

A *SLC24A2* Gene Variant Uncovered in Pancreatic Ductal Adenocarcinoma by Whole Exome Sequencing

Lei Wang,^{1,*} Zhuo Shao,^{2,*} Shiyue Chen,^{3,*} Lu Shi⁴ and Zhaoshen Li¹

¹Digestive Endoscopy Center, Department of Gastroenterology, Changhai Hospital, Second Military Medical University, Shanghai, China

²Department of General Surgery, Changhai Hospital, Second Military Medical University, Shanghai, China

³Department of Radiology, Changhai Hospital, Second Military Medical University, Shanghai, China

⁴Department of Public Health Sciences, Clemson University, SC, USA

Pancreatic ductal adenocarcinoma (PDAC) presents as an aggressive malignancy caused by environmental and genetic factors. In order to identify causal genes for PDAC, we performed whole exome sequencing (WES) to detect gene mutations in seven pairs of PDAC tissue and adjacent non-tumor tissue samples. Finally, we found a new nonsynonymous single nucleotide variant (nsSNV) in solute carrier 24 family member 2 (*SLC24A2*) gene resulting in the substitution of native glutamic acid (E) into aspartic acid (D) at position of 287 amino acid (E287D) in *SLC24A2* protein, and confirmed this variant by Sanger gene sequencing. *SLC24A2* is a potassium-dependent sodium-calcium exchanger and can transport metal ion across cell membrane. Multiple *in silico* variants' effects analyses methods including SIFT, PolyPhen, PROVEAN, and PANTHER demonstrated this variant had probably damaging effects, which was consistent with the results obtained from Mutation Taster software analysis with a probability of 0.99999997 to be "disease causing." The three dimension (3D) structure analysis results suggested this variant had little effects on the solubility and hydrophobicity of the protein; but it could decrease the protein stability by increasing the total protein structure energy (-8874.33 kJ/mol for the mutant and -8963.54 kJ/mol for the native) and by causing the mutant protein decreasing three stabilizing residues. Less stability of the mutant 287D protein than the native E287 protein was also supported by I-Mutant and Western-blotting analysis results. Overall, a new mutation in *SLC24A2* gene was identified to decrease the stability of *SLC24A2*, which may have potential clinical usages.

Keywords: next-generation sequencing; nonsynonymous single nucleotide variant; pancreatic ductal adenocarcinoma; *SLC24A2*; whole exome sequencing

Tohoku J. Exp. Med., 2017 April, 241 (4), 287-295. © 2017 Tohoku University Medical Press

Introduction

Pancreatic ductal adenocarcinoma (PDAC), the most common type of pancreatic cancer, is the fourth leading cause of cancer-related mortality in America and the sixth for men in China with an estimated 5-year overall survival rate of about 5%-7% (Siegel et al. 2015; Chen et al. 2016). Surgical resection is currently the only curative treatment, however, only 10-15% of patients with PDAC can be identified early enough to undergo surgical resection (Camacho et al. 2005). Various factors can cause the late diagnosis of PDAC, such as rapid local growth and metastasis of tumor cells, the high rate of recurrence, and resistance to chemotherapy or radiotherapy (Maitra and Hruban 2008). Thus, exploring the possible biomarkers for the early diagnosis and therapy of PDAC seems extremely urgent (Fang et al.

2016).

With the advent of human cancer genome map, some variants in certain genes have been identified to play crucial roles in driving carcinogenesis, and thus gain colossal importance as cancer prognostic biomarkers and therapy targets. Furthermore, previous population-based genetic association studies including genome-wide association studies (GWAS), which detect the association of genes or genetic polymorphisms with certain cancers, have obtained the inconsistent results (Cai et al. 2015, 2016; Huang et al. 2016). As carcinogenesis is complex and advancing gradually, it is usually initiated by the somatic mutation in several genes. Therefore, identifying and assessing variants in all genes would be more effective on finding possible cancer prognostic biomarkers and therapy targets.

The whole exome sequencing (WES), a kind of mas-

Received February 20, 2017; revised and accepted March 21, 2017. Published online April 11, 2017; doi: 10.1620/tjem.241.287.

*These authors contributed equally to this work.

Correspondence: Zhaoshen Li, Digestive Endoscopy Center, Department of Gastroenterology, Changhai Hospital, Second Military Medical University, 168 Changhai Road, Shanghai 200433, China.
e-mail: zsl2001@yeah.net

sively parallel next-generation sequencing (NGS) technology, provides new opportunities to identify novel mutations and establish new genetic basis of complex and rare diseases in sporadic cases (Chandler et al. 2016). It involves capture and sequencing of the almost all exons in genomes and thus has become a valuable method of choice for mutation identification. Moreover, several small-scale studies provide some evidence of the potential of WES approach in identifying cancer-driving mutations and genes (Li et al. 2015; Tomsic et al. 2015).

Here, due to the limited clinical genetic markers for PDAC and the rapid progress of NGS, the current study is trying to gain a landscape of the somatic mutations underlying PDAC using the WES and identify the damaging variants in certain genes.

Methods and Materials

Study subjects

Tumor tissues and adjacent non-tumor tissues were collected from 7 PDAC Chinese patients, who underwent the precise laser micro-surgical resection from January 2015 to October 2015. All patients were diagnosis as primary PDAC without other cancers and were followed up every month after surgery. Summary characteristics of patients are described in Table 1. Clinic staging was determined on the basis of Union for International Cancer Control (UICC) Tumor Node Metastasis (TNM) staging system v6.

The study was approved by the human ethics committee of Changhai hospital and informed consent of patients was obtained. All procedures performed in the studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Whole-exome sequencing (WES)

The bisected tissue samples were stored in -80°C freezer. Genomic deoxyribonucleic acid (DNA) was isolated according to the widely-used standard protocols (Cai et al. 2013; Hao et al. 2014; Jiang et al. 2016). The genomic DNA libraries were constructed based on Illumina protocols. About 50 Mb of sequences from the whole exons and their flanking DNA regions were enriched from the fragmented genomic DNA using Agilent SureSelect DNA Capture Array. The captured DNA libraries were then sequenced using the Illumina HiSeq 2500 Genome Analyzer with an overall $>10\text{X}$ coverage depth.

Mutation identification

Sequence data in FASTQ format files were quality-checked using the software of FASTQC v0.11.4 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The filtered sequence reads were mapped to hg19 reference sequence using Burrows-Wheeler Aligner (BWA) v0.7.12 (<http://sourceforge.net/projects/bio-bwa/>) and SAM files were generated. Local realignment around indels, duplicate removal, and base quality score recalibration was performed using the software of Genome Analysis Toolkit (GATK) v3.5 (<https://www.broadinstitute.org/gatk/>). Alignment SAM files were converted to BAM files using Samtools v1.3 (<http://www.htslib.org/>). Somatic

variants were called using the software of MuTect v2.0 (<https://www.broadinstitute.org/gatk/download/auth?package=MuTect>) with default parameters. High confidence variants were those with a minimum of 10 reads covered in the tumor tissue and 8 reads in the adjacent non-tumor tissues, and the maximum alternative allele frequency in the adjacent non-tumor tissues less than 0.3 of that in the tumor tissue. Somatic mutations which are those found in the tumor tissue but not in the adjacent non-tumor tissues DNA samples were annotated using the latest version of annovar main package (<http://annovar.openbioinformatics.org/en/latest/>). Somatic single nucleotide variants (SNVs) were then combined and searched in the dbSNP database and COSMIC database. The new variant identified by WES was verified by Sanger sequencing. Primers used in polymerase chain reaction (PCR) and Sanger sequencing were 5'-TGT CAT CAT GTG GTG GGA-3' as the forward primer and 5'-ACG ACC TTA TTG CGG TTT-3' as the reverse primer. T_m was 51°C for PCR.

In silico analysis

To identify the hitherto unknown mutations, the frequencies of detected somatic mutations were double-checked in available databases, which include ESP database (<http://evs.gs.washington.edu/EVS/>), 1000 Genomes database (<http://www.1000genomes.org/>), and ExAC database (<http://exac.broadinstitute.org/>).

The effects of the SNVs were firstly evaluated using the software of Mutation Taster (<http://www.mutationtaster.org/>) (Schwarz et al. 2010), which gave types of predictions, i.e. "disease_causing_automatic," "disease_causing," "polymorphism," and "polymorphism_automatic," and the probability P-value of the prediction. The P value close to 1 indicates a high possibility. A transcript ID and a sequence covering each SNV were required as input.

SIFT (Sorting Tolerant From Intolerant, <http://sift.jcvi.org/>) algorithm uses a query sequence and multiple alignment information to predict roles of variants (Ng and Henikoff 2003). The alleles with calculated probabilities less than 0.05 are predicted to be deleterious; those greater than or equal to 0.05 are tolerated (Ng and Henikoff 2006). The protein amino acid (AA) sequence and SNVs were used as the input.

PolyPhen-2 (Polymorphism Phenotyping v2.0, <http://genetics.bwh.harvard.edu/pph2/>) analysis tool predicts influence of variants through automatically selecting eight sequence-based and three structure-based predictive features by an iterative greedy algorithm (Adzhubei et al. 2010). If the classifier probability is smaller than 0.15, the allele is predicted to be benign, if the probability is larger than or equal to 0.15, the allele is predicted to be probably or possibly damaging. Protein AA sequence, AA position and variants were used as input.

The software of PROVEAN (Protein Variation Effect Analyzer) is a newly developed and useful tool for predicting biological function impact of an AA substitution through calculating pairwise sequence alignment scores (Choi et al. 2012). If the score for a nsSNV is less than or equal to default score threshold of -4.100 , this SNV is predicted to be deleterious. Protein AA sequence, AA position and variants were used as input.

The software of PANTHER (<http://www.pantherdb.org/tools/csnpscoreForm.jsp>) employs position-specific evolutionary preservation (PSEP) algorithm to calculate the length of time that a given AA has been preserved in the lineage. The longer a position has been preserved, the more likely that it will have a deleterious effect (Thomas et al. 2003). Usually, "probably damaging" (time > 450

million of years) means a false positive rate of ~ 0.2 ; “possibly damaging” (450 million of years $>$ time $>$ 200 million of years) means a false positive rate of ~ 0.4 and “probably benign” has the time less than $<$ 200 million of years. Protein AA sequence, AA position and variants were used as input.

The software of I-Mutant v2.0 (<http://folding.biofold.org/i-mutant/i-mutant2.0.html>) is a neural network based tool for the routine analysis of protein stability through providing the scores for the Gibbs free energy change (ddG), calculated with the FOLD-X (Capriotti et al. 2005). The FOLD-X analysis tool provides the comparison between wild type and mutant models in the form of van der Waals clashes, which greatly influence the energy decomposition. The protein AA sequence was used as the query sequence.

Modeling protein structures

To study the effects of newly identified mutations in detail, the 3D (three-dimensional) structure models for *SLC24A2* was constructed using the I-TASSER (Iterative Threading ASSEmbly Refinement) server (Zhang 2009; Cai et al. 2011). For a given protein, I-TASSER firstly uses multiple threading alignment (MTA) methods to search its structural templates in the PDB database, then constructs its structure models by iterative fragment assembly simulations, and last provides a confidence score (C-score) of the models' global accuracy estimation. The C-score is calculated according to both the significance of threading template alignments and the convergence parameters of the structure assembly simulations. We selected the top 1 structure model with the optimal C-score for the next step analysis.

The residues mutation was constructed by using the software of SWISSPDB viewer and energy minimization for three dimension (3D) structures was processed by the NOMAD-Ref server (http://lorenz.immstr.pasteur.fr/gromacs/minimization_submission.php) (Lindahl et al. 2006). The Gromacs was used in this server as the default force field for energy minimization based on the selected method of conjugate gradient. The deviation between the two structures is evaluated by their root-mean-square deviation (RMSD) values.

Furthermore, we used the server SRide (<http://sride.enzim.hu/>) (Magyar et al. 2005) for identifying the stabilizing residues in native protein and mutant models. Stabilizing residues were computed based on the 3D protein structure using parameters such as: long-range order (LRO), surrounding hydrophobicity, conservation score, and stabilization center.

Vector construction

Based on the lentiviral vector construction method (Cai et al. 2010; Fang et al. 2016), the pGCSIL-hU6.flag-hSLC24A2.CMV.eGFP, pGCSIL-hU6.flag-hSLC24A2mu.CMV.eGFP, and pGCSIL-hU6.mock.CMV.eGFP were constructed. The hSLC24A2 wide-type cDNA and mutant cDNA were respectively cloned by polymerase chain reaction (PCR) from adjacent non-tumor tissue and tumor tissue of PDAC5 patient using the following primers: 5'-TACCTACTGCAACAAAGCACCACC-3' and 5'-CTAGATGGA GACGGGGCATGTAAG-3'. The cloned hSLC24A2 cDNA sequences were confirmed by Sanger sequencing. Recombinant lentiviruses were prepared by plasmids co-transfection, and then pancreatic cancer PANC-1 cells (Shanghai GeneChem Co., Shanghai, China) were infected with the recombinant lentivirus (Cai et al. 2010). The eGFP indicated that the infection efficiency in PANC-1

cells was around 90% and at a multiplicity of infection (MOI) of 30.

Real-time PCR and western blotting

The flag-hSLC24A2 gene expression was determined to assess the exogenous hSLC24A2's stability. Total ribonucleic acid (RNA) was extracted 48 hr after transfection from 2×10^6 transfected PANC-1 cells by using TRIzol Reagents (Invitrogen, Beijing, China). RNA 6000 Nano Kit (Agilent, Santa Clara, USA) was used to evaluate RNA quantity and quality by the Agilent Bioanalyzer. The reverse-transcription (RT) reaction was performed using the Promega RT kit (Promega, Madison, USA) with 1 μ g RNA. The 20 μ l PCR mixture contained 10 μ l of 2 \times Fast SYBR Green MasterMix (Roche Diagnostics, Basel, Switzerland), 1 μ l diluted cDNA, and 10 μ M primer pairs as following: GAPDH, 5'-GACAGTCAGCCGCAT CTTCT-3', and 5'-TTAAAAGCAGCCCTGGTGAC-3'; Flag-hSLC24A2, 5'-CGACTACAAGGATGACGA-3', and 5'-AAG GCACTGATTGAAAATG-3'. PCR amplification was performed using the LightCycler 480 (Roche Diagnostics, Basel, Switzerland) with the following procedures: 94°C for 3 min, 40 cycles at 94°C for 20 s, 50°C for 20 s and 72°C for 20 s. The relative mRNA expression level of flag-hSLC24A2 mutant in PANC-1 cells was calculated based on the expression level of flag-hSLC24A2 wild type according to the method of $2^{-\Delta\Delta Ct}$. Statistical analysis was carried out by student t-test from three independent experiments.

Total proteins were extracted at the meantime of RNA extraction. Totally 20 μ g boiled protein in loading buffer was separated on 10% SDS-polyacrylamide gels, electro-transferred to polyvinylidene fluoride membranes and last probed with 1:3,500 diluted mouse anti-flag and 1:6,000 diluted mouse anti- β -actin (Santa Cruz Biotechnology, Inc., CA, USA) antibodies. A 1:5,000 diluted goat anti-mouse immunoglobulin G (IgG) peroxidase-conjugated secondary antibody (Sigma Aldrich, St. Louis, MO, USA) and the Amersham enhanced chemiluminescence (ECL) plus western blotting detection system (Amersham Pharmacia Biotech, NJ, USA) were used to detect the target protein.

Results

Subject Characteristics

Totally 7 PDAC patients were collected with informed consent to contribute their tumor tissues and adjacent non-tumor tissues. Among them, three subjects were female, the age range for all subjects was from 52 to 61 years with a mean range of 56.9 ± 3.02 (Table 1). Most tumor tissues were located on the head of pancreas and absent of lymph node metastasis.

WES and mutation identification

Exon regions of about 20,000 genes were sequenced from DNA isolated from adjacent non-tumor tissue and primary tumor tissue of each subject. The mean depth of DNA sequencing was 138.34 ± 10.08 for adjacent non-tumor and 142.99 ± 11.04 for tumor, respectively. Details of WES data summary for each patient are presented in Table 2. Coding regions of genomes in the 7 subjects contained 308 somatic variants, of which 61 (19.8%) were predicted to be synonymous, and 173 (56.2%) to be missense (Table 3). Average mutation rate per Mb was estimated to be 2.15 ± 0.62 (Table 3).

Table 1. Clinical characteristics of seven patients with pancreatic ductal adenocarcinoma (PDAC).

Patient ID	Age	Sex	Tumor location	Tumor size	Lymph node metastasis	TNM stage
PDAC1	52	Female	Head	T2	Absent	T2BN0M0
PDAC2	54	Male	Head	T3	Absent	T3aN0M0
PDAC3	58	Female	Head	T3	Absent	T3aN0M0
PDAC4	57	Male	Body/tail	T3	Absent	T3aN0M0
PDAC5	59	Female	Head	T4	Present	T4N2Mx
PDAC6	57	Male	Head	T3	Present	T3bN1M0
PDAC7	61	Male	Body/tail	T3	Absent	T3N0M0

Table 2. Whole exome sequencing data summary for 7 patients.

Items	tumor tissues (7)	adjacent non-tumor tissues (7)
Reads passed filtration	101755135 ± 5458170	99881298 ± 3027867
Reads mapped (%)	99.89 (99.88-99.91)	99.84 (99.77-99.89)
Capture Efficiency (%) ¹	75.26 (70.6-78.2)	74.96 (71-77.8)
Mean Depth*	142.99 ± 11.04	138.34 ± 10.08
nu1Coverage (%) ²	96.8 (96.6-97.0)	96.8 (96.6-97.0)
≥ 10Coverage (%) ²	94.76 (94.6-95.0)	94.74 (94.5-94.9)
≥ 20Coverage (%) ²	92.4 (91.8-93.0)	92.36 (91.8-92.7)
≥ 50Coverage (%) ²	82.08 (79.9-84.8)	82.2 (80.7-83.9)

*Presented as mean ± standard deviation.

¹Capture efficiency (%), ratio of the read number mapped to the exon to the total read number.

²Coverage (%), ratio of the bases number that had been sequenced (≥ 1, 10, 20, 50) times to the total base number of the exome.

Table 3. Summary of somatic mutation types and prevalence in seven patients with pancreatic ductal adenocarcinoma (PDAC).

Samples	SNV in exons			Total	Mutation per Mb DNA	NS/S
	Synonymous	Missense	Others			
PDAC1	6	13	5	24	0.99	2.17
PDAC2	9	21	16	46	2.19	2.33
PDAC3	10	32	15	57	2.51	3.20
PDAC4	10	33	3	46	1.67	3.30
PDAC5	9	23	6	38	2.38	2.56
PDAC6	4	10	9	23	2.72	2.50
PDAC7	13	41	20	74	2.59	3.15
Total	61	173	74	308		

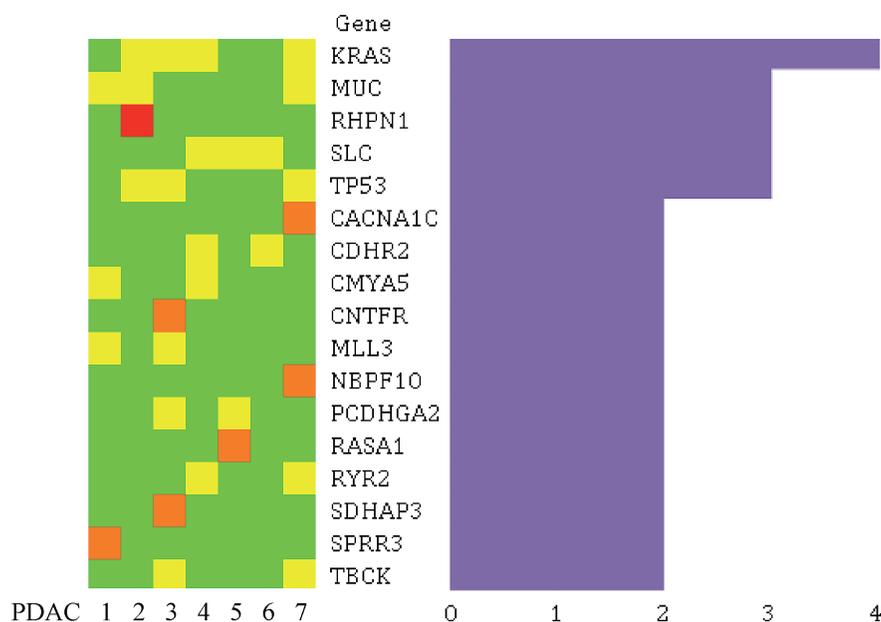


Fig. 1. Main gene mutation landscape in seven patients.

Mutations in a gene with two or more cases (left), colored by the number of the mutation, red means 3, orange means 2, yellow means 1, green means 0; and number of subjects with mutations in a given gene (right). Each column represents a subject, and each row represents a gene.

KRAS genes were mutated in at least four samples and 17 genes in two or more samples (Fig. 1). Among them, solute carrier (SLC) family contains three somatic mutations. SLC family members primarily act as transporters of compounds across biological membranes and have been suggested to play important roles in PDAC (Bhutia et al. 2016). The three somatic variants are following: a newly identified variant (chr9:19786004T>G) in *SLC24A2* gene of PDAC5 patient without any records in the databases of ESP, 1000 Genomes, and ExAC, a variant (chr 5:128364085G>T) in *SLC27A6* of PDAC6 with a frequency of 7.32E-05 in ExAC database, and a variant (chr 10:18250567) in *SLC39A12* of PDAC6 with a frequency of 8.13E-06 in ExAC database. Then, we confirmed the newly identified mutation (chr9:19786004T>G) in *SLC24A2* gene using the Sanger sequencing (Fig. 2). This nonsynonymous SNP (nsSNP) is T>G mutation and results in the replacement of native glutamic acid (E) to aspartic acid (D) at the position of 287.

Evaluation of the variant

We initially used Mutation Taster to evaluate the functional effect of the nsSNV chr9:19786004T>G, which is predicted to be “disease causing” with the probability of 0.99999997 and suggested to potentially affect protein features.

Then, SIFT, which predicts likelihood of damaging substitution at specific position of input sequence with multiple alignment score (Ng and Henikoff 2003), indicated that this nsSNV was damaged (score = 0.01, Table 4). PolyPhen-2, which predicts the impact of an allele replacement using naive Bayes classifier, indicated that this nsSNV

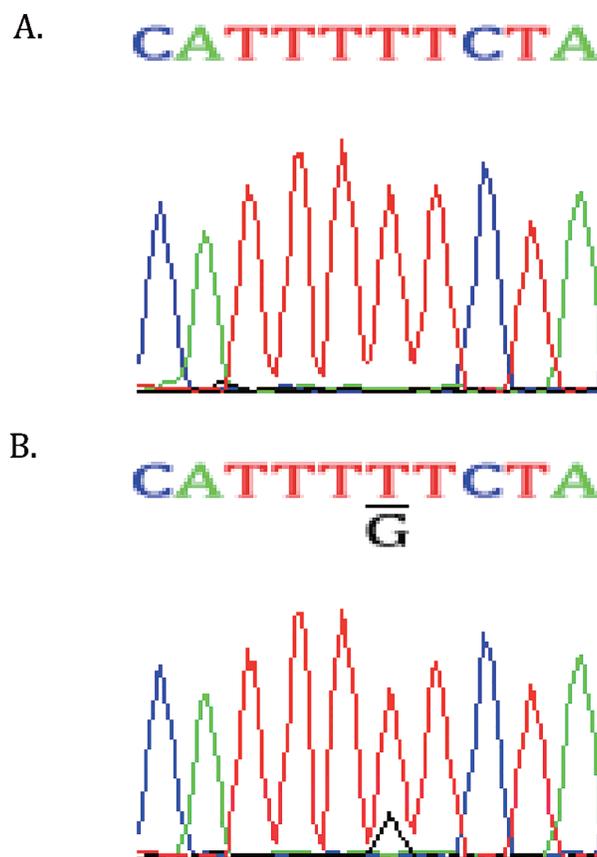


Fig. 2. Sanger sequencing electropherograms.

A newly identified *SLC24A2* variant at position chr9:19786004T>G, causing the Glu-to-Asp substitution (E287D), is not found in adjacent non-tumor tissue (A), but found in tumor tissue (B).

Table 4. Analysis results of four analysis tools, i.e. SIFT, PolyPhen, PROVEAN, PANTHER, and I-Mutant.

SNV ID	AA change	SIFT (score)	PolyPhen (PSIC score)	PROVEAN (score)	PANTHER (time)	I-Mutant (ddG ¹)
chr9:19786004T>G*	E287D	Damaging (0.01)	Probably damaging (0.987)	Deleterious (-2.83)	Probably damaging (1037)	Decreasing stability (-0.31)

*Based on the GRCh37 Ensembl 55.

¹ddG (Kcal/mol) = the unfolding Gibbs free energy value of the mutant – the unfolding Gibbs free energy value of the native.

PSIC, means position-specific independent counts; AA, means Amino acid; SNV, means single nucleotide variant.

was “probably damaging” with the score of 0.987. A newly developed and useful tool of PROVEAN indicated that this nsSNV was deleterious with the score of -2.83. PANTHER, which is based on PSEP algorithm, suggested that this nsSNV was “probably damaging” with a false positive rate of less than 0.2. I-Mutant v2.0, which is a neural network based tool for the analysis of protein stability alterations, indicated that it can decrease protein stability by the ddG of -0.31.

I-Mutant v2.0, which is a neural network based tool for the routine analysis of protein stability and alterations caused by the single-site mutations, indicated that it can decrease protein stability with the ddG of -0.31.

3D Structure analysis

Thereafter, the 3D structure of SLC24A2 was simulated with the C-score of -0.15 through I-TASSER to check the effects of E287D mutation. The residue mutation of SLC24A2 was then superposed by using the SWISSPDB viewer and energy minimization for 3D structures was performed by the NOMAD-Ref server (Fig. 3) (Lindahl et al. 2006). We used the conjugate gradient method for optimizing the 3D structures. Total energy for the native structure was -8,963.54 kJ/mol, and that for the mutant structure (287D) was -8,874.33 kJ/mol. The difference between the wild type structure and the 287D structure is evaluated by their (RMSD) value of 0.51.

The computed total solvent accessible surface area of wild type protein is 24,026 Å² and the hydrophobic surface patches are 16,764 Å² while total solvent accessible surface area of 287D protein is 24,360 Å², and the hydrophobic surface patches are 16,745 Å². The mutant amino acid increases the solvent accessible surface area by 334 Å² for E287D, and reduces solvent excluded surface area by 19 Å² in spite of having more number of atoms as compared to wild type. The reduction in surface area may have little effects on the solubility of the protein and thus its function. The hydrophobicity slightly decreases for the mutant 287D protein.

The SRide server (Magyar et al. 2005) was used to identify the stabilizing residues of the native structure and mutant modeled structures with the conservation score threshold of 6, LRO threshold of 0.01, and surrounding hydrophobicity threshold of 20. Three stabilizing residues were identified in the native structure, which were Arg 38, Ser55, and Ile230. The mutant model (E287D) has no identified stabilizing residues. This analysis revealed that three

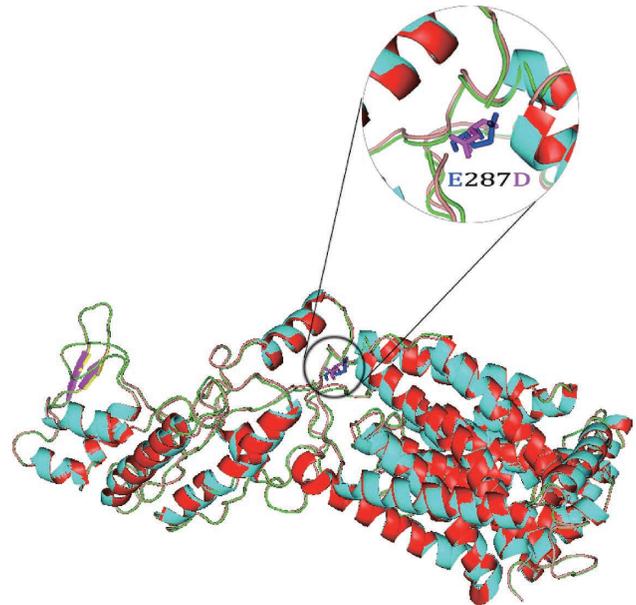


Fig. 3. Superimposed 3D structures of proteins.

Structures of native SLC24A2 (green) and mutant SLC24A2 (salmon) are shown, and the mutated residue 287D is presented in the circle.

stabilizing residues in the E287D-mutant model less than the native protein structure may decrease the stability of E287D-mutant protein.

Validation of mutant's effects

In order to validate the effects of the identified mutation on the stability of hSLC24A2, we transfected recombinant lentiviruses containing the exogenous flag-hSLC24A2 E287 and flag-hSLC24A2 287D gene respectively into PANC-1 human PDAC cells and found that hSLC24A2 287D mutant had the similar mRNA level as hSLC24A2 E287 ($p = 0.23$) (Fig. 4A). While hSLC24A2 287D mutant had less protein expression level than the native protein in PDAC cells (Fig. 4B).

Discussion

PDAC is related to a largely unfavorable outcome and presents as an aggressive and recurrent malignancy. Till now, there are no effective treatments for this disease. Thus, identification of early prognostic factors, which is adequate to distinguish between patients with a high and low risk of disease incidence is crucial. It has been reported that approximately 10% of patients with PDAC have a first

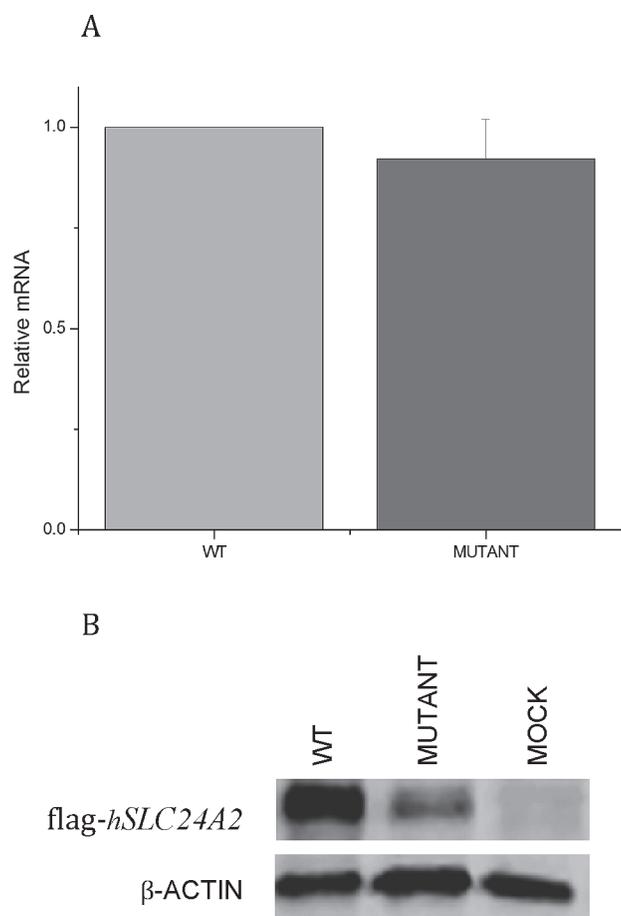


Fig. 4. Validation of protein stability by molecular biology experiments.

Real-time PCR of exogenous flag-*hSLC24A2* E287 wild type (WT) and flag-*hSLC24A2* 287D in PANC-1 cells; $p = 0.23$ (A). Western blotting analysis of exogenous flag-*hSLC24A2* E287 wild type (WT) and flag-*hSLC24A2* 287D gene expression in PANC-1 cells (B).

degree relative with PDAC, supporting that genetic factors have an important contribution to PDAC (Ueno et al. 2015). Moreover, susceptibility genes of PDAC have been described in numerous studies (Chen et al. 2014; Tang et al. 2015). Here, we used WES to explore new gene mutations underlying PDAC, and finally, a *SLC24A2* missense mutation was identified.

WES has been suggested as an effective method to bring a more comprehensive dissection to genomic variation within gene region, and it is being a routine to reveal subtle gene mutations in cancer genomes and provide new potential genetic biomarkers for tumor diagnosis (Chandler et al. 2016). However, there are some technical problems for WES: one is the identified mutations without underlying the mendelian genetic rules; the other is the mutant alleles located in the coding regions that are not well covered by WES; the last is the obscured specific copy to which the variant maps because of the presence of pseudogenes or repetitive regions (Yang et al. 2013). Thus, we only studied gene point variants in the current research and confirmed

them with Sanger sequencing.

SLC24A2 is firstly cloned from rat brain and assigned to 9p22, which belongs to the solute carrier (SLC) family (Lin et al. 2015). SLC family members are primarily involved in the transport of a broad range of compounds across biological membranes and into cells. Most of the SLC transporters have a similar protein structure of 12 crucial transmembrane segments (TMs), which may be critical for substrate binding and correct processing of the transporter protein. To date, 395 SLC transporters that organized into 52 subfamilies have been identified in humans. *SLC24A2* is further categorized into potassium-dependent sodium-calcium exchangers (NCKX) subfamily. NCKX proteins including *SLC24A1*, *SLC24A2*, *SLC24A3*, *SLC24A4*, and *SLC24A5*, are suggested to transport one intracellular Ca^{2+} and one K^{+} in exchange for four extracellular Na^{+} , and to be co-expressed with cyclic guanosine monophosphate (cGMP) gated channels or other ligand-gated channels (Schnetkamp 2004). Although five members of the NCKX gene family have been identified, a clear understanding of NCKX physiology is limited to the role of *SLC24A1* and *SLC24A2* in retinal rod and cone photoreceptors. The human *SLC24A2* (644 residues) has a high degree of sequence identity with *SLC24A1* (1,069 residues) in the hydrophobic regions and shares significant functional similarity as well (Cai et al. 2002). As demonstrated before (Schnetkamp 2013), the basic characteristics of *SLC24A1* and *SLC24A2* are similar: 1. transport stoichiometry is $1\text{Ca}^{2+}+1\text{K}^{+}:4\text{Na}^{+}$. 2. They can mediate both forward and reverse $\text{Na}^{+}/\text{Ca}^{2+}-\text{K}^{+}$ exchange. 3. Na^{+} may not be replaced by other cation. Ca^{2+} may be replaced by Sr^{2+} , K^{+} may be replaced by NH_4^{+} . Nowadays, increasing evidence suggests that SLC family members play important contribution to the carcinogenesis and progression of tumors (Thangaraju et al. 2006; Alshahrani et al. 2012; Bhutia et al. 2016). For example, gene-silencing of *SLC5A8* has been detected in breast, colon, brain, thyroid, gastric, and prostate tumors (Thangaraju et al. 2006). Moreover, *SLC5A8* has been suggested to function as a tumor suppressor gene in the initiation and development of pancreatic cancer (Park et al. 2008).

Here, we reported three variants in *SLC* family genes. Among them, the variant (chr9:19786004T>G) in *SLC24A2* gene was totally new. To evaluate the impact of this variant, we constructed the 3D structures of both native and mutant *SLC24A2* protein. Furthermore, multiple in silico analysis was performed to identify the roles of this nsSNV. SIFT is a so advantageous analysis tool that it can distinguish the damaging SNPs with only about 20% false positive error and 90% true positive prediction (Ng and Henikoff 2003). PolyPhen-2 is a structural modification analysis tool and can achieve true positive prediction rates of 92% and 73% on training dataset and test dataset at a false positive rate of 20% (Adzhubei et al. 2010). Both SIFT and PolyPhen suggest that this variant has damaging or probably damaging effects on *SLC24A2* gene, which is

supported by the analysis results from PROVEAN and PANTHER. These results are consistent with that obtained from Mutation Taster analysis with a probability of 0.99999997 to be “disease causing.” Our 3D structure models suggest this variant has little effects on the solubility and hydrophobicity of the protein, but can decrease the stability of protein by increasing the total energy after minimization and decreasing the stabilizing residues of the protein. Furthermore, I-Mutant supports that this variant can decrease protein stability with the ddG of -0.31 .

In conclusion, we have newly identified a variant (chr9:19786004T>G) in *SLC24A2* gene by WES, which was then reconfirmed by Sanger sequencing. With multiple in silico analysis and 3D structure simulation, we suggest that this variant has probably damaging effects on the function of *SLC24A2* gene through decreasing protein stability, which is validated by cell experiment results. Thus, this variant may have potential clinical usages.

Acknowledgments

This study was funded by Shanghai Leading Academic Discipline Project (No. S30206).

Author Contributions

Conceived and designed the experiments: L.W., L.S., Z.S.L.. Performed the experiments: L.W., Z.S., S.Y.C.. Analyzed the data: Z.S., S.Y.C., L.S.. Wrote the paper: L.W., Z.S., S.Y.C., L.S., Z.S.L.. All authors gave final approval for publication.

Conflict of Interest

The authors declare no conflict of interest.

References

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. & Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248-249.
- Alshahrani, S., Alvarez-Leefmans, F.J. & Di Fulvio, M. (2012) Expression of the *Slc12a1* gene in pancreatic β -cells: molecular characterization and in silico analysis. *Cell. Physiol. Biochem.*, **30**, 95-112.
- Bhutia, Y.D., Babu, E., Ramachandran, S., Yang, S., Thangaraju, M. & Ganapathy, V. (2016) SLC transporters as a novel class of tumour suppressors: identity, function and molecular mechanisms. *Biochem. J.*, **473**, 1113-1124.
- Cai, L., Cai, M.H., Wang, M.Y., Xu, Y.F., Chen, W.Z., Qin, S.Y., Wan, C.L. & He, L. (2015) Meta-Analysis-Based Preliminary Exploration of the Connection between *ATDIL1* and Schizophrenia by *GSTM1/T1* Gene Polymorphisms. *PLoS One*, **10**, e0128643.
- Cai, L., Deng, S.L., Liang, L., Pan, H., Zhou, J., Wang, M.Y., Yue, J., Wan, C.L., He, G. & He, L. (2013) Identification of genetic associations of *SP110/MYBBP1A/RELA* with pulmonary tuberculosis in the Chinese Han population. *Hum. Genet.*, **132**, 265-273.
- Cai, L., Pan, H., Trzcinski, K., Thompson, C.M., Wu, Q. & Kramnik, I. (2010) *MYBBP1A*: a new *Ipr1*'s binding protein in mice. *Mol. Biol. Rep.*, **37**, 3863-3868.
- Cai, L., Wang, Y., Wang, J.F. & Chou, K.C. (2011) Identification of proteins interacting with human *SP110* during the process of viral infections. *Med. Chem.*, **7**, 121-126.
- Cai, L., Yang, Y.H., He, L. & Chou, K.C. (2016) Modulation of cytokine network in the comorbidity of schizophrenia and tuberculosis. *Curr. Top. Med. Chem.*, **16**, 655-665.
- Cai, X., Zhang, K. & Lytton, J. (2002) A novel topology and redox regulation of the rat brain K^+ -dependent Na^+/Ca^{2+} exchanger, *NCKX2*. *J. Biol. Chem.*, **277**, 48923-48930.
- Camacho, D., Reichenbach, D., Duerr, G.D., Venema, T.L., Sweeney, J.F. & Fisher, W.E. (2005) Value of laparoscopy in the staging of pancreatic cancer. *JOP*, **6**, 552-561.
- Capriotti, E., Fariselli, P. & Casadio, R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, **33**, W306-310.
- Chandler, M.R., Bilgili, E.P. & Merner, N.D. (2016) A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Hum. Mutat.*, **37**, 835-846.
- Chen, L., Fan, J., Chen, H., Meng, Z., Chen, Z., Wang, P. & Liu, L. (2014) The IL-8/CXCR1 axis is associated with cancer stem cell-like properties and correlates with clinical prognosis in human pancreatic cancer cases. *Sci. Rep.*, **4**, 5911.
- Chen, W., Zheng, R., Baade, P.D., Zhang, S., Zeng, H., Bray, F., Jemal, A., Yu, X.Q. & He, J. (2016) Cancer statistics in China, 2015. *CA Cancer J. Clin.*, **66**, 115-132.
- Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. & Chan, A.P. (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One*, **7**, e46688.
- Fang, S., Zhang, Y., Xu, M., Xue, C., He, L., Cai, L. & Xing, X. (2016) Identification of Damaging nsSNVs in Human ERCC2 Gene. *Chem. Biol. Drug Des.*, **88**, 441-450.
- Hao, C.Q., Zhou, Y., Wang, J.P., Peng, M.J., Xie, Y.M., Kang, W.Z., Sun, L., Wang, P.Z., Wan, C.L., He, L., Cai, L. & Jia, Z.S. (2014) Role of Nogo-A in the regulation of hepatocellular carcinoma SMMC-7721 cell apoptosis. *Mol. Med., Rep.*, **9**, 1743-1748.
- Huang, T., Liu, C.L., Li, L.L., Cai, M.H., Chen, W.Z., Xu, Y.F., O'Reilly, P.F., Cai, L. & He, L. (2016) A new method for identifying causal genes of schizophrenia and anti-tuberculosis drug-induced hepatotoxicity. *Sci. Rep.*, **6**, 32571.
- Jiang, S.Y., Li, L.L., Yue, J., Chen, W.Z., Yang, C., Wan, C.L., He, L., Cai, L. & Deng, S.L. (2016) The effects of *SP110*'s associated genes on fresh cavitary pulmonary tuberculosis in Han Chinese population. *Clin. Exp. Med.*, **16**, 219-225.
- Li, C., Gao, Z., Li, F., Li, X., Sun, Y., Wang, M., Li, D., Wang, R., Li, F., Fang, R., Pan, Y., Luo, X., He, J., Zheng, L., Xia, J., et al. (2015) Whole exome sequencing identifies frequent somatic mutations in cell-cell adhesion genes in Chinese patients with lung squamous cell carcinoma. *Sci. Rep.*, **5**, 14237.
- Lin, L., Yee, S.W., Kim, R.B. & Giacomini, K.M. (2015) SLC transporters as therapeutic targets: emerging opportunities. *Nat. Rev. Drug Discov.*, **14**, 543-560.
- Lindahl, E., Azuara, C., Koehl, P. & Delarue, M. (2006) NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.*, **34**, W52-56.
- Magyar, C., Gromiha, M.M., Pujadas, G., Tusnady, G.E. & Simon, I. (2005) SRide: a server for identifying stabilizing residues in proteins. *Nucleic Acids Res.*, **33**, W303-305.
- Maitra, A. & Hruban, R.H. (2008) Pancreatic cancer. *Annu. Rev. Pathol.*, **3**, 157-188.
- Ng, P.C. & Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.*, **31**, 3812-3814.
- Ng, P.C. & Henikoff, S. (2006) Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.*, **7**, 61-80.
- Park, J.Y., Helm, J.F., Zheng, W., Ly, Q.P., Hodul, P.J., Centeno, B.A. & Malafa, M.P. (2008) Silencing of the candidate tumor suppressor gene solute carrier family 5 member 8 (*SLC5A8*)

- in human pancreatic cancer. *Pancreas.*, **36**, e32-39.
- Schnetkamp, P.P. (2004) The SLC24 Na⁺/Ca²⁺-K⁺ exchanger family: vision and beyond. *Pflugers Arch.*, **447**, 683-688.
- Schnetkamp, P.P. (2013) The SLC24 gene family of Na⁺/Ca²⁺-K⁺ exchangers: from sight and smell to memory consolidation and skin pigmentation. *Mol. Aspects Med.*, **34**, 455-464.
- Schwarz, J.M., Rodelsperger, C., Schuelke, M. & Seelow, D. (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods*, **7**, 575-576.
- Siegel, R.L., Miller, K.D. & Jemal, A. (2015) Cancer statistics, 2015. *CA Cancer J. Clin.*, **65**, 5-29.
- Tang, B., Li, Y., Qi, G., Yuan, S., Wang, Z., Yu, S., Li, B. & He, S. (2015) Clinicopathological Significance of CDKN2A Promoter Hypermethylation Frequency with Pancreatic Cancer. *Sci. Rep.*, **5**, 13563.
- Thangaraju, M., Gopal, E., Martin, P.M., Ananth, S., Smith, S.B., Prasad, P.D., Sterneck, E. & Ganapathy, V. (2006) SLC5A8 triggers tumor cell apoptosis through pyruvate-dependent inhibition of histone deacetylases. *Cancer Res.*, **66**, 11560-11564.
- Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. & Narechania, A. (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.*, **13**, 2129-2141.
- Tomsic, J., He, H., Akagi, K., Liyanarachchi, S., Pan, Q., Bertani, B., Nagy, R., Symer, D.E., Blencowe, B.J. & de la Chapelle, A. (2015) A germline mutation in SRRM2, a splicing factor gene, is implicated in papillary thyroid carcinoma predisposition. *Sci. Rep.*, **5**, 10566.
- Ueno, M., Ohkawa, S., Morimoto, M., Ishii, H., Matsuyama, M., Kuruma, S., Egawa, N., Nakao, H., Mori, M., Matsuo, K., Hosono, S., Nojima, M., Wakai, K., Nakamura, K., Tamakoshi, A., et al. (2015) Genome-wide association study-identified SNPs (rs3790844, rs3790843) in the NR5A2 gene and risk of pancreatic cancer in Japanese. *Sci. Rep.*, **5**, 17018.
- Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M.R., Leduc, M.S., Kirby, A., et al. (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.*, **369**, 1502-1511.
- Zhang, Y. (2009) I-TASSER: fully automated protein structure prediction in CASP8. *Proteins*, **77** Suppl 9, 100-113.
-