 2D plots were generated using the LigPlot software packages [31]. The electrostatic potential (ESP) surfaces were calculated using the APBS (Adaptive Poisson-Boltzmann Solver) software, and the PDB2PQR software was used to assign the charges and radii to protein atoms [32].

Calculation of binding free energy

The Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) calculation of free energies and energy contribution by individual residues were carried out to analyze the impact of amino-acid substitutions on the different structures using the last 100ns of MD trajectories and the g_mmpbsa package [33]. Therefore, the interacting energy was calculated using the following equation:

$$\Delta G_{int} = G_{Prot} - (G_1)$$

Where the term G_1 is the free energy of the different sites of the protein, and G_{Prot} is the free energy of entire 3D structure. In this context, the free energy of each term was calculated as follow:

$$G_x = E_{MM} + G_{solv} - TS$$

where E_{MM} is the standard mechanical energy (MM) produced from bonded interactions, electrostatic interactions, and van der Waals interactions. G_{solv}, is the solvation energy that includes the free energy contributions of the polar and nonpolar terms. The TS term refers the entropic contribution and was not included in this calculation due to the computational costs [33,34]. Finally, 309 Kelvin (K) of temperature was used as the default parameter in all our calculations.

Results

1. Genetic analyses

Genealogic investigations allowed us to identify 5 members of a Peruvian family with strong clinical history of ADRD including dementia, Alzheimer´s disease and schizophrenia across two generations. To detect genetic risk factors associated to the development of the neurologic disorders observed in this family, we performed a whole genome sequencing analysis on affected and healthy members (n=14) of the family. By using the BGISEQ-500 platform, we obtained an average of 113,895.38 Mb of raw bases. After removing low-quality reads, we obtained an average of 106,676.25 Mb clean reads, identifying a total of 3,933,470 SNPs. Then, we selected coding variants that met the two following criteria. First, candidate variants harbored at least one “disruptive” or missense variant. Second, variants that were present in affected probands, but not in unaffected members of the family. As a result, we identified 3 coding variants that segregated across two generations of affected individuals. These variants were found to be located at chr6:167728791 (*UNC93A*; rs7739897), at chr6:170047902 (*WDR27*; rs61740334), and at chr6:170058374 (*WDR27*; rs3800544) (Fig. 1A-B). In addition, 2 different Sanger PCR sequencing platforms were used to validate the presence of these SNPs (Fig. 1C). In this line, several studies have also found SNPs in genes located on chromosome 6q with significant linkage with neurological diseases [35,36]. However, the *UNC93A* and *WDR27* variants have not yet been associated with ADRD. And it is worthy of note that none of the already known Alzheimer´s disease-associated variants were present in this Peruvian family with clinical history of ADRD.

To further confirm an association between the *UNC93A* and *WDR27* variants and familial genetics risk for neurodegenerative disorders we analyzed their presence in unrelated healthy Peruvians (n=50) and unrelated individuals with neurological disorders (probable Alzheimer´s disease, n=8). The UNC93A variant (V409I) was present in 1/50 of the healthy group, while the WDR27 variants (Thr542S and Arg467His) were present in 2/50 of the controls (Table 1). Interestingly, the 3 variants did not co-exist in any healthy individuals and were absence in volunteers diagnosed with probable Alzheimer´s disease with

not familial history of ADRD. These findings suggest the presence of a rare digenic mutation that are related to neurological disorders in a Peruvian family.

2. Structural analysis of the WDR27 and UNC93A variants

UNC93A gene encode a transmembrane protein (457 amino acids) that has 11 alpha-helices and is mainly expressed in the brain, kidney, and liver [37]. The *WDR27* gene encode a scaffold protein with multiple WD repeat domains and is ubiquitously expressed in the human body. We sought to use *in-silico* approaches to provide insights into the molecular and structural impact of UNC93A (V409I) and WDR27 (R467H and T542S) variants associated with familial ADRD. Therefore, we built the human UNC93A and WDR27 protein structures by homology comparative and performed a site direct mutagenesis to generate the mutated proteins using the UCSF Chimera software (Fig. 2A-B). Molecular dynamics simulations (MDS) for 500 ns were performed to stabilize the physical motions of atoms in both proteins to mimic physiological conditions. Importantly, we observed that at the beginning of the MDS (0-50 ns) the residue V409 of the wild-type UNC93 protein interacts with the lipid bilayer of the cell membrane. After 500ns of MDS the V409 is internalized toward to the protein active transmembrane conduct region where the subsequent interactions with ligands or ionic exchanges occur. These movements were characterized by a high structural vibration and epitope exposure of the UNC93A's ectodomain (aa 200-300) of the protein towards the surface of cell membrane (Fig. 2C). In contrast the mutant I409 block the internalization of this residue and loss its capacity to move towards to the protein active transmembrane conduct region due to the loss of about 50% of the global residual vibration (Fig. 2D). As results of this the amino acid change the UNC93A loss its capacity of ion exchange and interaction with potential ligands or partner. Regarding the WDR27 protein, both H467 and S542 variants induce the internalization of these residues that part of the hinge domain of the protein predicting the loss of function and capacity to interact with other partners (Fig. 2E). We also observed that the wild-type and mutant proteins are very stable due their closely residual vibration and epitope exposure patterns (Fig. 2F). The solvent accessible surface area (SASA) analysis demonstrates that the UNC93A (I409) protein increased the surface area of the mutated amino acid and its environment (Fig. 2G). While the

WDR27 variants reduced its SASA (Fig. 2H). The impact on the amino acid substitutions for both proteins was carried out by the Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) calculation of free energies and energy contribution using the last 150ns of MD trajectories. Remarkably the I409 increase the affinity of the protein to the membrane indicating a reduced capacity of internalization. While the WDR27 variants reduce the protein stabilization of both H467 and S542 mutated amino acids. Together these findings indicate a strong effect of the UNC93A and WDR27 variants on the functionality and ability to interact with their environment and thus potentially impacting the brain homeostasis.

3. Functional analysis of loss of WDR27 and UNC93A gene expression *in-vitro*

We next investigated how co-occurring inhibition of *UNC93A* and *WDR27* gene expression affect the cellular homeostasis of brain cell lines including, neurons, pericytes, astrocytes and VSMCs. In order to achieve this goal, we simultaneously silenced both *UNC93A* and *WDR27* genes using a siRNA approach and performed a high throughput RNA sequencing. Differential expression analysis demonstrates that VSMCs showed the highest numbers of differentially expressed genes (DEG=2231) followed by pericytes (DEG=2091). While astrocytes (DEG=226) and neurons (DEG=191) showed a modest change in gene expression compared with control groups (Fig. 3A), suggesting that mutations promoting the loss of function of *UNC93A* and *WDR27* genes impact the neurovascular unit of the brain. In this line, pathway analysis of VSMC and pericyte gene signatures identified enrichment for multiple known molecular pathways associated with neurodegenerative disorders such as impaired autophagy pathways signaling [38], ubiquitin-mediated proteolysis [39], unproductive metabolism signaling [40] inflammation and necroptosis [41] (Fig. 3B). While neurons showed a gene signature enriched for cellular senescence, cytokine-cytokine interactions and apoptosis. However, astrocytes showed a modest enrichment for necroptosis and proteasomal degradation pathway (Fig. 3B). In accordance with our results previous *in vivo* studies have reported the direct connection between autophagy activation and UNC93A

levels in healthy brain of mice under starvation indicating a potential role of UNC93A in the metabolic stability, energy uptake and nutrients transport in the brain.[37] Interestingly several neurogenerative disease have been associated with defects on autophagy and metabolism homeostasis including Alzheimer's disease, Parkinson's, and Huntington's diseases [42–44]. Moreover, we observed that multiple genes previously identified in GWS [45], and meta-analysis [46] studies were also dysregulated in brain cells deficient on UNC93A and WDR27 (Fig. 3C). Regarding the VSMCs and pericytes genes related with Alzheimer´s disease like CLU [47] SQSTM1 [48] GPC6 [49] and APCA7 [50] were modulated. Interestingly, the strongest genetic risk factor for Alzheimer´s disease, APOE, was also modulated in pericytes and neurons.

Discussion

Recent studies have shown racial disparities in Alzheimer's disease diagnosis between white and minority groups. Evidence suggests that there may be racial differences in risk factors associated with the development of Alzheimer's disease and related dementias (ADRD) [51–53]. Risk factors such as genetics, age, lifestyle, and co-morbid cardiovascular disease can be useful to understand the incidence, prevalence, and predisposition of an individual to ADRD. Despite the progress in research on racial differences in Alzheimer´s disease in developed countries, the diagnosis of Alzheimer´s disease in developing countries (e.g., Asian, African, and South American countries) deserves more recognition for its contribution in the global burden of Alzheimer´s disease [54]. In addition, the limited resources to address the mental health issue, the lack of adequate technology to diagnose ADRD and few funding agencies to support research studies are the major challenges that public health systems from developing countries are facing nowadays. According to the latest WHO report in 2018 ADRD deaths in Peru reached an astonishing 5.20% of total deaths (https://www.worldlifeexpectancy.com/peru-alzheimers-dementia) [55]. Remarkably the Peruvian population had a strong Amerindian ancestral background (80%), over other Latin American populations [56–58], indicating an opportunity to identify ancestry-specific genetic modifiers that associate with the development of ADRD.

Here we report an inheritance risk factor for ADRD in a Peruvian family with Amerindian ancestral background. We identified the combination of 3 pathogenic variants at the heterozygous state (*UNC93A*: rs7739897 and *WDR27*: rs61740334; rs3800544) that segregated across two generations in a family with a strong clinical history of ADRD. Remarkably, the combination of these variants was present in members with neurological disorders but absent in healthy individuals within the family. These findings were interesting because neurological disorders are commonly caused by monogenic or rarely by digenic variants as reported in Parkinson's disease, AD phenotype and frontotemporal dementia (FTD) [59]. Importantly, although these three SNPs are fairly common in European American and African American ancestry population (MAF = 1.78 -17.09) (https://evs.gs.washington.edu/EVS/) the combinatory effect of these variants were not previously associated with development of ADRD. Thus, these findings suggest that the combination of these variants is necessary to manifest the disease.

Our *in-silico* analysis of the 3D structure of the mutant UNC93A (V409I) and WDR27 (R467H and T542S) proteins demonstrates that changes in the amino acid sequences cause a dramatic impact on the conformational structure, predicting the loss of function of both proteins. However, the exact biological role of the UNC93A protein remains unknown. For instants, some studies identified a potential role of UNC93A as solute carrier and in ion homeostasis [60]. Its expression seemed to associate with increased metabolic activity in organs such as in the brain and kidneys [37]. In this context, we observed that the amino acid substitution (I409) in the UNC93A protein reduce its capacity of ion exchange and interaction with potential ligands or partner indicating a negative impact on the UNC93A bioactivity in the brain. Our *in-silico* analysis of the mutant WDR27 proteins demonstrated that both amino acid substitutions induce the internalization of the hinge domain of the protein impacting its segmental flexibility and ability to clamp down on its substrates or ligands. In the same manner, few is known on WDR27 biological functions in the brain. However, an SNP in the intergenic region adjoining WDR27 (rs924043) was associated with type 1 diabetes [61], and its duplication has been seen in obese

patients [62]. This evidence suggests the involvement of UNC93A and WRD27 in metabolic syndrome and related diseases.

Importantly, the brain is the most complex and metabolically active organ being equipped with a sophisticated network of specialized cell types such as neurons, microglia, astrocytes, pericytes, VSMCs and others. In the last years, diverse studies demonstrated the contribution of these cells to Alzheimer´s disease pathogenesis [63,64]. To investigate the potential impact of loss of function of UNC93A and WDR27, we used gene silencing technologies to reduce simultaneously the expression of both proteins in 4 brain cell types to mimic the clinical phenotype of member of the family with ADRD. Our KEGG pathway enrichment analysis showed that autophagy, mitophagy and metabolic pathways are the most impacted in both UNC93A and WDR27 inhibitory conditions. Interestingly, these pathways play an important role in Aβ clearance and thus its dysfunction may lead to the development Alzheimer´s disease. Furthermore, as a reduced autophagy activity was related to enhance cell death in response to intracellular stress [65], these variants could have a negative impact in the BBB integrity.

Our study has several limitations. First, due to the lack of access to the imageological studies we were not able to correlate the variants with damaged areas of the brain. However, the MoCa test corroborated that brain areas associated with cognitive domains, predominantly temporal and frontal lobe areas, are damaged. Second, we could not find a validation family for the combination of these variants. However, is possible that these combinatory variants are only present in the reported family. It could be a similar case reported for the PSEN1 (E280A) mutation that only affects a Colombian family descendant of a Spanish conquistador [66], or the mutation in the PSEN2 gene (N141I) that is only present in families with German's descendants that emigrated in the 1760s to a southern Volga region in Russia [67]. Despite these limitations, this study report for the first time a new genetic rick loci associated with ADRD and highly the importance of the UNC93A and WDR27 genes in brain biology.

List of abbreviations

ADRD: Alzheimer's disease and related dementias

APOE: apolipoprotein E

APP: amyloid precursor protein

DEGs: differentially expressed genes

EOAD: early onset Alzheimer's disease

GWAS: genome wide association studies

LOAD: late onset Alzheimer's disease (AD)

MDS: molecular dynamics simulations

PSEN: presenilin

SASA: solvent accessible surface area

VSMCs: vascular smooth muscle cells

WGS: whole-genome sequencing

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of research of the EsSalud Hospital (CIEI) of Arequipa-Peru. Saliva samples were obtained from participants that previously signed the informed consent.

Consent for publication

Not applicable.

Availability of data and materials

All raw data are stored in cloud storage and will be made available upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by Consejo Nacional de Ciencia, Tecnologia e Innovacion Tecnologica de Peru (grant N◦ 024-2019-Fondecyt-BM-INC.INV). Dr. Lino Cardenas was supported by the MGH Physician-Scientist Development Award and the Ruth L. Kirschstein National Research Service Award (5T32HL007208-43) and the Physician-Scientist Development Award (PSDA-MGH).

Authors' contributions

CLLC conceived the work with KLFA and GDDC. CLLC and KLFA designed the work. CLLC and KLFA performed the experiments. KLFA, MMOM, LDGM, MACF, and KJVL collected the samples. JAAP performed the *in silico* analysis. MFPC performed and analyzed the neurological test. PLM analyzed the neurological test. CLLC and KLFA analyzed and interpreted the data of the study. APM provide the clinical samples. GDDC and CLLC supervised the study. CLLC and KLFA wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

We are very grateful to all the volunteers who participated in this study. In addition, we thank Drs. Claudia Caracela-Zeballos, Jessica L. Lewis-Paredes, Juan M Pacheco-Salazar, Froilán Huaraya, Rita Montesinos-Nieto, and Badhin Gomez for their contribution to this study.

References

1. Kochunov P, Zavaliangos-Petropulu A, Jahanshad N, Thompson PM, Ryan MC, Chiappelli J, et al. A white matter connection of schizophrenia and Alzheimer's Disease. Schizophr Bull. 2021;47:197–206.

2. Report AA. 2021 Alzheimer's disease facts and figures. Alzheimer's Dement. 2021;17:327–406.

3. Mendez MF. Early-onset Alzheimer Disease. Neurol Clin. 2017;35:263–81.

4. Rabinovici BGD. Late-onset Alzheimer Disease. Conitunuum (Mineapp Minn). 2019;14–33.

5. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry. 2006;63:168–74.

6. Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, et al. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. Nature. 1995. p. 775–8.

7. Sherrington R, Rogaev EI, Liang Y, Rogaeva E, Levesque. G, Ikeda. M, et al. Cloning of a gene bearing missense mutations in early onset familial Alzheimer's disease. Funct. Neurol. 1995. p. 754–60.

8. Goate A, Chartier-Harlin M-C, Mullan M, Brown J, Crawford F, Fidani L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer´s disease. Nature. 1991;354:704–6.

9. Guerreiro RJ, Gustafson DR, Hardy J. The genetic architecture of

Alzheimer's disease: Beyond APP, PSENS and APOE. Neurobiol Aging. 2012;33:437–56.

10. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: A meta-analysis. J Am Med Assoc. 1997;278:1349–56.

11. Granot-Hershkovitz E, Tarraf W, Kurniansyah N, Daviglus M, Isasi CR, Kaplan R, et al. APOE alleles' association with cognitive function differs across Hispanic/Latino groups and genetic ancestry in the study of Latinos-investigation of neurocognitive aging (HCHS/SOL). Alzheimer's Dement. 2021;17:466–74.

12. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's Disease. N Engl J Med. 2013;368:117–27.

13. Miyashita A, Wen Y, Kitamura N, Matsubara E, Kawarabayashi T, Shoji M, et al. Lack of genetic association between TREM2 and late-onset Alzheimer's disease in a Japanese population. J Alzheimer's Dis. 2014;41:1031–8.

14. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019;51:414–30.

15. Bellenguez C, Grenier-Boley B, Lambert JC. Genetics of Alzheimer's disease: where we are, and where we are going. Curr Opin Neurobiol. 2020;61:40–8.

16. Cady J, Koval ED, Benitez BA, Zaidman C, Jockel-balsarotti J, Allred P, et al. The TREM2 variant p.R47H is a risk factor for sporadic amyotrophic lateral sclerosis. 2014;71:449–53.

17. Drange OK, Bjerkehagen Smeland O, Shadrin AA, Finseth PI, Witoelar A, Frei O, et al. Genetic overlap between alzheimer's disease and bipolar disorder implicates the MARK2 and VAC14 genes. Front Neurosci. 2019;13:1–11.

18. McKhann G, Knopman DS, Chertkow H, Hyman BT, Jr CRJ, Kawas CH, et

al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging- Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:263–9.

19. WDR27 - WD repeat-containing protein 27 - Homo sapiens (Human) [Internet]. [cited 2020 Jun 10]. Available from: https://www.uniprot.org/uniprot/A2RRH5#A2RRH5-4

20. UNC119 - Protein unc-119 homolog A - Homo sapiens (Human) [Internet]. [cited 2020 Jun 10]. Available from: https://www.uniprot.org/uniprot/Q86WB7#Q86WB7-1

21. Zheng W, Zhang C, Li Y, Pearce R, Bell EW, Zhang Y. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. Cell Reports Methods 2021;1:100014.

22. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem. 2004;25:1605–12.

23. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J [Internet]. Biophysical Society; 2011;101:2525–34.

24. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. J Comput Chem. 2019;40:2418–31.

25. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc. 1996;118:11225–36.

26. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79:926–35.

27. Hockney RW, Goel SP, Eastwood JW. Quiet high-resolution computer models of a plasma. J Comput Phys. 1974;14:148–58.

28. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81:3684–90.

29. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126.

30. Hess B. P-LINCS: A parallel linear constraint solver for molecular simulation. J Chem Theory Comput. 2008;4:116–22.

31. Wallace AC, Laskowski RA, Thornton JM. Ligplot: A program to generate schematic diagrams of protein-ligand interactions. Protein Eng Des Sel. 1995;8:127–34.

32. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 2001;98:10037–41.

33. Kumari R, Kumar R, Lynn A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model. 2014;54:1951–62.

34. Rastelli G, Del Rio A, Gianluca D, Miriam S. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. Wiley Intersci. 2010;31:797–810.

35. Kohn Y, Lerer B. Excitement and confusion on chromosome 6q: The challenges of neuropsychiatric genetics in microcosm. Mol Psychiatry. 2005;10:1062–73.

36. Naj AC, Beecham GW, Martin ER, Gallins PJ, Powell EH, Konidari I, et al. Dementia revealed: Novel chromosome 6 locus for Late-onset alzheimer disease provides genetic evidence for Folate-pathway abnormalities. PLoS Genet. 2010;6.

37. Ceder MM, Lekholm E, Hellsten S V., Perland E, Fredriksson R. The Neuronal and Peripheral Expressed Membrane-Bound UNC93A Respond to Nutrient Availability in Mice. Front Mol Neurosci. 2017;10.

38. Li Q, Liu Y, Sun M. Autophagy and Alzheimer’s Disease. Cell Mol Neurobiol. Springer US; 2016;37:377–88.

39. Upadhya SC, Hegde AN. Role of the ubiquitin proteasome system in Alzheimer's disease. BMC Biochem. 2007;8.

40. Van Der Velpen V, Teav T, Gallart-Ayala H, Mehl F, Konz I, Clark C, et al. Systemic and central nervous system metabolic alterations in Alzheimer's disease. Alzheimer's Res Ther. Alzheimer's Research & Therapy; 2019;11.

41. Zhang G, Zhang Y, Shen Y, Wang Y, Zhao M, Sun L. The potential role of ferroptosis in Alzheimer's Disease. J Alzheimer's Dis. 2021;907–25.

42. Xu Y, Propson NE, Du S, Xiong W, Zheng H. Autophagy deficiency modulates microglial lipid homeostasis and aggravates tau pathology and spreading. Proc Natl Acad Sci U S A. 2021;118:1–10.

43. Wang B, Abraham N, Gao G, Yang Q. Dysregulation of autophagy and mitochondrial function in Parkinson's disease. Transl Neurodegener [Internet]. Translational Neurodegeneration; 2016;5:1–9.

44. Croce KR, Yamamoto A. A role for autophagy in Huntington's disease Katherine. Neurobiol Dis. 2019;122:16–22.

45. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat Genet. 2021;53:1276–82.

46. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45:1452–8.

47. Jun G, Naj AC, Beecham GW, Wang LS, Buros J, Gallins PJ, et al. Meta-analysis confirms CR1, CLU, and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes. Arch Neurol. 2010;67:1473–84.

48. Cuyvers E, van der Zee J, Bettens K, Engelborghs S, Vandenbulcke M, Robberecht C, et al. Genetic variability in SQSTM1 and risk of early-onset Alzheimer dementia: A European early-onset dementia consortium study. Neurobiol Aging. 2015;36:2005.e15-2005.e22.

49. Kunkle BW, Schmidt M, Klein HU, Naj AC, Hamilton-Nelson KL, Larson EB, et al. Novel Alzheimer Disease risk loci and pathways in african american

individuals using the African genome resources panel: A meta-analysis. JAMA Neurol. 2021;78:102–13.

50. De Roeck A, Van Broeckhoven C, Sleegers K. The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. Acta Neuropathol [Internet]. Springer Berlin Heidelberg; 2019;138:201–20.

51. Barnes LL. Alzheimer disease in African American individuals: increased incidence or not enough data? Nat Rev Neurol. Springer US; 2022;18:56–62.

52. Mayeda, Elizabeth Rose, M Maria Glymor CPQ, Whitmer RAW. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. Physiol Behav. 2017;176:139–48.

53. Brewster P, Barnes L, Haan M, Johnson JK, Manly JJ, Nápoles AM, et al. Progress and future challenges in aging and diversity research in the United States. Alzheimers Dement. 2019;15:995–1003.

54. Chávez-Fumagalli MA, Shrivastava P, Aguilar-Pineda JA, Nieto-Montesinos R, Del-Carpio GD, Peralta-Mestas A, et al. Diagnosis of Alzheimer's disease in developed and developing countries: systematic review and meta-analysis of diagnostic test accuracy. J Alzheimer's Dis Reports. 2021;5:15–30.

55. World heatlh rankings [Internet]. [cited 2021 Dec 15]. Available from: https://www.worldlifeexpectancy.com/peru-alzheimers-dementia

56. Homburger JR, Moreno-Estrada A, Gignoux CR, Nelson D, Sanchez E, Ortiz-Tello P, et al. Genomic insights into the ancestry and demographic history of South America. PLoS Genet. 2015;11:1–26.

57. Norris ET, Rishishwar L, Chande AT, Conley AB, Ye K, Valderrama-Aguirre A, et al. Admixture-enabled selection for rapid adaptive evolution in the Americas. Genome Biol. Genome Biology; 2020;21:1–12.

58. Norris ET, Wang L, Conley AB, Rishishwar L, Mariño-Ramírez L, Valderrama-Aguirre A, et al. Genetic ancestry, admixture and health determinants in Latin America. BMC Genomics. 2017;19.

59. Tang B, Xiong H, Sun P, Zhang Y, Wang D, Hu Z, et al. Association of

PINK1 and DJ-1 confers digenic inheritance of early-onset Parkinson's disease. Hum Mol Genet. 2006;15:1816–25.

60. Ceder MM, Aggarwal T, Hosseini K, Maturi V, Patil S, Perland E, et al. CG4928 is vital for renal function in fruit flies and membrane potential in cells: A first in-depth characterization of the putative solute carrier UNC93A. Front Cell Dev Biol. 2020;8.

61. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet. 2011;7:e1002293.

62. D'Angelo CS, Varela MC, De Castro CIE, Otto PA, Perez ABA, Lourenço CM, et al. Chromosomal microarray analysis in the genetic evaluation of 279 patients with syndromic obesity. Mol Cytogenet. Molecular Cytogenetics; 2018;11.

63. Zenaro E, Piacentino G, Constantin G. The blood-brain barrier in Alzheimer's disease. Neurobiol Dis. 2017;107:41–56.

64. Aguilar-Pineda JA, Vera-Lopez KJ, Shrivastava P, Chávez-Fumagalli MA, Nieto-Montesinos R, Alvarez-Fernandez KL, et al. Vascular smooth muscle cell dysfunction contribute to neuroinflammation and Tau hyperphosphorylation in Alzheimer disease. iScience. 2021;24.

65. Galati S, Boni C, Gerra MC, Lazzaretti M, Buschini A. Autophagy: a player in response to oxidative stress and DNA damage. Oxid Med Cell Longev. Hindawi; 2019;Article ID.

66. Lall MA, Cox HC, Arcila ML, Cadavid L, Moreno S, Garcia G, et al. Origin of the PSEN1 E280A mutation causing early–onset Alzheimer's disease. Alzheimers Dement [Internet]. 2014;10:S277–S283.e10.

67. Tomita T, Maruyama K, Saido TC, Kume H, Shinozaki K, Tokuhiro S, et al. The presenilin 2 mutation (N141I) linked to familial Alzheimer disease (Volga German families) increases the secretion of amyloid β protein ending at the 42nd (or 43rd) residue. Proc Natl Acad Sci U S A. 1997;94:2025–30.

Table 1. Variants in unrelated volunteers.

Gene	Variant	Reference SNP number	Unrelated Controls (n=50)	Unrelated volunteers with ADRD (n=8)
UNC93A	V409I	rs7739897	1	0
WDR27	Thr542S	rs61740334	2	0
WDR27	Arg467His	rs3800544	2	0

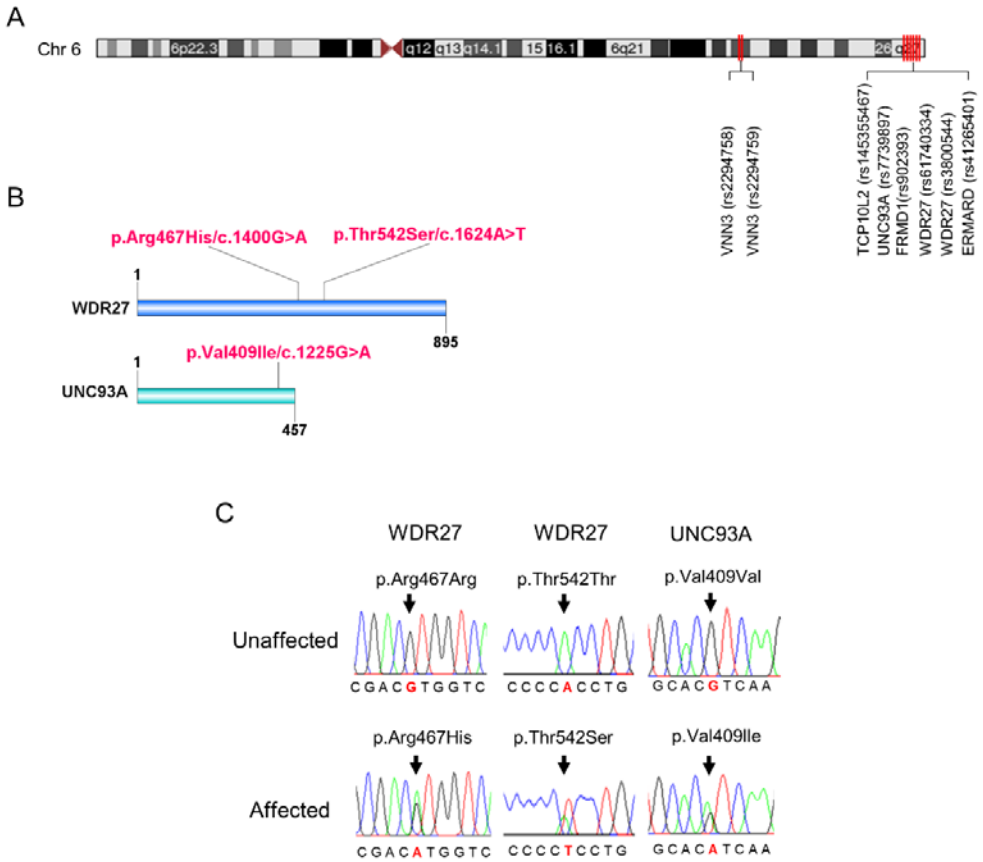
SNP: single nucleotide polymorphism; rs: reference sequence

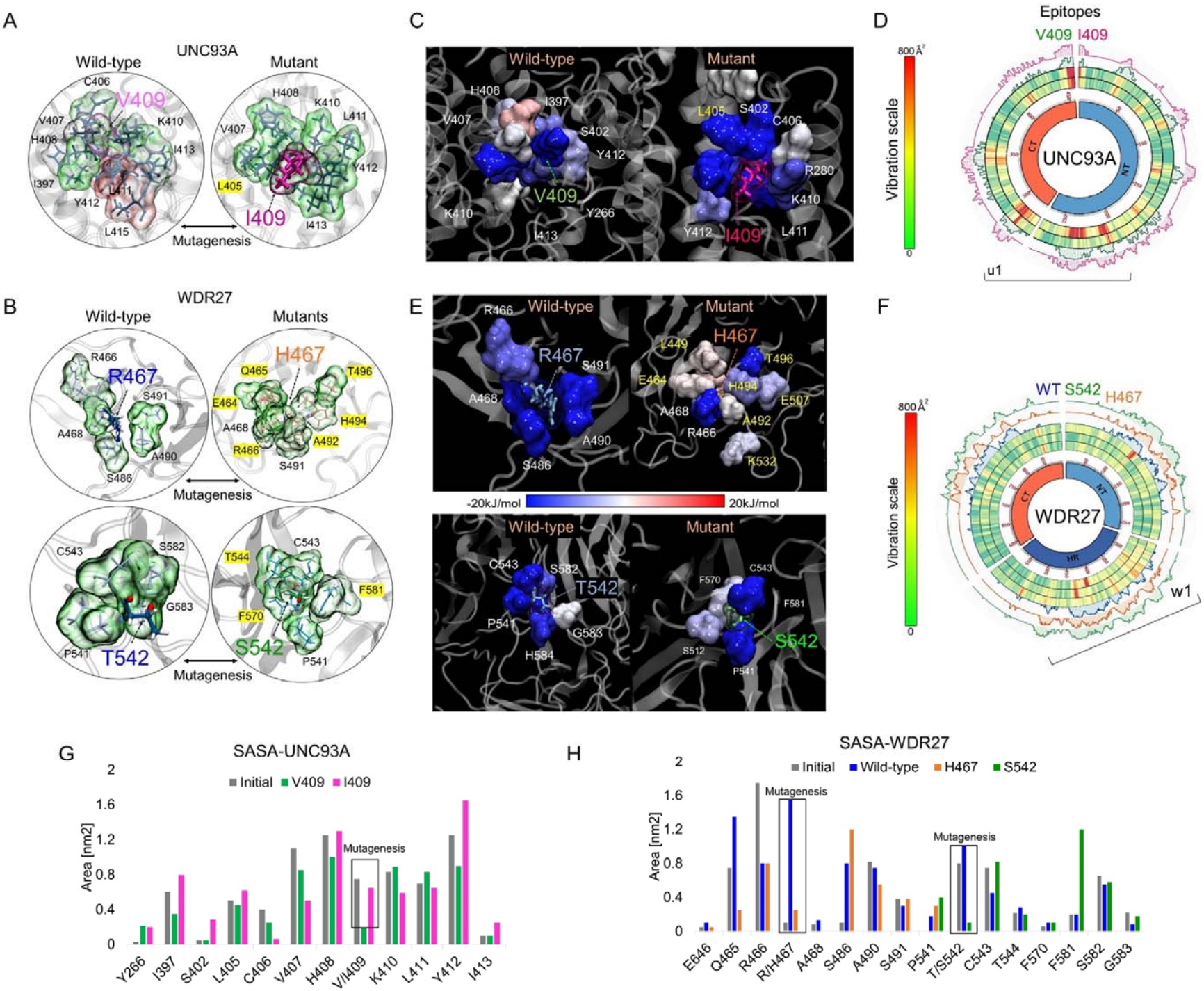
Figure legends

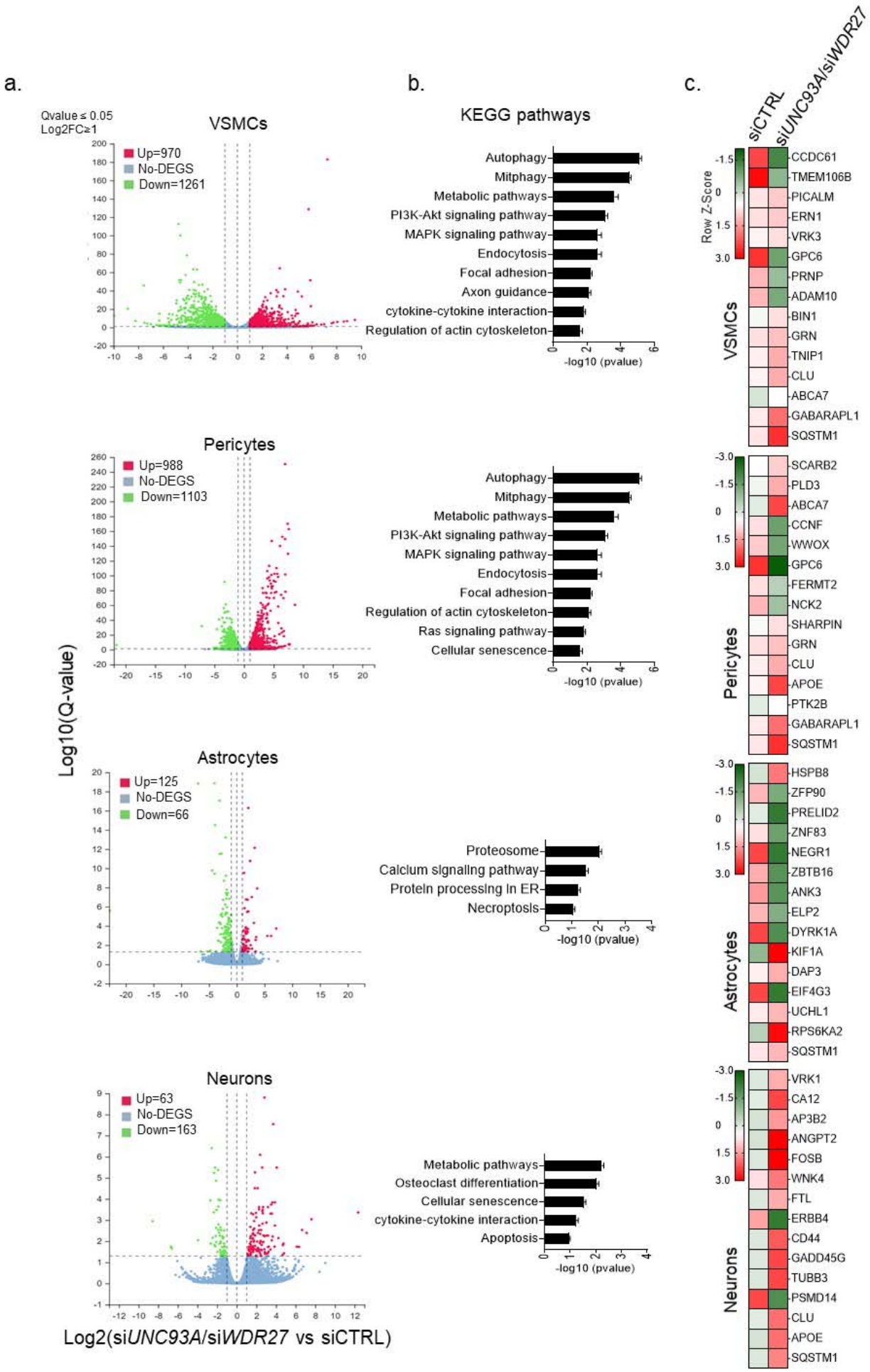
Figure 1. Variants identified in family affected members. (A) Chromosomal position of UNC93A and WDR27 genes. (B) Location of a mutation in the protein structure. (C) An electropherogram from affected probands showing a base pair change in the UNC93A gene (Val409Ile/c.1225G>A) and in the WDR27 gene (Thr542Ser/c.1624A>T; Arg467His/c.1400G>A), compared to healthy controls.

Figure 2. In silico analysis of UNC93A and WDR27 amino acid substitutions predict loss of protein function. (A and B) show the UNC93A and WDR27 wild-type and mutant proteins**.** (C and E) show the MM-PBSA calculation of main energetic interactions of residues at mutation sites. Blue color indicates favorable energies and red color unfavorable energies. (D and F) show Circos plot of the wild-type and mutant full-length protein. The heat map represents the vibrational movement of each residue throughout MD simulations at the scale bar values. The outer histograms show the region's most likely (>50%) to be epitopes. (G and H) show the solvent-accessible surface area (SASA) average values of the wild-type and mutation residues and their neighbors.

Figure 3. Loss of UNC93A and WDR27 impact the global transcriptomic signature of brain cell lines in vitro. (A) Volcano plot of dysregulated genes in 4 different brain cell types. (B) Shows the KEGG pathways analysis. (C) Heatmap of dysregulated genes previously associated with neurodegenerative disease.


A
Chr 6
6p22.3
q12
q13
q14.1
15
16.1
6q21
26
VNN3 (rs2294758)
VNN3 (rs2294759)
TCP10L2 (rs145355467)
UNC93A (rs77739897)
FRMD1(rs902393)
WDR27 (rs61740334)
WDR27 (rs3800544)
ERMARD (rs41265401)
B
p.Arg467His/c.1400G>A
p.Thr542Ser/c.1624A>T
1
WDR27
895
p.Val409Ile/c.1225G>A
1
UNC93A
457
C
WDR27
WDR27
UNC93A
p.Arg467Arg
p.Thr542Thr
p.Val409Val
Unaffected
CGACGTGGTC
CCCCACCTG
GCACGTCAA
p.Arg467His
p.Thr542Ser
p.Val409Ile
Affected
CGACATGGTC
CCCCTCCTG
GCACATCAA


A
UNC93A
Wild-type
Mutant
V409
I409
Mutagenesis
B
WDR27
Wild-type
Mutants
R467
H467
T542
S542
Mutagenesis
C
Wild-type
Mutant
V409
I409
D
Epitopes
V409 I409
Vibration scale
UNC93A
u1
E
Wild-type
Mutant
R467
H467
-20kJ/mol
20kJ/mol
T542
S542
F
WT S542 H467
Vibration scale
WDR27
w1
G
SASA-UNC93A
Initial V409 I409
Area [nm2]
Mutagenesis
Y266 I397 S402 L405 C406 V407 H408 V/I409 K410 L411 Y412 I413
H
SASA-WDR27
Initial Wild-type H467 S542
Area [nm2]
Mutagenesis
E646 Q465 R466 R/H467 A468 S486 A490 S491 P541 T/S542 C543 T544 F570 F581 S582 G583


a.
Qvalue ≤ 0.05
Log2FC≥1
VSMCs
Up=970
No-DEGS
Down=1261
Pericytes
Up=988
No-DEGS
Down=1103
Astrocytes
Up=125
No-DEGS
Down=66
Neurons
Up=63
No-DEGS
Down=163
Log10(Q-value)
Log2(siUNC93A/siWDR27 vs siCTRL)
b.
KEGG pathways
Autophagy
Mitphagy
Metabolic pathways
PI3K-Akt signaling pathway
MAPK signaling pathway
Endocytosis
Focal adhesion
Axon guidance
cytokine-cytokine interaction
Regulation of actin cytoskeleton
Ras signaling pathway
Cellular senescence
Proteosome
Calcium signaling pathway
Protein processing in ER
Necroptosis
Osteoclast differentiation
Apoptosis
-log10 (pvalue)
c.
siCTRL
siUNC93A/siWDR27
Row Z-Score
VSMCs
CCDC61
TMEM106B
PICALM
ERN1
VRK3
GPC6
PRNP
ADAM10
BIN1
GRN
TNIP1
CLU
ABCA7
GABARAPL1
SQSTM1
Pericytes
SCARB2
PLD3
ABCA7
CCNF
WWOX
GPC6
FERMT2
NCK2
SHARPIN
GRN
CLU
APOE
PTK2B
GABARAPL1
SQSTM1
Astrocytes
HSPB8
ZFP90
PRELID2
ZNF83
NEGR1
ZBTB16
ANK3
ELP2
DYRK1A
KIF1A
DAP3
EIF4G3
UCHL1
RPS6KA2
SQSTM1
Neurons
VRK1
CA12
AP3B2
ANGPT2
FOSB
WNK4
FTL
ERBB4
CD44
GADD45G
TUBB3
PSMD14
CLU
APOE
SQSTM1