ARTICLE

A novel heterozygous missense variant of the *ARID4A* gene identified in Han Chinese families with schizophrenia-diagnosed siblings that interferes with DNA-binding activity

Decheng Ren[1,2,3,5], Xiaoxi Wei[1,2,3,5], Lin Lin[1,2,3,5], Fan Yuan[1,2,3], Yan Bi[1,2,3], Zhenming Guo[1,2,3], Liangjie Liu[1,2,3], Lei Ji[1,2,3], Xiao Yang[1,2,3], Ke Han[1,2,3], Fengping Yang[1,2,3], Xi Wu[1,2,3], Xingwang Li[1,2,3], Zhenghui Yi[3], Yifeng Xu[3], Changgun Cai[4], Peng Wang[4], Weidong Li[1,2,3], Lin He[1,2,3], Daizhan Zhou[1,2,3], Tao Yu[1,2,3], Yi Shi[1,2,3], Qing Lu[1,2,3]✉ and Guang He[1,2,3]✉

ARID4A plays an important role in regulating gene expression and cell proliferation. ARID4A belongs to the AT-rich interaction domain (ARID)-containing family, and a PWWP domain immediately precedes its ARID region. The molecular mechanism and structural basis of ARID4A are largely unknown. Whole-exome sequencing (WES) revealed that a novel heterozygous missense variant, *ARID4A* c.1231 C > G (p.His411Asp), was associated with schizophrenia (SCZ) in this study. We determined the crystal structure of the PWWP-ARID tandem at 2.05 Å, revealing an unexpected mode in which ARID4A assembles with its PWWP and ARID from a structural and functional supramodule. Our results further showed that compared with the wild type, the p.His411Asp ARID mutant protein adopts a less compact conformation and exhibits a weaker dsDNA-binding ability. The p.His411Asp mutation decreased the number of cells that were arrested in the G0-G1 phase and caused more cells to progress to the G2-M phase. In addition, the missense mutation promoted the proliferation of HEK293T cells. In conclusion, our data provide evidence that ARID4A p.His411Asp could cause a conformational change in the ARID4A ARID domain, influence the DNA binding function, and subsequently disturb the cell cycle arrest in the G1 phase. *ARID4A* is likely a susceptibility gene for SCZ; thus, these findings provide new insight into the role of *ARID4A* in psychiatric disorders.

Molecular Psychiatry; https://doi.org/10.1038/s41380-022-01530-w

INTRODUCTION

Schizophrenia (SCZ) is a severe, chronic mental disorder that deeply affects the behavior of individuals and their cognitive and emotional processes [1]. The main clinical features in patients include delusions, hallucinations, and disorganized speech and behavior [2]. SCZ is a complex multifactorial disorder. Although studies on SCZ have been conducted for many years, the underlying pathophysiology of SCZ remains largely elusive. Previous studies have found that both genes and a range of environmental factors play significant roles in the development of SCZ. The heritability of SCZ has been estimated to be ~80% [3]. The high heritability of SCZ suggests that genetic risk factors contribute to a significant proportion of the etiology. Although scientists have identified many SCZ-related variants, functional studies of these mutation sites are still missing.

The 4A AT-rich interaction domain (ARID4A) is a DNA-binding protein that modulates the activities of several transcription factors, including retinoblastoma-associated protein (RB1) and androgen receptor (AR) [4, 5]. Previous studies suggested that ARID4A can function as a component in the mSin3A repressor complex, which suppresses gene expression and regulates epigenetic imprinting [6–8]. In cancer research, ARID4A may serve as a tumor suppressor in various human cancers, such as prostate cancer and breast cancer [9–11]. ARID4A also contributes to the regulation of epigenetic programming in leukemia and the regulation of the epigenetics in Prader–Willi syndrome [7, 12], which is known as an imprinted neurodevelopmental disorder and is characterized by metabolic, sleep, and neuropsychiatric features [13, 14]. Mutations in *ARID4A* can be associated with bipolar and major depression conditions according to the data of the Genebass database [15], which contains 3,817 phenotypes based on exome sequencing of 28,1852 samples in UK Biobank [16]. The ARID4A protein is ubiquitously expressed in almost all tissues, indicating that this protein participates in numerous biological processes. Moreover, according to the Human Protein Atlas database, the high level of ARID4A expression in the cerebral cortex and neuronal cells suggests that it may be vital for neural development and function.

ARID4A belongs to the ARID-containing family, which comprises 15 distinct human proteins with a conserved ARID domain [17, 18]. ARID proteins are implicated in the control of cell growth, proliferation, and differentiation [19, 20]. The AT-Rich Interaction Domain (ARID) is an

[1]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Shanghai Jiao Tong University, Shanghai, China. [2]The Collaborative Innovation Center for Brain Science, and Brain Science and Technology Research Center, Shanghai Jiao Tong University, Shanghai, China. [3]Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [4]Wuhu Fourth People's Hospital, Wuhu, Anhui, China. [5]These authors contributed equally: Decheng Ren, Xiaoxi Wei, Lin Lin. ✉email: luqing67@sjtu.edu.cn; heguangbiox@163.com

Received: 1 September 2021 Revised: 10 March 2022 Accepted: 16 March 2022
Published online: 01 April 2022

ancient helix-turn-helix motif-based DNA-binding domain [21, 22]. Although the ARID domain was first identified as a DNA-binding domain with AT-rich selectivity, ARID family members exhibit divergent DNA-binding preferences. The ARID domain of ARID4A showed no DNA sequence preference [17, 23]. In addition to the ARID domain, ARID4A contains several other functional domains, including a Tudor domain, a PWWP domain (also known as the HATH domain or RBB1NT domain), a chromo barrel domain, and a C-terminal R2 domain [23–25]. A previous study reported that the Tudor domain of ARID4A interacted with DNA without sequential specificity [26]. The PWWP domain interacts with chromatin by synergistically binding to histones and DNA [27]. Recently, to identify SCZ susceptibility genes, we performed a systematic family-based study using whole-exome sequencing (WES) in Han Chinese patient-sibling family-based cohorts. We found that the *ARID4A* c. C1231G (p. His411Asp) was associated with SCZ. The single point mutation p. His411Asp on ARID4A is located at the C-terminal in the ARID domain. Owing to the lack of structural and biochemical data on the ARID4A ARID domain, it is difficult to know how point mutations perturb the physiological function of ARID4A. Therefore, solving the structure of the ARID4A ARID domain is crucial to understanding the structural mechanism of its repression activity.

Even though many functional studies have been performed with *ARID4A*, the structural data available for the ARID4A ARID domain is lacking, impeding a better understanding of its repression activity mechanism. Moreover, a previous study found that a certain single-point mutation could also influence the gene function [28]. Considering the lack of structural and point mutation studies on the ARID4A ARID domain, we aimed to investigate the function of the ARID4A p.His411Asp mutation in this study.

MATERIALS AND METHODS

Subjects

Two families with SCZ-diagnosed siblings were selected from the mental disease biobank of Bio-X institutes. All the subjects were of Han Chinese origin. Family_1 contained two siblings, and in this family, both siblings and one of their parents were diagnosed with SCZ; family_2 contained three siblings, and two siblings and one of their parents were diagnosed with SCZ (Supplementary Material Table 1). The families were recruited from Shanghai Mental Health Center, Shanghai Changning Mental Health Center, and Wuhu Fourth People's Hospital from 2001 to 2003. All of the SCZ patients were interviewed by two independent psychiatrists and diagnosed according to DSM-IV criteria. All participants were informed of the details of the study and provided informed consent, which was approved by the ethics review committee of human-related scientific and technological research of Shanghai Jiao Tong University.

Whole-exome sequencing and data analysis

Whole blood samples were collected with EDTA Blood Collection Tubes (Becton, Dickinson and Company, NJ, USA), and DNA was extracted by standard procedures using FlexiGene DNA kits (Fuji, Tokyo, Japan). The DNA samples from the siblings and parents in each family were used for follow-up experiments. DNA libraries were prepared using protocols recommended by Illumina (Illumina, San Diego, CA). Whole-exome enrichment was performed using the TruSeq Exome Enrichment Kit (Illumina, San Diego, CA). Captured DNA libraries were sequenced with HiSeq2500 (Illumina, San Diego, CA, USA). All sequencing reads were mapped to the NCBI reference genome with Burrows–Wheeler Aligner (BWA). Quality score recalibration, Indel realignment, and duplicate removal were performed with the Genome Analysis Toolkit (GATK). The variants were matched to 1000 Genomes (http://www.1000genomes.org/) and were annotated to RefSeq hg19 and evaluated with different algorithms (Sift, PROVEAN, Polyphen-2, etc.) to predict the functional severity. Only variants with MAF < 0.01 were retained for further analysis. Sanger sequencing was used to verify the existence of the variants.

Mutagenesis

The coding sequence of ARID4A (NP_002883.3) was PCR amplified from cDNA and cloned into a p3xFlag-CMV-14 vector between the BamH 1 Hind III sites by using the primers 5′-CCAAGCTTGCCACCATGAAGGCGGCAG AT-3′ (forward primer) and 5′-CGCGGATCCCCTGCATTCTACAGCAAGTA-3′ (reverse primer). Plasmids encoding human *ARID4A* with the p. His411Asp variant were constructed with the Hieff Mut™ Site-Directed Mutagenesis Kit (YEASEN) using the primers 5′-AGTTCAGAACTGTTGATCACCATGAA-3′ (forward primer) and 5′- TTTGGTTCATGGTGATCAACAGTTCT -3′ (reverse primer).

Constructs and protein expression

The coding sequences of *ARID4A* His411 and *ARID4A* His411Asp were PCR amplified from the above p3xFlag-CMV-14-*ARID4A* or p3xFlag-CMV-14-*ARID4A* p. His411Asp vector and were cloned into a pET vector between the BamH1 and Xho1 sites. Proteins were expressed in BL21 (DE3) Escherichia coli cells. The cells were grown in LB medium at 37 °C until the optical density at 600 nm (A600) reached 0.7–0.8, then protein production was induced by the addition of 0.25 mM isopropyl β-D-thiogalactoside and grown at 16 °C for ~24 h. The His6-tagged proteins were purified with a Ni2 + nitrilotriacetic acid–agarose column followed by size-exclusion chromatography.

FPLC coupled with static light scattering

Protein samples (100 µL at a concentration of 60 µM, pre-equilibrated with corresponding column buffer) were injected into an AKTA FPLC system with a Superose 12 10/300 GL column (GE Healthcare) that contained 50 mM Tris-HCl, 100 mM NaCl, and 1 mM DTT column buffers at pH 8.0. The chromatography system was coupled to a static light-scattering detector (miniDawn, Wyatt) and differential refractive index detector (Optilab, Wyatt). Data were analyzed with ASTRA 7 (Wyatt).

Crystallography, data collection and structure determination

ARID4A crystals (12 mg/ml in 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DTT) were obtained using the hanging-drop method at 16 °C by mixing 1 µL of protein sample with an equal volume of 2.4 M ammonium phosphate dibasic and 0.1 M Tris-HCl, pH 8.5. Crystals were transferred to the reservoir solution containing 20% glycerol as the cryoprotectant and flash cooled with liquid nitrogen. Diffraction data were collected at the Shanghai Synchrotron Radiation Facility (SSRF, China). The diffraction data were processed by XDS. The structure of ARID4A was solved by the molecular replacement method using the structure of the ARID domain of ARID3A (PDB: 4LJX) as a searching model by PHASER. CCP4 Autobuild was used to improve the structure completeness. Further refinement was iteratively performed using CCP4 Refmac and Coot. Structural diagrams were prepared by PyMOL (http://www.pymol.org).

Isothermal titration calorimetry assay

Synthetic dsDNA as an interacting ligand used in the titration contained two complementary DNA sequences (15 bp) 5′-CCTGTATTGATGTGG-3′ and 3′-GGACATAACTACACC-5′. Isothermal titration calorimetry (ITC) measurements were carried out on a MicroCal-ITC200 at 25 °C. The titration buffer contained 50 mM Tris-HCl, 100 mM NaCl, and 1 mM DTT, pH 8.0, and total valid injections were 20, 2 µL of each volume, reaction time lasted 4 s, and 120 s of spacing time between reaction intervals to ensure the titration peak returned to the baseline. The titration data were analyzed by Origin 7.0 (MicroCal).

Cell culture and transfection

HEK293T cells were maintained in high glucose Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin–streptomycin (Invitrogen). Cells were plated in a 6-well plate the day before transfection at a density of 5 × 10^5 cells per well and were transiently transfected with 3.3 µg p3xFlag-CMV-14-ARID4A or p3xFlag-CMV-14-ARID4A p. His411Asp vector concentration was in the range from 800 to 1000 ng/µL of each plasmid per well using FuGENE® HD Transfection Reagent (Promega). All cells were cultured at 37 °C in a humidity chamber with 5% CO2 and harvested 48 h posttransfection.

5-Ethynyl-2′-deoxyuridine assay

For 5-ethynyl-2′-deoxyuridine (EdU) analysis, after transfection and synchronization, the cells were harvested, and an EdU assay was performed using a BeyoClick™ EdU-594 Cell Proliferation Kit (Beyotime) according to the manufacturer's instructions. Images were taken by using an upright fluorescence microscope (Axio Imager M2) and the cells were

counted by using ImageJ software (version 1.8.0.112). EdU-positive cells were stained red and were calculated as (EdU add-in cells/Hoechst stained cells) × 100%.

Cell cycle analysis

For the cell cycle analysis, after transfection and synchronization, the cells were harvested from each well by trypsinization and fixed in 70% (v/v) cold ethanol at 4 °C overnight. After washing with ice-cold PBS, the fixed-cell pellets were collected by centrifugation. Cells were prepared by using Cell Cycle and Apoptosis Analysis Kit (Beyotime) following procedures and resuspended in Propidium/RNase Staining Buffer for staining of DNA and finally analyzed using flow cytometry (Beckman Cytoflex).

Detection of apoptosis

To quantify apoptotic death, after the indicated transfections, the cells were collected, and the apoptotic portion was identified using the Annexin V-FITC/PI Apoptosis Detection Kit (Beyotime) following the manufacturer's instructions. The number of apoptotic cells was then quantified by flow cytometry (Beckman Cytoflex), and the analysis was carried out using FlowJo software (version 10.5.2).

Statistical analysis

Normal distribution was tested using the Kolmogorov- Smirnov test and variance was compared. All statistical analyses in this study were performed by using the Graphpad Prism 8 for macOS (version 8.4.3). Unless otherwise stated, the independent t-test evaluates whether the means for two independent groups are statistically different from each other. Continuous variables were represented as mean ± SEM. Differences were considered statistically significant when the P value was <0.05.

RESULTS

A novel heterozygous variant is shared by two families with suspected SCZ

Two families with SCZ-diagnosed siblings were selected from the mental disease biobank of Bio-X institutes. Each family had three SCZ patients and one or two normal individuals (Fig. 1A). We performed a systematic family-based study using whole-exome sequencing in these two Han Chinese families. Given the pattern of dominant inheritance in the family, only rare heterozygous variants (allele frequency <1%) that were shared by the patients but were not found in the normal controls in the same family were considered candidate mutations. Finally, we found that the *ARID4A* c.1231 C > G (NM_002892.4) variant was co-segregated with SCZ in both families. A nucleotide change (c.1231 C > G) resulted in an amino acid change of histidine to aspartic acid at position p.His411Asp. The p.His411Asp missense mutation was predicted to be a damaging or a disease-causing mutation by MutationTaster and PolyPhen-2 software. In addition, the c.1231 C > G heterozygous variant was validated in the SCZ siblings by Sanger sequencing (Fig. 1B). The gene-based burden test, in which the aggregate burden of rare, protein-altering variants in the *ARID4A* gene was compared among 5418 cases and 68724 controls from the UK Biobank, showed that the *ARID4A* gene putative loss-of-function variants were associated with bipolar and major depression conditions (P *value* SKATO = 0.0164, P *value* burden = 0.0159, P *value* SKAT = 0.0178) (Fig. 1C). According to the Genome-Wide Association Studies (GWAS) that used data from the UK biobank, a total of 36 missense mutations in the *ARID4A* gene were found to be associated with mental disorders ($P < 0.05$) (Fig. 1D). Two of 36 loci (p. Pro467Leu and Arg1017Gln) were associated with both bipolar and major depression conditions and SCZ, and p.Ser1059Thr was associated with both depression and SCZ.

The overall structure of ARID4A PWWP-ARID tandem

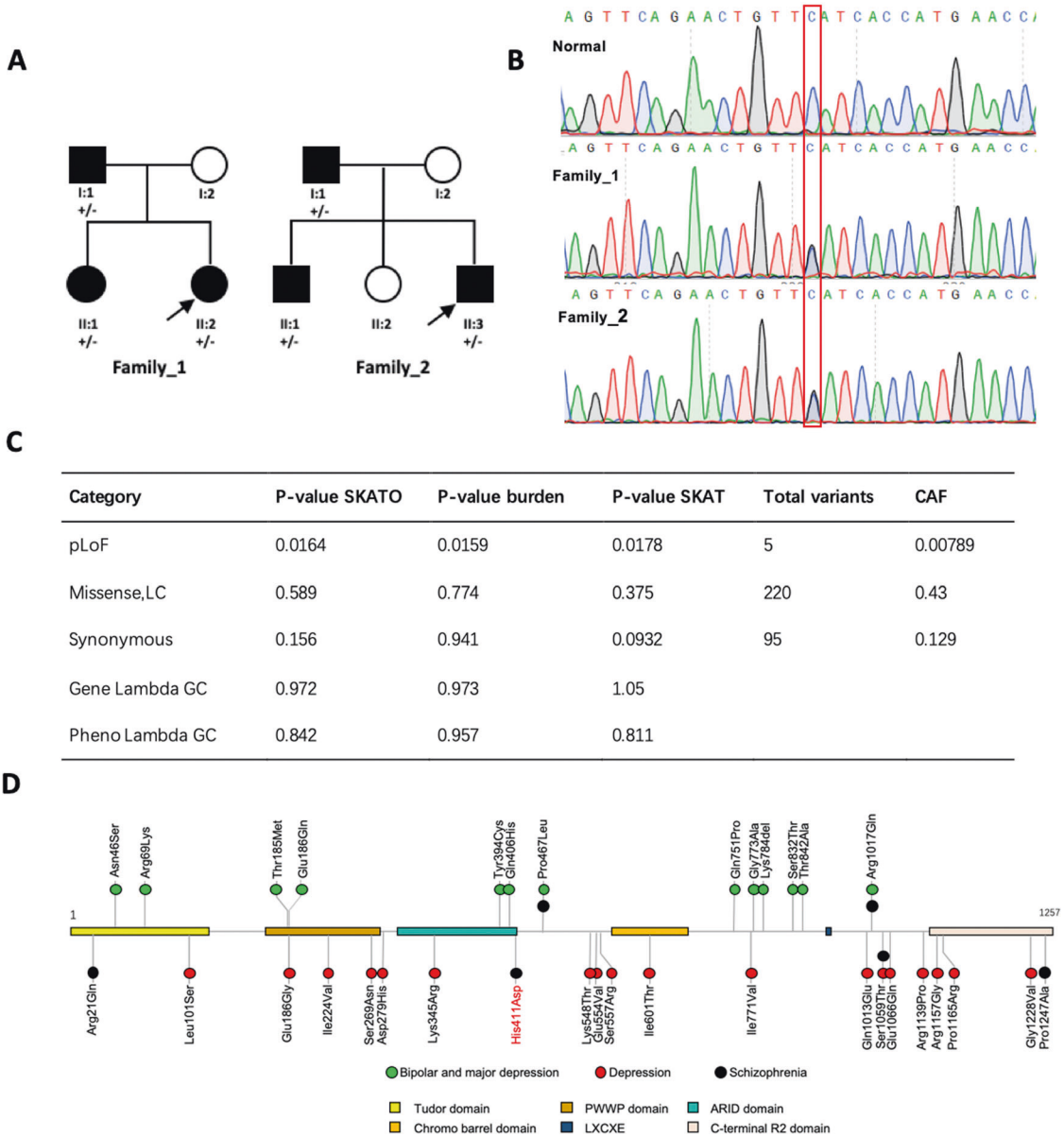
ARID4A, from the N-terminus to the C-terminus, contains the following structural domains: a Tudor domain, PWWP domain, ARID domain, chromo barrel domain, LXCXE motif, and R2 domain (Fig. 2A). Based on sequence analysis, the SCZ-associated mutation p.His411Asp in ARID4A is located at the C-terminal region in the ARID domain. To better understand the functional role of the ARID domain and rationalize the structural effects of the mutation in psychiatric disorders, we set out to characterize its three-dimensional structure. First, we purified the protein containing the predicted ARID domain (residues 307-420). However, the protein sample tended to form oligomers in gel filtration, probably due to nonspecific sample aggregation (data not shown). Despite extensive efforts in condition screening, we could not obtain crystals with the ARID domain. We next purified an extended ARID4 construct that includes the predicted PWWP and ARID domains. The PWWP-ARID protein was of good quality, and fortunately, crystals that diffracted to 2.05 Å were obtained. The structure of the tandem PWWP-ARID was solved through molecular replacement by using the ARID domain of ARID3A (PDB: 4LJX) as the search model. In the structure, the PWWP and ARID domains are located side-by-side and form a compact dual-domain cassette (Fig. 2B). The PWWP domain (a.a. 169-272) adopts a canonical fold that contains a β-barrel of five antiparallel β-strands (βP1-βP5, "P" denotes elements of PWWP), with an α-helix at the very N-termini (αP1), a short α-helix between βP2 and βP3, and three α-helices following the β-barrel (αP3-αP5) (Fig. 2B). The linker region (a.a. 273-302) between PWWP and ARID is missing in the structure, possibly due to the disordered conformation. The ARID domain (a.a. 303-415) conformation is similar to the core structure of the other solved ARID domains. In general, the domain comprises six helices (αA1–αA6; "A" denotes elements of ARID), with three of the helices (αA2, αA3, and αA4) forming a U shape and two of the helices (αA1 and αA6) packing perpendicularly (Fig. 2B).

The interface between PWWP and ARID

The interaction between the PWWP and ARID domains is mediated by hydrophobic and polar interactions. In the binding interface of PWWP-ARID, the majority of contacts are mediated by αA1 and αA6 of ARID and αP2, αP4, αP5, βP2, and βP3 in the PWWP domain (Fig. 2C–E). PWWP and ARID share a contact surface area of 964 Å, which can be subdivided into three binding sites/regions. In detail, the hydrophobic face of PWWP involving Leu194, Ile196, Phe221, Phe254, Val260, and Trp264 packs tightly against Phe322 at α1A and Tyr399 and Ile405 at the α6A of ARID (Fig. 2C). In addition, Arg214 forms charge–charge interactions and hydrogen bonds with Asp325 and Tyr399 (Fig. 2C). Furthermore, Met266, Ile268, and Ile271 at αP5 form hydrophobic interactions with Leu316 and Tyr320 at αA1 (Fig. 2D). The hydrogen bonds formed by Ser199 and Glu398 and by the main chain of Cys200 and Arg236 further reinforce the formation of the PWWP-ARID cassette (Fig. 2E). Consistent with these structural analyses, the protein from the isolated ARID domain is not very soluble or aggregated, indicating that PWWP is important for the stability of the ARID domain.

Biochemical characterization of disease mutation

Based on the alignment of the C-terminal in the ARID domain among vertebrates, we found that the residue His411 (shown in the red frame) in ARID4A is highly conserved among these species but exhibits a much lower identity in other ARID-family proteins (Fig. 3A). This implied that the mutation site is unique in ARID4A and possibly plays important functional roles in disorders by weakening the binding affinities between ARID4A and the target DNA. We performed a series of detailed biochemical and biophysical analyses of the interactions between dsDNA and the PWWP-ARID of ARID4A in humans. Fast protein liquid chromatography (FPLC) coupled with light-scattering studies (Fig. 3B) showed that the molecular mass in the wild type (WT) and disease mutation (p.His411Asp, H411D) groups were both the same (approaching 30 kDa), indicating that the WT and p.His411Asp of *ARID4A* were stable monomers in liquid



Category	P-value SKATO	P-value burden	P-value SKAT	Total variants	CAF
pLoF	0.0164	0.0159	0.0178	5	0.00789
Missense,LC	0.589	0.774	0.375	220	0.43
Synonymous	0.156	0.941	0.0932	95	0.129
Gene Lambda GC	0.972	0.973	1.05		
Pheno Lambda GC	0.842	0.957	0.811		

Fig. 1 Genetic and burden analysis of *ARID4A* in mental disorders. A Novel heterozygous variant shared by two families with SCZ-diagnosed siblings. The black arrow indicates the proposita in family_1 and propositus in family_2. **B** Normal sequence and c.1231 C > G (p.His411Asp) heterozygous variant in SCZ siblings as determined by Sanger sequencing. **C** *ARID4A* gene burden associations with bipolar and major depression according to the previous GWAS. CAF cumulative allele frequency, pLoF putative loss-of-function. **D** *ARID4A* single variants are associated with mental disorders. The variant found in this study is marked in red.

solutions. Unexpectedly, compared to that of the WT, the position of the potential difference peak in p.His411Asp is ahead. Together, these biochemical results reveal that the conformation of p. His411Asp is less compact than that of the WT. Furthermore, our isothermal titration calorimetry (ITC)-based assay (Fig. 3C, D) also confirmed that WT has a strong affinity (Kd~0.14 μM) for dsDNA, but p.His411Asp has no binding affinity for dsDNA. We therefore conjecture that residue His411 is a crucial position for the binding of ARID4A to dsDNA and that the disease mutation can disrupt the interaction between the protein and DNA.

Structural comparison of ARIDs of ARID4A and Mrf2

A ribbon diagram of the ARID4A PWWP-ARID (Fig. 3E) is shown along with the Mrf-2 ARID (PDB: 1IG6, Fig. 3G) and Mrf-2 ARID-DNA complex (PDB: 2OEH, Fig. 3H) structures. The core region (from helices α1 to α6) of the ARID4A ARID is similar to those of apo and DNA-complexed Mrf-2 and can be well superimposed with the rmsd on Cα of 2.7 Å and 2.6 Å, respectively. In contrast to the similarity of the central region, the ARIDs of ARID4A, apo Mrf-2, and DNA-complexed Mrf-2 have significant conformational differences at the C termini (Fig. 3I). The orientation of the C-terminus of ARID4A is similar to that of apo Mrf2, but in DNA-complexed Mrf2, the C-terminus is positioned toward DNA and contacts the DNA backbone. The structures of Mrf2, Dri, p270, and ARID5 ARIDs have been previously solved in complex with DNA. All the solved structures indicate that the core region of ARID binds DNA via similar modes. Two loop regions of ARID, namely, L1 and L2, mainly contribute to DNA recognition. Loop L1 (between α1 and α2) interacts with the phosphate backbone or the minor groove of DNA, and loops L2 (between α4 and α5) and α5 insert into and contact the major groove of DNA. In the Mrf2-DNA complex structure, the positively charged residues at the

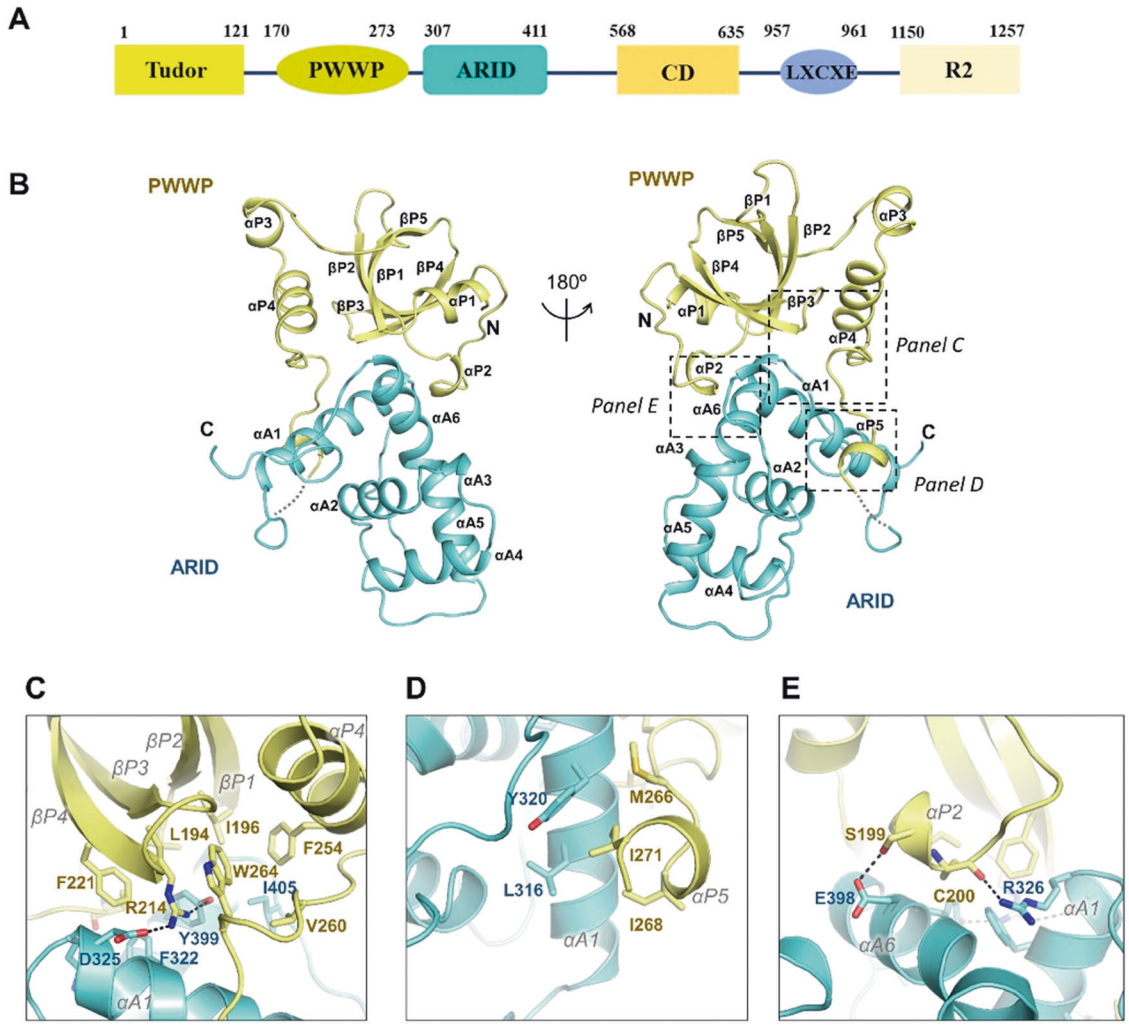


Fig. 2 The overall structure of the ARID4A PWWP-ARID domain. A Schematic diagrams showing the domain organization of human ARID4A. **B** Ribbon diagram representation of the ARID4A PWWP-ARID structure as viewed from the front (left) and the back (right). Secondary structure elements are labeled. PWWP and ARID are yellow and cyan, respectively. **C-E** Detailed interface between the PWWP and ARID domains. The residues involved in the domain-domain interactions are shown with the stick mode. Hydrogen bonds are shown as black dashed lines.

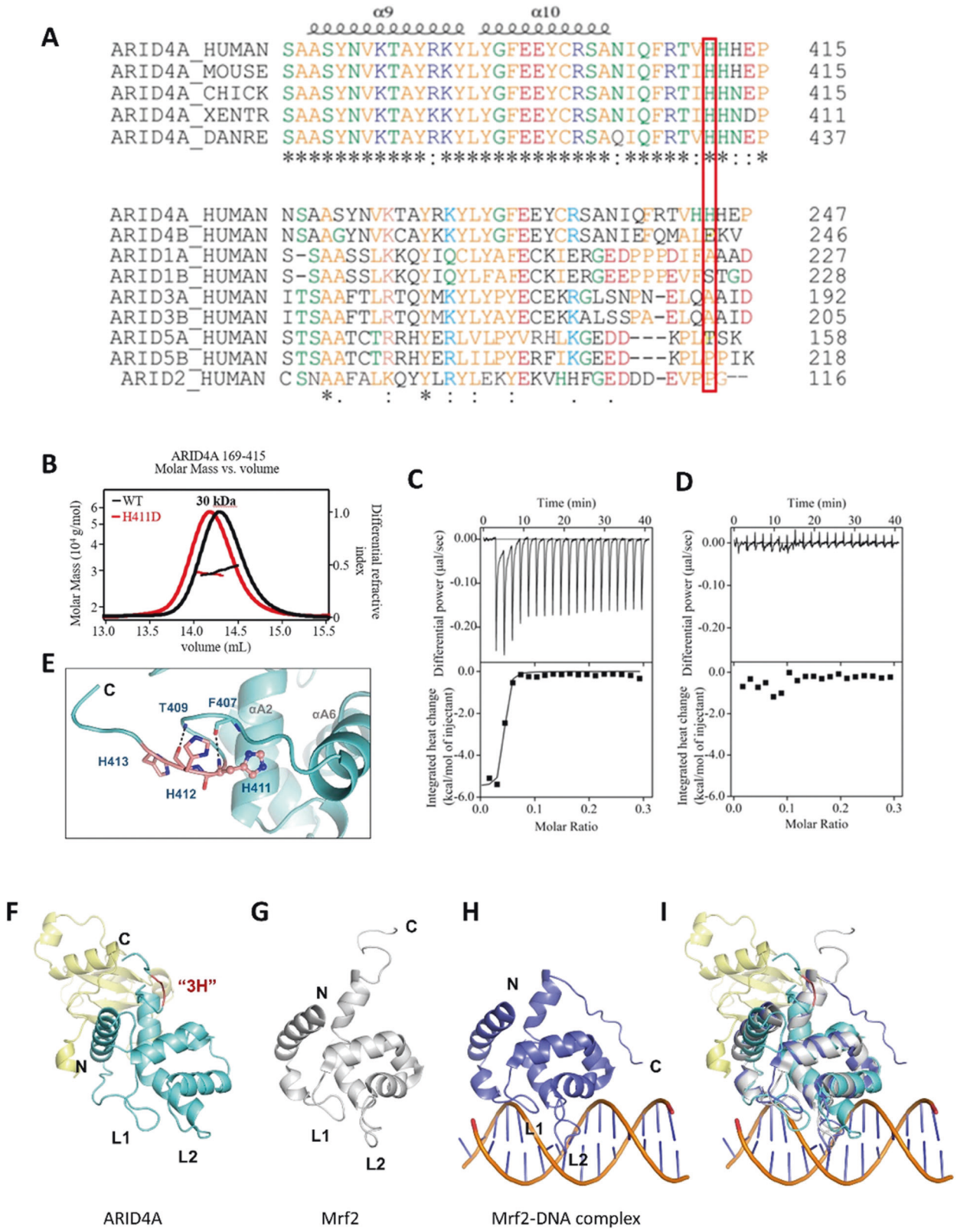
flexible C-terminus of the ARID domain also make contact with DNA. Deleting the C-terminus abolishes its DNA-binding ability, indicating that the C-terminus is also required for DNA binding. Our biochemical data show that His411 at the C-terminus of ARID, which is equivalent to the positively charged C-terminal DNA-binding site of Mrf2, plays an important role in DNA binding. However, in our structure, the orientation of His411 is unfavorable for making contact with DNA, possibly due to crystal packing. The conformation of the C-terminus in the crystal structure is stabilized by the hydrogen bonds formed by the main chains of His411-Phe407 and His412-Thr409 (Fig. 3E). Structural rearrangement of the C-terminus of ARID may occur during DNA binding. Further structural studies of the ARID4A PWWP-ARID in complex with DNA will be helpful for understanding its mechanism of DNA recognition. Table 1 shows the data collection and refinement statistics related to Fig. 3.

p.His411Asp promotes proliferation and cell cycle progression

To investigate the effect of the ARID4A p. His411Asp variant on cell proliferation, HEK293T cells were transfected with p3xFlag-CMV-14-ARID4A p. His411Asp plasmid or p3xFlag-CMV-ARID4A p. His411 plasmid, respectively, and were then examined by an EdU assay. As a result, the EdU incorporation rate (percentage of cells that underwent cell division) was elevated in the cells transfected with p. His411Asp plasmid compared with the wild type (Wt) ($P = 0.001$, Wt: 35.29 ± 1.25, p.His411Asp: 43.78 ± 1.509) (Fig. 4A, B). Moreover, p. His411Asp facilitated cell cycle progression, in which fewer cells arrested in the G0-G1 phase ($P = 0.029$, Wt: 37.36 ± 1.500, p.His411Asp: 31.37 ± 1.913) and more cells progressed to the G2-M phase ($P = 0.006$, Wt: 12.960 ± 1.370, p.His411Asp: 18.580 ± 0.814) (Fig. 4C, D). Collectively, our findings indicate that the *ARID4A* p. His411Asp promotes the proliferation of HEK293T cells, which probably results from the progression of the cell cycle. To determine whether the mutation could influence cell apoptosis, we assessed the apoptosis of HEK293T cells with transfections using fluorescence-activated cell sorting (FACS). We found that the apoptotic rate was not significantly changed ($P = 0.679$, Wt: 5.247 ± 0.489, p.His411Asp: 5.480 ± 0.501) (Fig. 4E, F).

DISCUSSION

SCZ is a severe psychiatric disorder that profoundly affects cognitive, behavioral, and emotional processes, yet its etiology and pathophysiology are still largely unknown [29–31]. SCZ is considered a neurodevelopmental disease based on epidemiologic, histologic, and genetic evidence. The cellular and molecular basis for altered brain development remains poorly understood [32]. In this study, we found that the *ARID4A* c.1231 C > G (NM_002892.4, p.His411Asp)



mutation is associated with SCZ. The *ARID4A* p. His411Asp mutation is located in the ARID domain, which was reported to be responsible for DNA-binding activity. We performed a structural and functional study to investigate the roles of the *ARID4A* p. His411Asp variant.

In mammals, sequence relationships reveal seven distinct subfamilies of ARID-containing proteins in metazoans, namely, ARID1, ARID2, ARID3, ARID4, ARID5, JARID1, and JARID2 [17, 33]. The precise function of all human ARID proteins is still unclear.

Fig. 3 Biochemical characterization of disease mutation. A Structure-based sequence alignment of the C-terminus in the ARID domain. At the top, the sequence alignment from various species is shown as follows: HUMAN, Homo sapiens; MOUSE, Mus musculus; CHICK, Gallus gallus; XENTR, Xenopus (Silurana) tropicalis; DANRE, Danio rerio. At the bottom, the sequence alignment of C-terminal AD with other ARID-family proteins in Homo sapiens is shown. The structure-based alignments were made using secondary structure matching. The residue of disease mutation is shown in a red frame. **B** FPLC coupled with static light-scattering analysis of the ARID4A PWWP-ARID domain, showing the state in the solution of WT (black) and disease mutation (red). ITC titration curves showing the binding affinities of the WT PD-AD (**C**) and the disease mutation (**D**) with dsDNA. **E** Ribbon diagram of the C termini of ARID4A PWWP-ARID. H411 is shown in ball-stick mode. Ribbon diagram representations of the ARID4A PWWP-ARID (**F**), apo Mrf2 ARID (**G**), and DNA-complexed Mrf2 ARID (**H**) from the same view. **I** Structural superimposition of ARID4A, apo, and DNA-complexed Mrf2.

Table 1. Data collection and refinement statistics.

Data collection and processing	
Crystal	ARID4A PWWP-ARID
Source	SSRF-BL19U1
Wavelength(Å)	0.97853
Space group	P212121
Unit cell (^{a,b,c}, Å)	54.668, 54.818, 180.755
Unit cell (α,β,γ,°)	90, 90, 90
Resolution range (Å)	50–2.05 (2.13–2.05)
No. of unique reflections	34353 (2942)
Redundancy	12.5 (9.1)
I/σ(I)	20.83 (2.48)
Completeness (%)	98.40 (86.12)
Rmerge (%)[a]	10.79 (90.73)
CC1/2	99.9 (81.5)
Wilson B	33.06
Structure refinement	
Resolution (Å)	46.87–2.05
Rworkb/Rfreec (%)	19.58/23.14
rmsd bonds (Å)/angles (°)	0.010/1.040
Number of reflections	
Working set	34,345
Test set	1994
Number of protein atoms	3420
Number of solvent atoms	169
Average B factor (Å2) (Protein/solvent)	34.85/35.72
Ramachandran plot(%)	
Most favored regions	98.1
Additionally allowed	1.9
Outlier	0

Numbers in parentheses represent the value for the highest resolution shell.
[a]Rmerge = Σ|Ii - Im|/ΣIi, where Ii is the intensity of the measured reflection and Im is the mean intensity of all symmetry-related reflections.
[b]Rwork = Σ||Fobs|- |Fcalc||/Σ|Fobs|, where Fobs and Fcalc are observed and calculated structure factors.
[c]Rfree = ΣT||Fobs|- |Fcalc||/ΣT|Fobs|, where T is a test data set of about 5% of the total reflections randomly chosen and set aside prior to refinement.

Most ARID DNA binding domains show transcriptional activation and/or repression functions [34–36]. *ARID1B* is important in neural development, as studies have reported that *ARID1B* haploinsufficiency disrupts the development of cortical interneurons and is related to neurodevelopmental and psychiatric disorders [37–39]. According to GWAS using data from the UK biobank, a total of 36 missense mutations in the *ARID4A* gene showed that the allele frequencies were different in the control and mental disorder cases. Thirteen (36.1%) loci were associated with bipolar and major depression conditions (5418 cases and 68724 controls), 18 (50%) loci with depression (16242 cases and 265301 controls), and 5 (13.9%) loci with SCZ (560 cases and 281025 controls). It is suggested that the *ARID4A* gene plays a significant role in the occurrence and development of psychiatric diseases. SCZ has long been thought to be clinically heterogeneous. A range of studies suggests that this is due to genetic heterogeneity [40, 41]. However, sample sizes directly influence the findings of research, and the small sample size of cases (560 cases) in the GWAS of SCZ is a nonnegligible reason for finding so few risk variants. Although over 600 thousand subjects were recruited in three GWASs, only a few *ARID4A* mutations were found to be weakly associated with mental disorders. Considering that GWAS is usually used to find common variants, family-based WES is more suitable and efficient than GWAS to find rare variants in patients with SCZ. Our findings provide new insights for association studies between *ARID4A* and SCZ.

The structural conformation of the PWWP domain of human RBBP1 (ARID4A) was solved by RIKEN Structural Genomics/Proteomics Initiative (RSGI) through NMR measurement (PDB code 2YRV) [12], while we determined the crystal structure of the tandem PWWP-ARID domain of ARID4A at 2.05-Å resolution. In the structure, the PWWP and ARID domains are located side-by-side and form a compact dual-domain cassette; these structural data are crucial for explaining the structural-based mechanism of its physiological functions. Moreover, we performed a series of detailed biochemical and biophysical analyses of the interactions between dsDNA and the PD-AD of ARID4A in humans. The results indicated that the WT and the His411Asp of ARID4A are stable monomers in liquid, but the structure of p.His411Asp is looser than that of the WT, possibly due to the positively charged histidine residue being replaced by a negatively charged aspartic acid residue. The changed amino charge then influences the local spatial conformation of the ARID domain, leading to a looser structure. Furthermore, our ITC-based assay experiment also confirmed that the WT has a strong affinity for dsDNA, but p.His411Asp has no binding affinity for dsDNA. The above results indicate that the residue His411 might be a crucial position for ARID4A binding to dsDNA; this variant might lead to conformational changes in the ARID4A ARID domain DNA binding site, which then disrupts the interaction between protein and DNA.

Previous studies have shown that the PWWP domain may potentially recognize methylated lysines of histones [42], and the ARID domain can bind to DNA in a nonspecific manner and exert the repressor activity of mSin3-HDAC to inhibit the transcription of E2F-regulated genes, thus inducing cell cycle stagnation [17]. ARID4A associates with the pocket region of pRb and can repress transcription from E2F-dependent promoters. Two repression domains (R1 and R2) have been mapped within ARID4A. R2 is a COOH-terminal and can associate with the mSin3-HDAC complex [8, 19]. ARID4A recruits this complex to the pRb pocket and can repress E2F-mediated transcription in an HDAC-dependent manner. R1 maps to a region that includes the ARID domain and represses transcription in an HDAC-independent manner [19, 43]. Cell proliferation can be inhibited by the RB/E2F complex,

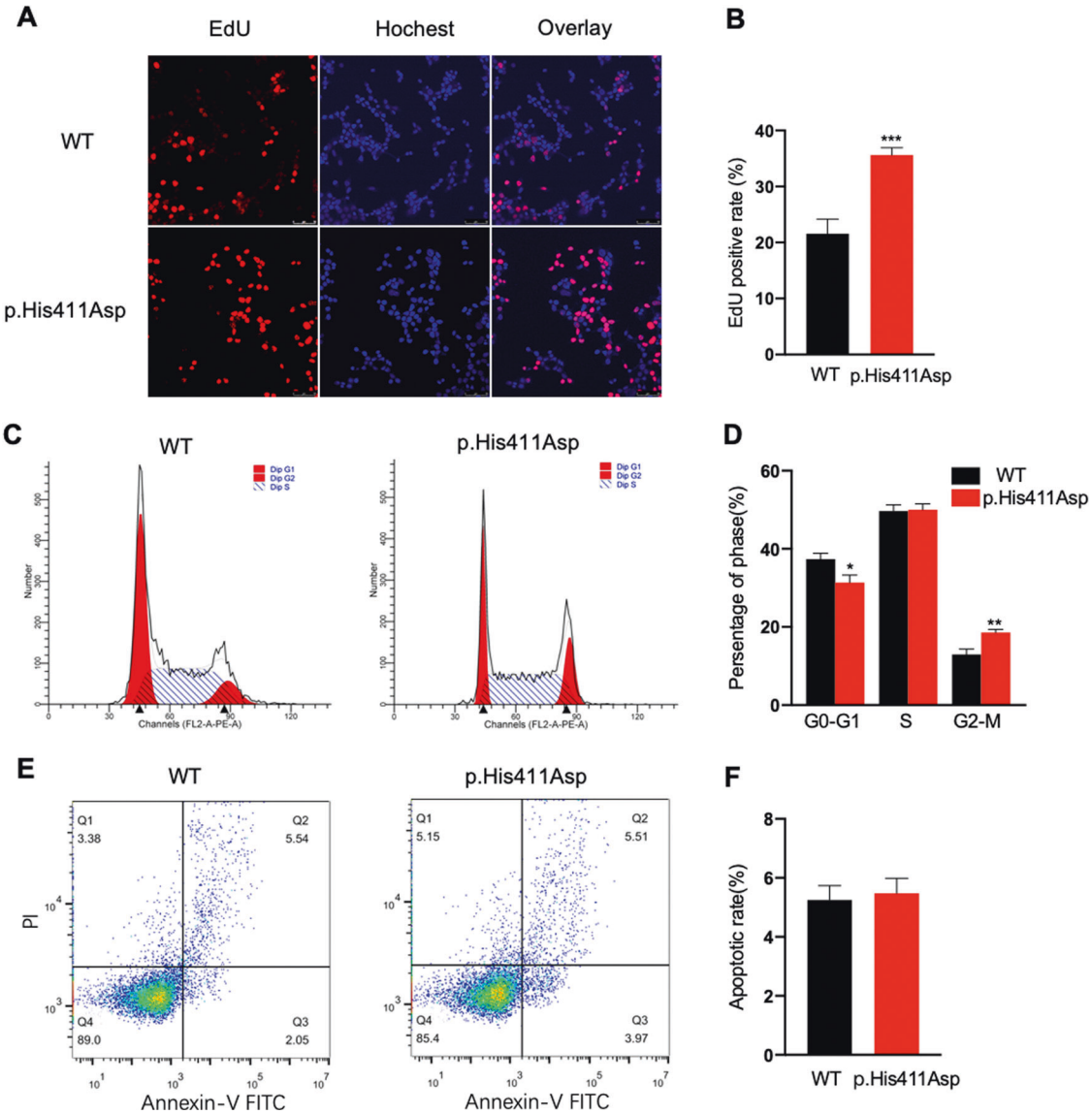


Fig. 4 ***ARID4A* p. His411Asp promotes the proliferation and cell cycle progression of HEK293T cells. A, B** Detection of EdU add-in cells (red) 48 h after transfection. Nuclei were counterstained with Hoechst (blue). **C, D** Cell cycle distributions of HEK293T cells were analyzed 48 h after different transfections as indicated. **E, F** FACS analysis of apoptosis. The cells were collected, stained with Annexin V and PI and analyzed by FACS. The data are expressed as the means ± SEM. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

which arrests cells in the G1 phase [44]. The function of repressing transcription and stagnating the cell cycle relied on the combination between ARID and dsDNA. The broken interaction between protein and DNA may lead to weakened repressor activity and increased cell proliferation. Then, we performed experiments to explore the function of the ARID4A p.His411Asp in vitro. The p.His411Asp-transfected HEK293T cells exhibited abnormal cell cycles and cell proliferation. Although the molecular and cellular mechanisms behind the functional and structural changes in SCZ remain unclear, scientists believe that changes in cell proliferation and neurogenesis play critical roles in SCZ. Katherine M Allen et al. found a decrease in cell proliferation in the anterior hippocampus of people with SCZ, confirming a deficit in hippocampal cell proliferation and neurogenesis in SCZ [45]. Reif et al. suggested that reduced neural stem cell proliferation may contribute to the pathogenesis of SCZ [46]. Edit Hathy et al. observed increased cellular proliferation in neuronal progenitor cells derived from patients with SCZ [47]. Yongjun Fan et al. reported that the cell population proliferation rate was increased in SCZ due to the larger pool of proliferating progenitors and reduced cell cycle period [32]. SCZ is increasingly believed to be a disorder involving abnormal neurodevelopment. Several recent research findings indicate that SCZ may begin with abnormal neurogenesis from embryonic neural stem cells (NSCs) and that this process may be particularly vulnerable to a number of genetic and/or environmental disturbances during early brain development [48]. These studies implied that abnormal cell proliferation is closely related to the occurrence of SCZ. The high level of *ARID4A* expression in the cerebral cortex and neuronal cells suggested that it is important for the regulation of cell proliferation and the cell cycle during different stages of neural development. The abnormal cell proliferation caused by the ARID4A p.His411Asp might play an important role in early brain development and the occurrence and development of SCZ.

In summary, we found that *ARID4A* c.1231 C > G (p.His411Asp) was associated with SCZ by a family-based WES. The ARID4A p. His411Asp could cause conformational changes in the ARID4A ARID domain and then lead to the loss of normal function. Our

study provided evidence that the ARID4A p.His411Asp plays an important role in suppressing the cell cycle and proliferation. Considering the small number of families recruited in this study, targeted sequencing and more experimental validations are needed in future studies to confirm the association between the *ARID4A* gene and SCZ.

REFERENCES

1. Mueser KT, McGurk SR. Schizophrenia. Lancet. 2004;363:2063–72.
2. Patel KR, Cherian J, Gohil K, Atkinson D. Schizophrenia: overview and treatment options. P t. 2014;39:638–45.
3. Degenhardt F, Priebe L, Meier S, Lennertz L, Streit F, Witt SH, et al. Duplications in RB1CC1 are associated with schizophrenia; identification in large European sample sets. Transl Psychiatry. 2013;3:e326.
4. Wu MY, Tsai TF, Beaudet AL. Deficiency of Rbbp1/Arid4a and Rbbp1l1/Arid4b alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain. Genes Dev. 2006;20:2859–70.
5. Wu RC, Jiang M, Beaudet AL, Wu MY. ARID4A and ARID4B regulate male fertility, a functional link to the AR and RB pathways. Proc Natl Acad Sci USA. 2013;110:4616–21.
6. Wu RC, Zeng Y, Pan IW, Wu MY. Androgen Receptor Coactivator ARID4B Is Required for the Function of Sertoli Cells in Spermatogenesis. Mol Endocrinol. 2015;29:1334–46.
7. Wu MY, Eldin KW, Beaudet AL. Identification of chromatin remodeling genes Arid4a and Arid4b as leukemia suppressor genes. J Natl Cancer Inst. 2008;100:1247–59.
8. Lai A, Kennedy BK, Barbie DA, Bertos NR, Yang XJ, Theberge MC, et al. RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest. Mol Cell Biol. 2001;21:2918–32.
9. Liang YK, Han ZD, Lu JM, Liu ZZ, Zhuo YJ, Zhu XJ, et al. Downregulation of ARID4A and ARID4B promote tumor progression and directly regulated by microRNA-30d in patient with prostate cancer. J Cell Biochem. 2018;119:7245–55.
10. Zhang J, Hou S, You Z, Li G, Xu S, Li X, et al. Expression and prognostic values of ARID family members in breast cancer. Aging (Albany NY). 2021;13:5621–37.
11. Hurst DR, Xie Y, Vaidya KS, Mehta A, Moore BP, Accavitti-Loper MA, et al. Alterations of BRMS1-ARID4A interaction modify gene expression but still suppress metastasis in human breast cancer cells. J Biol Chem. 2008;283:7438–44.
12. Gong W, Yao X, Liang Q, Tong Y, Perrett S, Feng Y. Resonance assignments for the tandem PWWP-ARID domains of human RBBP1. Biomol NMR Assign. 2019;13:177–81.
13. Coulson RL, Powell WT, Yasui DH, Dileep G, Resnick J, LaSalle JM. Prader-Willi locus Snord116 RNA processing requires an active endogenous allele and neuron-specific splicing by Rbfox3/NeuN. Hum Mol Genet. 2018;27:4051–60.
14. Cataldi M, Arnaldi D, Tucci V, De Carli F, Patti G, Napoli F, et al. Sleep disorders in Prader-Willi syndrome, evidence from animal models and humans. Sleep Med Rev. 2021;57:101432.
15. Systematic single-variant and gene-based association testing of 3,700 phenotypes in 281,850 UK Biobank exomes. https://genebass.org/, 2021, Accessed 2021.
16. Rusk N. The UK Biobank. Nat Methods. 2018;15:1001.
17. Patsialou A, Wilsker D, Moran E. DNA-binding properties of ARID family proteins. Nucleic Acids Res. 2005;33:66–80.
18. Cajuso T, Hänninen UA, Kondelin J, Gylfe AE, Tanskanen T, Katainen R, et al. Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer. Int J Cancer. 2014;135:611–23.
19. Wilsker D, Patsialou A, Dallas PB, Moran E. ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. Cell Growth Differ. 2002;13:95–106.
20. Wang X, Nagl NG, Wilsker D, Van Scoy M, Pacchione S, Yaciuk P, et al. Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. Biochem J. 2004;383:319–25.
21. Maulik A, Giri M, Singh M. Molecular determinants of complex formation between DNA and the AT-rich interaction domain of BAF250a. FEBS Lett. 2019;593:2716–29.
22. Wilsker D, Probst L, Wain HM, Maltais L, Tucker PW, Moran E. Nomenclature of the ARID family of DNA-binding proteins. Genomics. 2005;86:242–51.
23. Lei M, Feng Y, Zhou M, Yang Y, Loppnau P, Li Y, et al. Crystal structure of chromo barrel domain of RBBP1. Biochem Biophys Res Commun. 2018;496:1344–8.
24. Ren J, Yao H, Hu W, Perrett S, Gong W, Feng Y. Structural basis for the DNA-binding activity of human ARID4B Tudor domain. J Biol Chem. 2021;296:100506.
25. Cadet JL, Brannock C, Jayanthi S, Krasnova IN. Transcriptional and epigenetic substrates of methamphetamine addiction and withdrawal: evidence from a long-access self-administration model in the rat. Mol Neurobiol. 2015;51:696–717.
26. Gong W, Wang J, Perrett S, Feng Y. Retinoblastoma-binding protein 1 has an interdigitated double Tudor domain with DNA binding activity. J Biol Chem. 2014;289:4882–95.
27. Wu H, Zeng H, Lam R, Tempel W, Amaya MF, Xu C, et al. Structural and histone binding ability characterizations of human PWWP domains. PLoS One. 2011;6:e18919.
28. Yang C, Wang J, Liu J, Sun Y, Guo Y, Jiang Q, et al. Functional haplotypes of ARID4A affect promoter activity and semen quality of bulls. Anim Reprod Sci. 2018;197:257–67.
29. Li M, Shen L, Chen L, Huai C, Huang H, Wu X, et al. Novel genetic susceptibility loci identified by family based whole exome sequencing in Han Chinese schizophrenia patients. Transl Psychiatry. 2020;10:5.
30. Jouan L, Girard SL, Dobrzeniecka S, Ambalavanan A, Krebs MO, Joober R, et al. Investigation of rare variants in LRP1, KPNA1, ALS2CL and ZNF480 genes in schizophrenia patients reflects genetic heterogeneity of the disease. Behav Brain Funct. 2013;9:9.
31. Anticevic A, Repovs G, Corlett PR, Barch DM. Negative and nonemotional interference with visual working memory in schizophrenia. Biol Psychiatry. 2011;70:1159–68.
32. Fan Y, Abrahamsen G, McGrath JJ, Mackay-Sim A. Altered cell cycle dynamics in schizophrenia. Biol Psychiatry. 2012;71:129–35.
33. Hung PS, Chen CY, Chen WT, Kuo CY, Fang WL, Huang KH, et al. miR-376c promotes carcinogenesis and serves as a plasma marker for gastric carcinoma. PLoS One. 2017;12:e0177346.
34. Chan SW, Hong W. Retinoblastoma-binding protein 2 (Rbp2) potentiates nuclear hormone receptor-mediated transcription. J Biol Chem. 2001;276:28402–12.
35. Kim TG, Kraus JC, Chen J, Lee Y. JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. J Biol Chem. 2003;278:42247–55.
36. Wang X, Nagl NG Jr., Flowers S, Zweitzig D, Dallas PB, Moran E. Expression of p270 (ARID1A), a component of human SWI/SNF complexes, in human tumors. Int J Cancer. 2004;112:636.
37. Jung EM, Moffat JJ, Liu J, Dravid SM, Gurumurthy CB, Kim WY. Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior. Nat Neurosci. 2017;20:1694–707.
38. Moffat JJ, Jung EM, Ka M, Smith AL, Jeon BT, Santen GWE, et al. The role of ARID1B, a BAF chromatin remodeling complex subunit, in neural development and behavior. Prog Neuropsychopharmacol Biol Psychiatry. 2019;89:30–38.
39. Smith AL, Jung EM, Jeon BT, Kim WY. Arid1b haploinsufficiency in parvalbumin- or somatostatin-expressing interneurons leads to distinct ASD-like and ID-like behavior. Sci Rep. 2020;10:7834.
40. Fanous AH, Kendler KS. Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: searching for a framework. Mol Psychiatry. 2005;10:6–13.
41. Lefort-Besnard J, Varoquaux G, Derntl B, Gruber O, Aleman A, Jardri R, et al. Patterns of schizophrenia symptoms: hidden structure in the PANSS questionnaire. Transl Psychiatry. 2018;8:237.
42. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci. 2003;28:69–74.
43. Lai A, Lee JM, Yang WM, DeCaprio JA, Kaelin WG Jr, Seto E, et al. RBP1 recruits both histone deacetylase-dependent and -independent repression activities to retinoblastoma family proteins. Mol Cell Biol. 1999;19:6632–41.
44. Zhang HS, Postigo AA, Dean DC. Active transcriptional repression by the Rb-E2F complex mediates G1 arrest triggered by p16INK4a, TGFbeta, and contact inhibition. Cell. 1999;97:53–61.
45. Allen KM, Fung SJ, Weickert CS. Cell proliferation is reduced in the hippocampus in schizophrenia. Aust N. Z J Psychiatry. 2016;50:473–80.
46. Reif A, Fritzen S, Finger M, Strobel A, Lauer M, Schmitt A, et al. Neural stem cell proliferation is decreased in schizophrenia, but not in depression. Mol Psychiatry. 2006;11:514–22.
47. Hathy E, Szabó E, Varga N, Erdei Z, Tordai C, Czehlár B, et al. Investigation of de novo mutations in a schizophrenia case-parent trio by induced pluripotent stem cell-based in vitro disease modeling: convergence of schizophrenia- and autism-related cellular phenotypes. Stem Cell Res Ther. 2020;11:504.
48. Iannitelli A, Quartini A, Tirassa P, Bersani G. Schizophrenia and neurogenesis: A stem cell approach. Neurosci Biobehav Rev. 2017;80:414–42.

ACKNOWLEDGEMENTS

This project is supported by the National Key Research and Development Program (2016YFC0906400, 2018YF1410600), Innovation Funding in Shanghai (20JC1418600, 18JC1413100), the National Natural Science Foundation of China (82071262, 81671326, 19Z103150073), Natural Science Foundation of Shanghai (20ZR1427200, 20511101900), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), the Shanghai Leading Academic Discipline Project (B205).

AUTHOR CONTRIBUTIONS

GH and QL conceived the concept; DR, GH and QL designed experiments; GH led the project with assistance from DR, LL, LH, YS and QL; ZY, YX, CC, PW, DZ, XW and TY collected samples; DR performed experiments with assistance from XW, FY, YB, ZG, LL, LJ, XY, KH and WL; DR, LL, FY and QL analyzed data; DR, LL, QL and GH wrote the manuscript with input from all coauthors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-022-01530-w.

Correspondence and requests for materials should be addressed to Qing Lu or Guang He.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.