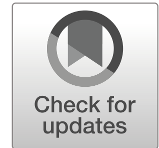
Pathology & Oncology Research (2020) 26:2237–2246
https://doi.org/10.1007/s12253-020-00815-1

ORIGINAL ARTICLE



AT-rich Interaction Domain 1A Gene Variations: Genetic Associations and Susceptibility to Gastric Cancer Risk

Jasiya Qadir[1] · Sabhiya Majid[1] · Mosin S. Khan[1] · Fouzia Rashid[2] · Mumtaz Din Wani[3] · Inshah Din[1] · Haamid Bashir[1]

Received: 16 April 2020 / Accepted: 24 April 2020 / Published online: 6 May 2020
© Arányi Lajos Foundation 2020

Abstract

AT-rich interaction domain containing protein 1A (*ARID1A*), has recently emerged as a novel class of gene which acts as a potent tumor suppressor in numerous types of cancers such as Gastric, Breast, Ovarian, Colorectal, Lung cancers. *ARID1A* is involved in the regulation of various cellular processes such as proliferation, differentiation and DNA repair, yet its association with the susceptibility of cancer remains unknown. Here, we aimed to analyse the association of the *ARID1A* variants (*Pro912Thr*, *Gln944Lys* and *Gln920Ter*) with the risk of Gastric cancer (GC) in Kashmiri population. The study included 103 confirmed cases of GC and 163 normal controls. The genotypes were studied using Polymerase Chain Reaction. Different bioinformatic predictive tools were also used to analyse the possible effect of these *SNP's* on the resultant protein. The *Pro912Thr* and *Gln920Ter* variants of *ARID1A* showed significant difference in genotypic and allelic frequencies between the GC cases and controls ($P < 0.05$), whereas, the data did not reveal any correlation between *Gln944Lys* variant and Gastric cancer risk. Both *Pro912Thr* and *Gln920Ter SNP's* follow "Dominant mode of inheritance". *In Silico* analysis predicted that amino acid substitution of *Pro912Thr* SNP decreases the protein stability thus changing the functional properties of resultant protein, so backing the possibility of damaging effect of this SNP. Our study suggests that *Pro912Thr* and *Gln920Ter* SNP's of *ARD1A* gene are associated with increased risk of GC in Kashmiri population.

Keywords ARID1A; Gastric cancer; *Pro912Thr* · *Gln944Lys* · *Gln920Ter* · Kashmiri Population

Introduction

Worldwide, Gastric cancer (GC) is one of the most frequently diagnosed cancer and the third leading cause of cancer-related deaths [1]. In 2018, over one million new cases of GC were reported and 783,000 deaths were estimated [1]. In Kashmir valley (North India), GC is the most commonly encountered cancer in men (25.2%) and third leading cancer site in women (10.4%) [2, 3]. The environmental factors and accumulation of genetic and epigenetic alterations with susceptibility of oncogenic mutations, were found to be associated with the increased risk of GC [4]. In Kashmir valley, the high incidence rate of GC have been reported to be associated with potential exposure to some of the nitroso compounds, amines and nitrates, present in the local food stuffs such as dried fish, red meat, pickled vegetables and traditional hot salted tea [5]. Mechanistic understanding of these alterations and molecular mechanism will be critical for the improving diagnosis and prognosis of GC.

ARID1A (AT-rich Interaction Domain-containing protein 1A) gene is located on chromosome 1p36.11 [6, 7]. It encodes a protein of approximately 250KD that is expressed predominantly in the nucleus [8]. ARID1A is involved in a number of protein-protein interactions, however, its interaction with the SWI/SNF chromatin remodeling complexes have been acknowledged widely [6, 8]. Recent studies have revealed that *ARID1A* gene has high mutational frequency in a wide variety cancers and decreased or loss of protein expression is significantly associated with the increased risk of several cancers

Sabhiya Majid and Mosin S. Khan have contributed equally as second authors.

✉ Sabhiya Majid
sabuumajid@gmail.com

[1] Department of Biochemistry and Research Centre University of Kashmir, Government Medical College Srinagar and Associated Hospitals , Srinagar 190010, India

[2] Department of Clinical Biochemistry, University of Kashmir, 190006 Srinagar, India

[3] Department of Surgery, Government Medical College Srinagar and Associated Hospitals, 190010 Srinagar, India

including Ovarian clear cell carcinoma, Uterine low grade endometriod carcinoma, High-grade endometrial carcinoma, primary GC and Hepatocellular carcinoma [4, 6–14].

ARID1A gene has emerged as a novel tumor suppressor gene, essential in regulating the process of cell cycle and maintaining the genomic stability [4, 6, 8, 14]. Studies have reported that polymorphism in genes regulating cell cycle, DNA mismatch repair and other metabolisms increase risk of cancer [15]. Non-synonymous polymorphisms change the protein sequence that could modify the secondary structure of proteins effecting its interactions and functioning, thereby contributing in cancer progression [15, 16].

Several single nucleotide polymorphisms (*SNP's*) have been identified within *ARID1A* gene, however, whether these genetic variation affects the initiation or progression of cancer is unknown. To fill the gap in knowledge, we studied 103 GC patients and 163 controls to evaluate the possible association of three non-synonymous *SNP's* of the *ARID1A* gene with GC risk. The study included the two missense (*Pro912Thr* and *Gln944Lys*) and one nonsense (*Gln920Ter*) SNP in exon-9 of the *ARID1A* gene. To the best of our knowledge, till date no study has been conducted that unravels the role of *ARID1A* polymorphism in GC. This is the first comprehensive analysis that investigated the possible correlation between the *ARID1A* polymorphism and susceptibility to GC.

Materials and Methods

Study Design and Study Subjects

It was a case-control hospital based prospective study conducted by the Department of Biochemistry and General Surgery, Government Medical College Srinagar and Associated Shri-Maharaja Hari Singh (SMHS) Hospital, Kashmir (North India). The study has been approved by the Ethical Committee of Govt. Medical College (GMC), Srinagar, Kashmir. The study included 103 histopathologically confirmed GC cases and 150 healthy controls attending the Department of Surgery, Govt. Medical College, Srinagar between October 2015 and June 2019. The Controls were randomly selected from a pool of healthy volunteers who visited the hospital for health check-up during the same period. None of the patients had received any radiotherapy or chemotherapy prior to surgery. A well written consent was obtained from each study subject.

Sample Collection

5 ml of blood was collected from each GC patient and healthy control in EDTA vials; refrigerated at -80 till further processing.

DNA Extraction and Polymerase Chain Reaction

Genomic DNA was isolated from blood samples by using Gen Elute™ Blood Genomic DNA kit (Sigma-Aldrich, USA) according to given protocol. The quality of DNA was checked by agarose gel electrophoresis whereas purity and concentration was measured by using the NanoDrop 2000c Spectrophotometer (ThermoScientific, USA). The Polymerase Chain Reaction (PCR) was performed to amplify the DNA segment pertaining to exon 9 using primers; Forward: 5 CACAGCACTATTTGGCTCCAG-3' and Reverse 5'- ATCATCTCTGGGCTGGCTG −3'. The PCR amplification was carried out in a 50 µl volume containing 5 µl of 50–150 ng genomic DNA; 12.5 µl of 2X PCR Master Mix (3B BlackBio, Biotech, India); 0.7 µl each of forward and reverse primers (7 pmol) (Eurofins Genomics, India Pvt Ltd). The PCR cycle conditions were as follows: initial denaturation at 94 °C for 7 min followed by 35 cycles of denaturation at 94 °C for 15 s, annealing at 58 °C for 30 s, extension for 72 °C for 30 s and final extension at 72 °C for 5 min. The PCR products of 343 bp were verified on 2% agarose gel (Fig. 1). All the amplified samples were sequencing, using the automated DNA sequencer ABI prism 310 (Applied Bio systems, USA) involving Sanger di deoxy method [17] (Fig. 2).

Computational Prediction Tools

In order to predict the possible effect of amino acid substitution on protein function different bioinformatic predictive tools were used. The missense variants were analysed using *PROVEAN, SIFT* and *MUpro*. and *Hope project tool* was used to collect the information about the phenotypes of the variants.

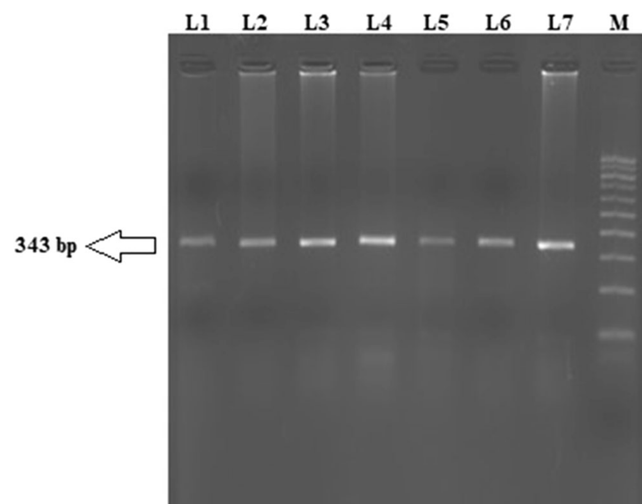


Fig. 1 Amplified PCR product of 343 bp ion size of ARID1A gene on 2.5% agarose gel
L1- L7: Lanes containing amplified PCR products with prominent/desired band, M: 100 bp Molecular size marker/Ladder

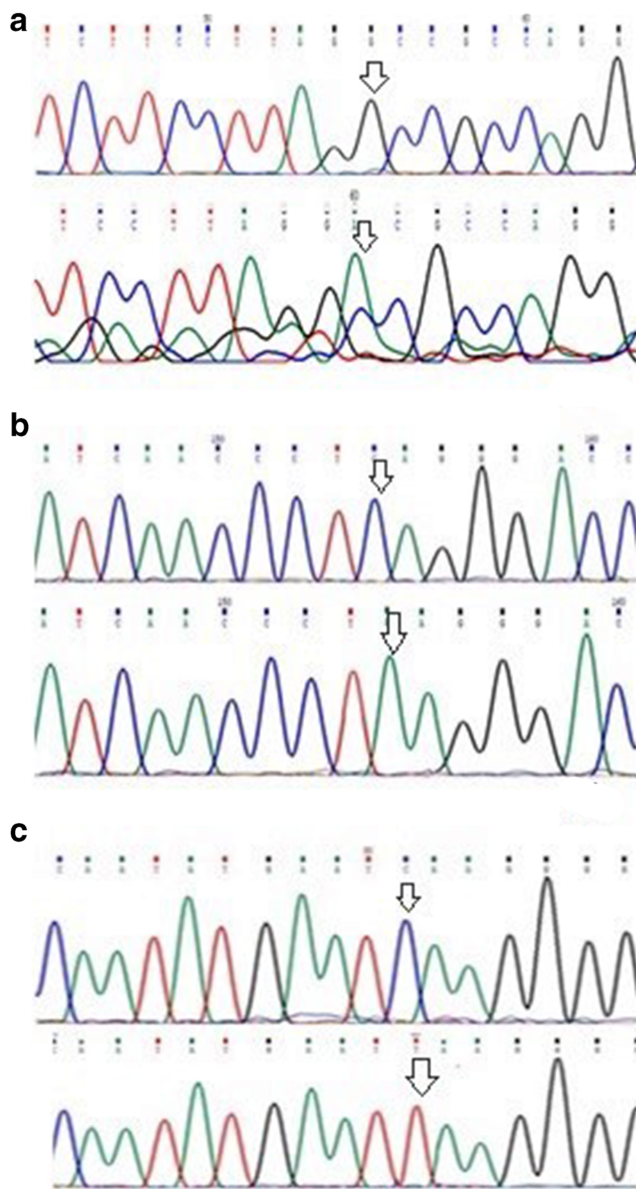


Fig. 2 Partial electrophoretograms (forward) of DNA sequences in exon 9 of *ARAD1A* gene in GC cases showing normal and mutated sequences. **A** Electrophoretogram showing C-to-A substitution (proline to threonine) at codon 912. **B** Electrophoretogram showing C-to-A substitution (glutamine to lysine) at codon 944. **C** Electrophoretogram showing C-to-T substitution resulting in generation of TAA stop codon at position 20

Statistical Analysis

Statistical analysis was done by SPSS 16.0 statistical package (SPSS Inc., Chicago IL, USA). To compare the categorical variables such as age, sex, smoking status, etc. between the cases and controls χ^2-test was used. The allelic and genotypic frequencies of cases and controls were compared by χ^2-test and Hardy-Weinberg equilibrium. The association of *VDR* genotypes with GC risk were estimated by odds ratios (OR) and 95% confidence intervals (95% CI), $P \leq 0.05$ was considered as significant.

Results

Patient Characteristics

Demographic and clinicopathological parameters of cases and controls were summarized in the Table 1. All cases and controls were matched as per their age, gender, dwelling and smoking status. The calculated mean age of the GC patients and control groups were 56.6 ± 12.1 and 53.04 ± 9.3 respectively. Interestingly, we found the incidence of GC was significantly high in patients with deranged BMI as compared to controls

Table 1 Demographic and Clinicopathological parameters of the study subjects

Variables	Cases (n = 103)	Controls (n = 163)	P value
Gender			
Male	61 (59.2%)	93 (57.0%)	> 0.05
Female	42 (40.7%)	70 (43.0%)	
Age (years)			
≥ 50	68 (66.0%)	104 (63.8%)	> 0.05
< 50	35 (34.0%)	59 (36.2%)	
Dwelling			
Rural	66 (64.0%)	99 (60.7%)	> 0.05
Urban	37 (36.0%)	64 (39.2%)	
Smoking status			
Never	59 (57.3%)	87 (53.3%)	> 0.05
Ever	44 (42.7%)	76 (46.6%)	
BMI (kg/m^2)			
Normal	54 (52.4%)	115 (70.5%)	**≤ 0.05**
Underweight	10 (9.7%)	11 (6.7%)	
Preobese	28 (27.2%)	25 (15.3%)	
Obese Class I	09 (8.7%)	10 (6.1%)	
Obese Class II	02 (1.9%)	2 (1.2%)	
Family history			
Yes	17 (16.5%)	12 (7.3%)	**≤ 0.05**
No	86 (83.5%)	151 (92.6%)	
Salt tea consumption			
< 5 cups/day	29 (28.1%)	72 (44.1%)	**0.01**
≥ 5 Cups/day	74 (71.8%)	91 (55.8%)	
H. Pylori			
Absent	65 (63.1%)		
Present	38 (36.9%)		
CEA levels			
Elevated	69 (67.0%)		
Normal	34 (33.0%)		
Stage			
I & II	70 (68.0%)		
III & IV	33 (32.0%)		
Grade			
WD	66 (64.0%)		
PD	37 (36.0%)		

BMI; Basal metabolic index (< 18.5 = Underweight, 18.5-24.99 = Normal, 25-29.99 = Preobese, 30-34.99 = Obese class I, 35-39.99 = Obese class II), CEA; Carcinoembryonic antigen, WD; Well differentiated, MD; Moderately differentiated, PD; Poorly differentiated. The p-values ≤ 0.05 are indicated in bold

(OR = 2.3, 95% CI = 1.2–4.5, $P \leq 0.05$). Furthermore, GC cases with positive family history had increased risk of GC (95% CI = 1.1–5.6, $P = 0.024$) compared to controls ($P \leq 0.05$). Helicobacter pylori (H. pylori) test was positive in 38 (36.9%) GC patients. The salted tea consumption rate was high (71.8%) in GC patients then those of controls (55.8%) and the difference was found significantly associated with the increased risk of GC (OR = 2.0, 95% CI = 1.2–3.4, $P = 0.01$).

Genotypes and Allele Distribution

The genotypic and allelic distribution of *ARID1A* gene polymorphisms (*Pro912Thr*, *Gln920Ter* and *Gln944Lys*) in GC case and controls were summarized in the Table 2. In the present study, distributions of genotype frequencies for the cases and control were in agreement with Hardy-Weinberg equilibrium ($P > 0.05$). Significant difference was observed in single-loci genotypic and allelic frequencies between the cases and controls of *ARID1A Pro912Thr* and *Gln920Ter* polymorphisms. Logistic regression revealed that patients with variant (ca. and *ca. + AA)* genotype of *Pro912Thr SNP* had 2.2 and 1.9-fold increased risk of GC than those with wild (CC) genotype. The frequency of the variant genotype (AA) was significantly high in cases compared to controls ($P \leq 0.05$). Similarly, in case of *Gln920Ter* SNP, there was 2.3 and 2.9-fold increased risk of GC among the patients with *CT* and *CT + TT* genotype compared to controls ($P \leq 0.05$). The genotypic frequency of less common variant (*TT*) was 5.8% in GC cases but altogether absent from control group. Furthermore, the *Pro912Thr* and *Gln920Ter SNP's* had significantly higher frequency of the rare alleles (*A* and *T*) in GC cases compared to control group ($P \leq 0.05$). No significant association was observed between the genotypes of *ARID1A Gln944Lys SNP* with the increased risk of GC ($P > 0.05$).

Stratification Analysis of VDR Polymorphisms and Risk of Gastric Cancer

To further assess the effect of *ARID1A Pro912Thr*, *Gln944Lys* and *Gln920Ter SNP's* on GC risk with respect to various

Table 2 Association between genotypic and allelic frequencies of the *ARID1A* polymorphisms in Gastric cancer cases and controls

Genotype/Allele	Cases (n = 103)	Controls (n = 163)	OR (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
Pro912Thr; C > A						
Genotype						
CC	48 (46.6)	106 (65.0)	1.00	**0.011**	1.00	**0.017**
ca.	42 (40.8)	46 (28.2)	2.1 (1.1–3.4)	**0.034**	2.2 (1.1–4.1)	**0.004**
AA	13 (12.6)	11 (6.7)	2.6 (1.0-6.3)	**0.003**	2.0 (1.2–3.3)	**0.002**
ca. + AA	55 (53.4)	57 (35.0)	2.1 (1.3–3.5)		1.9 (1.3-3.0)	
Allele type						
C	138 (68.4)	258 (79.1)	1.00	**0.002**	1.7 (1.2–2.2)	**> 0.001**
A	68 (31.5)	68 (20.8)	1.8 (1.2–2.8)			
Gln944Lys; C > A						
Genotype						
CC	61 (60.2)	102 (63.2)	1.00	0.80	1.0 (0.57–1.8)	0.98
ca.	33 (30.1)	49 (29.4)	1.0 (0.6–1.8)	0.48	1.3 (0.52–3.3)	0.56
AA	09 (9.7)	12 (7.3)	1.4 (0.5–3.4)	0.62	1.1 (0.65–1.8)	0.72
ca. + AA	42 (40.8)	61 (37.4)	1.1 (0.7–1.9)			
Allele type						
C	155 (75.2)	253 (77.6)	1.1 (0.7–1.7)	0.53	1.1 (0.75–1.7)	0.55
A	51 (24.7)	73 (22.4)				
Gln920Ter; C > T						
Genotype						
CC	71 (69.0)	143 (87.7)	1.00	**0.004**	1.00	**0.020**
CT	26 (25.2)	20 (12.3)	2.6 (1.3-5.0)	**0.014**	2.3 (1.1–4.6)	**0.003**
TT	06 (5.8)	00 (0.0)	14 (1.7–116)	**> 0.001**	13.8 (2.1-319.8)	**0.001**
CT + TT	32 (31.0)	20 (12.3)	3.2 (1.7–6.1)		2.9 (1.5–5.8)	
Allele type						
C	168 (81.5)	306 (93.8)	1.00	**> 0.001**	3.0 (1.9–4.9)	**> 0.001**
T	38 (18.4)	20 (6.1)	3.3 (1.9–5.8)			

demographic and clinicopathological parameters of GC cases and controls, stratification analysis was conducted as shown in Tables 3, 4 and 5 respectively. Due to the low frequency of the variant genotypes and increased risk of association, heterozygous and homozygous variants were compared against the wild genotype. For *Pro912Thr* and *Gln920Ter SNP's* the frequency of variant alleles (ca. + AA) and (CT + TT) was significantly high in the age group of ≥ 50 years (P = 0.022) (Tables 3 and 5). In case of *Pro912Thr* SNP, the frequency of variant allele (ca. + AA) in cases with family history was 6 times more as compared to controls (P = 0.034) (Table 3). In addition, frequency of variant allele (*CT* + *TT*) for *Gln920Ter SNP* was significantly high in preobese GC cases as compared to controls (P = 0.02) (Table 5). No significant association of genotypes was observed with any other clinicopathological characteristics of GC patients, both in case of *Pro912Thr* and *Gln920Ter* respectively (*P* > 0.05) *ARID1A Gln944Lys SNP* was not statistically significantly associated with GC risk (P > 0.05).

Table 3 Clinicopathological relevance of *Pro912Thr* ARID1A gene polymorphism in Gastric cancer

Pro912Thr	Case	CC	ca. + AA	Control	CC	ca. + AA	OR (95%CI)	*P* value
	n = 103	48 (46.6)	55 (53.4)	n = 163	106 (65.0)	57 (35.0)	1.9 (1.3-3.0)	0.002
Gender								
Male	61 (59.2)	26 (54.1)	35 (63.6)	93 (57.0)	55 (51.9)	38 (66.7)	1.9 (1.0-3.7)	**0.047**
Female	42 (40.8)	22 (45.8)	20 (36.3)	70 (43.0)	51 (48.1)	19 (33.3)	2.4 (1.1–5.5)	**0.031**
Age group								
≥ 50	68 (66.0)	27 (56.2)	41 (74.5)	104 (63.8)	60 (56.6)	44 (77.2)	2.0 (1.0-3.7)	**0.022**
< 50	35 (34.0)	21 (43.7)	14 (25.4)	59 (36.2)	46 (43.4)	13 (22.8)	2.6 (1.0-6.7)	0.072
Dwelling								
Rural	66 (64.0)	23 (48.0)	43 (78.2)	99 (60.7)	51 (48.1)	48 (84.2)	2.0 (1.0-3.8)	**0.036**
Urban	37 (36.0)	25 (52.0)	12 (21.8)	64 (39.2)	55 (51.9)	09 (15.8)	2.9 (1.0–8.0)	**0.035**
Smoking								
Never	59 (57.3)	29 (60.4)	30 (54.5)	87 (53.3)	58 (54.7)	29 (51.0)	2.0 (1.0-4.1)	**0.037**
Ever	44 (42.7)	19 (45.0)	25 (45.4)	76 (46.6)	48 (45.2)	28 (49.0)	2.2 (1.0-4.8)	**0.037**
BMI (kg/m^2)								
Normal	54 (52.4)	25 (52.1)	29 (52.7)	115 (70.5)	71 (67.0)	44 (77.2)	0.5 (0.3-1.0)	0.06
Underweight	10 (9.7)	4 (8.3)	06 (11.0)	11 (6.7)	08 (7.5)	03 (5.2)	3.7 (0.6–27.8)	0.17
Preobese	28 (27.2)	13 (27.1)	15 (27.3)	25 (15.3)	18 (17.0)	07 (12.3)	2.9 (1.0-9.7)	0.07
Obese Class I	09 (8.7)	05 (10.4)	04 (7.3)	10 (6.1)	08 (7.5)	02 (3.5)	3.0 (0.4–31.1)	0.31
Obese Class II	02 (1.9)	01 (2.1)	01 (1.8)	02 (1.2)	01 (1.0)	01 (1.7)	1.0 (0.0-76.5)	0.99
Family history								
Yes	17 (16.5)	03 (6.2)	14 (25.4)	12 (7.3)	07 (6.6)	05 (8.8)	6.0 (1.1–39.4)	**0.034**
No	86 (83.5)	45 (93.7)	41 (74.5)	151 (92.6)	99 (93.4)	52 (91.2)	0.8 (0.3–1.6)	0.53
Salt tea consumption								
< 5 cups/day	29 (28.1)	09 (18.7)	20 (36.3)	72 (44.1)	42 (39.6)	30 (52.6)	3.0 (1.2-8.0)	**0.014**
≥ 5 Cups/day	74 (71.8)	39 (81.2)	35 (63.6)	91 (55.8)	64 (60.4)	27 (47.3)	2.1 (1.1-4.0)	**0.020**
H. Pylori								
No	65 (63.1)	31 (64.6)	34 (61.8)	-	-	-	1.00	0.77
Yes	38 (36.9)	17 (38.0)	21 (38.1)				1.1 (0.5–2.5)	
CEA levels								
Normal	69 (67.0)	30 (62.5)	39 (71.0)	-	-	-	1.00	0.37
Elevated	34 (33.0)	18 (37.5)	16 (29.0)				0.7 (0.3–1.6)	
Stage								
I & II	70 (68.0	32 (66.7)	38 (69.0)	-	-	-	1.00	0.80
III & IV	33 (32.0)	16 (33.3)	17 (31.0)				1.0 (0.4–2.1)	
Grade								
WD	66 (64.0)	29 (60.4)	37 (67.3)	-	-	-	1.00	0.48
MD/PD	37 (36.0)	19 (39.6)	18 (32.7)				1.3 (0.6-3.0)	
mRNA expression								
Normal	57 (55.3)	31 (64.5)	26 (47.3)	-	-	-	1.00	0.08
Low	46 (44.6)	17 (35.4)	29 (52.7)				2.0 (1.0-4.5)	

*Adjusted for age, gender, dwelling, and smoking in multivariate unconditional logistic regression model. The p-values < 0.05 are indicated in bold.

Table 4 Clinicopathological relevance *Gln944Lys ARID1A* gene polymorphism in Gastric cancer

Gln944Lys	Case n = 103	*CC* 61 (60.2)	*ca. + AA* 42 (40.8)	Control n = 163	CC 102 (62.5)	ca. + AA 61 (37.4)	OR (95%CI) 1.1 (0.7–1.9)	*P* value 0.53
Gender								
Male	61 (59.2)	35 (57.4)	26 (62.0)	93 (57.0)	55 (54.0)	38 (62.3)	1.1 (0.5–2.1)	0.83
Female	42 (40.7)	26 (42.6)	16 (38.0)	70 (43.0)	47 (46.0)	23 (37.7)	1.2 (0.5–2.8)	0.58
Age group								
≥ 50	68 (66.0)	42 (68.8)	26 (62.0)	104 (63.8)	67 (65.7)	37 (60.6)	1.1 (0.6–2.1)	0.72
< 50	35 (34.0)	19 (31.1)	16 (38.0)	59 (36.2)	35 (34.3)	24 (39.3)	1.2 (0.5-3.0)	0.64
Dwelling								
Rural	66 (64.0)	39 (64.0)	27 (64.2)	99 (60.7)	66 (64.7)	33 (54.0)	1.4 (0.7–2.6)	0.33
Urban	37 (36.0)	22 (36.0)	15 (35.7)	64 (39.2)	36 (35.3)	28 (46.0)	1.0 (0.4-2.0)	0.76
Smoking								
Never	59 (57.3)	34 (55.7)	25 (59.5)	87 (53.3)	58 (56.8)	29 (47.5)	1.4 (0.7-3.0)	0.27
Ever	44 (42.7	27 (44.2)	17 (40.4)	76 (46.6)	44 (43.1)	32 (52.4)	1.0 (0.4–1.8)	0.71
BMI (kg/m^2)								
Normal	54 (52.4)	33 (54.1)	21 (50.0)	115 (70.5)	73 (71.5)	42 (69.0)	1.1 (0.5–2.1)	0.76
Underweight	10 (9.7)	07 (11.5)	03 (7.1)	11 (6.7)	06 (5.8)	05 (8.3)	0.4 (0.0-2.3)	0.51
Preobese	28 (27.2)	16 (24.2)	12 (28.6)	25 (15.3)	15 (16.7)	10 (16.7)	1.1 (0.3–3.4)	0.84
Obese Class I	09 (8.7)	04 (6.5)	05 (12.0)	10 (6.1)	06 (5.8)	04 (6.7)	1.8 (0.3–12.7)	0.54
Obese Class II	02 (1.9)	01 (1.6)	01 (2.4)	2 (1.2)	02 (1.9)	00 (0.0)	2.6 (0.1–116)	0.57
Family history								
Yes	17 (16.5)	05 (8.2)	12 (28.5)	12 (7.3)	06 (5.8)	6 (10.0)	2.3 (0.5–11.8)	0.29
No	86 (83.5)	56 (91.8)	30 (71.4)	151 (92.6)	96 (94.1)	55 (90.0)	1.0 (0.5–1.6)	0.81
Salt tea consumption								
< 5 cups/day	29 (28.1)	21 (34.4)	08 (19.0)	72 (44.1)	49 (48.0)	23 (37.7)	0.8 (0.3–2.1)	0.68
≥ 5 Cups/day	74 (71.8)	40 (65.5)	34 (80.9)	91 (55.8)	53 (51.9)	38 (62.3)	1.2 (0.6–2.2)	0.60
H. Pylori								
No	65 (63.1)	37 (60.6)	28 (66.7)	-	-	-	1.00	0.54
Yes	38 (36.9)	24 (39.3)	14 (33.3)				0.8 (0.3–1.7)	
CEA levels								
Normal	69 (67.0)	42 (68.8)	27 (64.3)	-	-	-	1.00	0.63
Elevated	34 (33.0)	19 (31.1)	15 (35.7)				1.2 (0.5–2.8)	
Stage								
I & II	70 (68.0	45 (74.0)	25 (59.5)	-	-	-	1.00	0.14
III & IV	33 (32.0)	16 (26.2)	17 (40.5)				1.9 (0.8–4.4)	
Grade								
WD	66 (64.0)	41 (67.2)	25 (59.5)	-	-	-	1.00	0.43
MD/PD	37 (36.0)	20 (32.8)	17 (40.5)				1.4 (0.6–3.2)	
mRNA expression								
Normal	57 (55.3)	35 (57.4)	22 (52.4)	-	-	-	1.00	0.62
Decreased	46 (44.6)	26 (42.6)	20 (47.6)				1.2 (0.5–2.7)	

*Adjusted for age, gender, dwelling, and smoking in multivariate unconditional logistic regression model. The p-values < 0.05 are indicated in bold

In Silico Analysis of ARID1A Gene Polymorphisms

In order to predict whether the amino acid substitution affects protein function, the significant missense variants *Pro912Thr* of *ARID1A* gene was analysed by *SIFT* and *PROVEN* computational tools. The SIFT predicted that amino acid substitution in *SNP Pro912Thr* was deleterious one with sift score ≤ 0.05. *PolyPhen-2* predicts the possibly damaging effect for *SNP Pro912Thr* with position-specific independent count (PSIC) score 0.33 (PSIC score 0 predicted benign effect and 1 predicted the greater damaging effect). *PROVEN* predicted that both *SNP's Pro912Thr* is deleterious with PROVEN score − 3.19 and − 2.68 respectively (PROVEN score ≤-2.5 is considered to be deleterious and >-2.5 neutral). In order to predict the changes in the protein stability, missense variant was subjected to *MUpro tool*. The MUpro tool predicts that the variants of *Pro912Thr* decreases the stability of protein structure compared to the

Table 5 Clinicopathological relevance of *Gln920Ter ARID1A* gene polymorphism in Gastric cancer

Gln920Ter	Case	CC	CT + TT	Control	CC	CT + TT	OR (95%CI)	*P* value
	n = 103	71 (69.0)	32 (31.0)	n = 163	143 (87.7)	20 (12.2)	3.2 (1.7–6.1)	> 0.01
Gender								
Male	61 (59.2)	41 (57.7)	20 (62.5)	93 (57.0)	79 (55.2)	14 (70.0)	2.7 (1.2-6.0)	**0.011**
Female	42 (40.7)	30 (42.2)	12 (37.5)	70 (43.0)	64 (44.7)	06 (30.0)	4.2 (1.4–13.2)	**0.008**
Age group								
≥ 50	68 (66.0)	46 (64.8)	22 (68.7)	104 (63.8)	91 (63.6)	13 (65.0)	3.3 (1.5–7.3)	**0.022**
< 50	35 (34.0)	25 (35.2)	10 (31.2)	59 (36.2)	52 (36.3)	07 (35.0)	2.9 (1.0–9.0)	0.052
Dwelling								
Rural	66 (64.0)	49 (69.0)	17 (53.1)	99 (60.7)	88 (61.5)	11 (55.0)	2.7 (1.2–6.5)	**0.020**
Urban	37 (36.0)	22 (31.0)	15 (46.9)	64 (39.2)	55 (38.4)	9 (45.0)	4.1 (1.5–11.1)	**0.004**
Smoking								
Never	59 (57.3)	39 (55.0)	20 (62.5)	87 (53.3)	74 (51.7)	13 (65.0)	2.8 (1.3–6.6)	**0.010**
Ever	44 (42.7)	32 (45.0)	12 (37.5)	76 (46.6)	69 (48.2)	07 (35.0)	3.6 (1.3–10.7)	**0.012**
BMI (kg/m^2)								
Normal	54 (52.4)	44 (62.0)	10 (31.2)	115 (70.5)	104 (72.7)	11 (55.0)	2.1 (0.8–5.5)	0.11
Underweight	10 (9.7)	4 (5.6)	06 (18.7)	11 (6.7)	09 (6.3)	02 (10)	6.0 (1.0-60.8)	0.07
Preobese	28 (27.2)	15 (21.1)	13 (40.6)	25 (15.3)	21 (14.6)	04 (20)	4.4 (1.2–18.5)	**0.02**
Obese Class I	09 (8.7)	06 (8.4)	03 (9.3)	10 (6.1)	08 (5.6)	02 (10)	1.9 (0.2–20.7)	0.56
Obese Class II	02 (1.9)	02 (2.8)	00 (0.0)	02 (1.2)	01 (0.7)	01 (5)	0.4 (0.0-8.3)	0.57
Family history								
Yes	17 (16.5)	05 (12.6)	12 (25)	12 (7.3)	09 (5.6)	03 (20.0)	6.6 (1.3–42.5)	**0.02**
No	86 (83.5)	66 (87.3)	20 (75)	151 (92.6)	134 (94.4)	17 (80.0)	2.3 (1.1–4.9)	**0.01**
Salt tea consumption								
< 5 cups/day	29 (28.1)	18 (25.3)	11 (34.4)	72 (44.1)	63 (44.0)	09 (45.0)	4.2 (1.5–12.1)	**0.006**
≥ 5 Cups/day	74 (71.8)	53 (74.6)	21 (65.6)	91 (55.8)	80 (56.0)	11 (55.0)	2.8 (1.3–6.6)	**0.010**
H. Pylori								
No	65 (63.1)	44 (62.0)	21 (71.9)	-	-	-	1.00	0.73
Yes	38 (36.9)	27 (38.0)	11 (28.1)				0.8 (0.3-2.0)	
CEA levels								
Normal	69 (67.0)	46 (64.8)	23 (71.9)	-	-	-	1.00	0.49
Elevated	34 (33.0)	25 (35.2)	09 (28.1				0.7 (0.2–1.8)	
Stage								
I & II	70 (68.0	47 (66.2)	23 (71.8)	-	-	-	1.00	0.58
III & IV	33 (32.0)	24 (33.8)	09 (28.1)				1.3 (0.5–3.4)	
Grade								
WD	66 (64.0)	46 (62.0)	20 (68.7)	-	-	-	1.00	0.51
MD/PD	37 (36.0)	25 (38.0)	12 (31.2)				1.3 (0.6–3.2)	
mRNA expression								
Normal	57 (55.3)	42 (59.1)	15 (47.0)	-	-	-	1.00	0.25
Low	46 (44.6)	29 (40.8)	17 (53.0)				1.6 (0.7–3.8)	

*Adjusted for age, gender, dwelling, and smoking in multivariate unconditional logistic regression model. The p-values < 0.05 are indicated in bold

wild type with ΔΔG= -1.45 and − 0.548 respectively. To furnish the phenotypic details the variant of *Pro912Thr* SNP was analysed by *Hope project tool*. The analysis revealed that wild type residue of *Pro912Thr* SNP has a greater hydrophobic nature compared to its variant. Further, prolines are very rigid and gives the protein backbone a special confirmation and any change may possibly disrupt the natural structure of proteins.

Genetic Association Study of ARID1A Gene Polymorphisms

For significant *SNP's* various inheritance models were applied. *Pro912Thr* and *Gln920Ter SNP's* seem to follow a Dominant inheritance model. Table 66 depicts the results of the association study for of *Pro912Thr* and *Gln920Ter* .

Table 6 Genetic association study of *ARID1A* gene polymorphism

Model	Genotype	Cases (n = 103)	Controls (n = 163)	OR (95% CI)	*P* value
Pro912Thr; C > A					
Co-Dominant	C/C	48 (46.6)	106 (65.0)	1.00 (Ref.)	0.0104
	C/A	42 (40.8)	46 (28.2)	2.2 (1.1–4.1)	
	A/A	13 (12.6)	11 (6.7)	2.0 (1.2–3.3)	
Dominant	C/C	48 (46.6)	106 (65.0)	1.0 (Ref.)	**0.002**
	C/A + A/A	55 (53.4)	57 (35.0)	1.9 (1.3-3.0)	
Recessive	CC + ca.	90 (87.4)	152 (93.2)	1.0 (Ref.)	0.11
	A/A	13 (12.6)	11 (6.7)	2.0 (0.8–4.7)	
Over-Dominant	C/C + A/A	61 (59.2)	117 (71.8)	1.0 (Ref.)	0.036
	C/A	42 (40.8)	46 (28.2)	1.7 (1.0-2.9)	
Additive	C/C	48 (46.6)	106 (65.0)	1.0 (Ref.)	0.004
	A/A	13 (12.6)	11 (6.7)	2.0 (1.2–3.3)	
Gln920Ter; C > T					
Co-Dominant	C/C	71 (69.0)	143 (87.7)	1.00	> 0.001
	C/T	26 (25.2)	20 (12.3)	2.3 (1.1–4.6)	
	T/T	06 (5.8)	00 (0.0)	26.1 (1.4-469.6)	
Dominant	C/C	71 (69.0)	143 (87.7)	1.0 (Ref.)	**0.001**
	C/T + T/T	32 (31.0)	20 (12.3)	2.9 (1.5–5.8)	
Recessive	CC + CT	97(94.2%)	163 (100%)	1.0 (Ref.)	0.007
	T/T	06 (5.8)	00 (0.0)	11.6 (1.7-267.4)	
Over-Dominant	C/C + T/T	77 (74.7%)	143 (87.7)	1.0 (Ref.)	0.008
	C/T	26 (25.2)	20 (12.3)	2.4 (1.2–4.6)	
Additive	C/C	71 (69.0)	143 (87.7)	1.0 (Ref.)	0.003
	T/T	06 (5.8)	00 (0.0)	13.8 (2.1-319.8)	

*Adjusted for age, gender, dwelling, and smoking in multivariate unconditional logistic regression model. The p-values < 0.05 are indicated in bold

Discussion

Gastric cancer is a global health issue that continues to challenge the world of medical sciences and demands constant action. ARID1A is a key non-catalytic component of SWI/SNF chromatin remodeling complex [4, 6, 7]. SWI/SNF complex is the most frequently dysregulated by ATP-dependant chromatin remodeler in cancer and its subunits are found to be missing in most tumors [4, 6]. *ARID1A* gene has recently emerged as a tumor suppressor gene that inhibits the uncontrolled proliferation of cancerous cells and participates in DNA damage repair in broad spectrum of cancers. [6]. Recent studies have reported that *ARID1A* gene has a high mutational frequency in a number of cancers including Bladder cancer, Gastric cancer, Uterine endometrioid carcinoma, Ovarian endometrioid and Clear cell carcinoma [4, 12, 18–20]. The expression of *ARID1A* gene varies during different phases of cell cycle, it is upregulating in G_0/G_1 and downregulated in S and G2/M phases, implying that *ARID1A* has a significant role in proper cell cycle arrest [4, 6, 8]. Furthermore, *ARID1A* is essential for maintaining the genomic stability via facilitating the DNA damage repair such as nucleotide excision repair and ATM regulated DNA double strand breaks repair [8].

To the best of our knowledge, this is the first study that investigates the correlation of three non-synonymous SNP's *Pro912Thr, Gln944Lys* and *Gln920Ter* of *ARID1A* gene with GC risk. We observed a strong association between the *Pro912Thr SNP* and the modulation of GC risk in Kashmiri population ($P \leq 0.05$). According to National Center for Biotechnology Information, the genetic polymorphism exhibited at position P912 of *ARID1A* is usually *Pro912Ser* (rs753300592 C > T), but in the present study we observed the predominance of *Pro912Thr* polymorphism in Kashmiri population (North India) [21]. The *Pro912Thr SNP* exhibits C to A transversion (CCG to ACG), resulting in the substitution of proline to threonine at position protein P912. Each amino acid has a unique size, charge and hydrophobicity-value that affects the conformational stability of proteins, therefore, amino acid substitution might have a significant impact on the functional properties of protein. Here, we used the Hope project tool which predicted the change in specific conformation and natural structure of proteins due to substitution of proline which has been proven in earlier studies also [22]. In line with previous studies, The MUpro tool predicted that amino acid

substitution decreases the protein stability as compared to wild type [23, 24] and might have a potential effect for altering the functional characteristics of the protein [25]. Interestingly, we observed an increased risk of GC among the older patients (≥ 50 of years) who carried the *Pro912Th*r variants of *ARID1A* gene which is in coherence with studies advocating the higher risk of Gastric cancer with advanced age [26–28]. In consistence with many studies, we observed a significantly higher frequency of variant (disease causing) allele (ca. + AA) in cases with family history of Gastrintestinal cancer (especially first degree relatives) as compared to controls [26, 29].

In the present study, we observed significant association of *Gln920Ter SNP* with increased risk of GC in Kashmiri population ($P \leq 0.05$). Our study is consistent with a study conducted in Iran that reported significantly higher prevalence of variant genotypes (*CT* and *TT*) of *Gln920Ter SNP* in patients with endometriosis compared to the control group [30]. The C to T transition of *Gln920Ter SNP* generates a premature termination codon at protein position 920, causing the premature termination of protein. The resultant protein may be completely or partially inactivated, resulting in altered or loss of the protein function. In cancer, mutations generating premature termination codons that causes the premature termination of a protein are common and accounts for 10–30% mutations in tumor suppressor genes [31]. There was a significant relationship of *Gln920Ter* SNP with the age group of ≥ 50 years which is in line with majority of studies on various cancers [32–34]. In case of *Gln920Ter* SNP, we observed a significant increased risk of 2.3-fold among preobese GC patients (BMI of 30-34.99) having variant allele (CT + TT) as compared to controls (P = 0.02). It has been reported that individuals with BMI of 30–35 have a 2-fold risk of developing GC cancer compared to individuals with BMI of < 25 [26].

GC is implicated by both genetic and environmental factors. Life style and Dietary factors play a critical role in the development of GC. The consumption of traditional salted tea is considered one of the potent factor contributing GC risk in Kashmir valley as it leads to exposure to some suspected carcinogens like nitrosamines, methylamine, ethylamine etc. [5]. In consistent with the above study, we found the rate of salted tea consumption (> 5cups/day) was significantly high in GC cases compared to controls but there was no statistical difference of GC risk between the low and high salt tea consuming groups as far as *Pro912Thr* and *Gln920Ter* SNPs are considered.

Both *Pro912Thr* and *Gln920Ter SNP's* follow Dominant mode of inheritance that assumes the wild genotypes are associated with lowest risk against the heterozygous and rare genotypes. In dominant inheritance, the carriers of heterozygous genotypes have a high risk of developing cancer compared to the wild genotype [35].

In conclusion, the GC is associated with a number of factors such as gene, environment and life-style. Our findings suggest that gene polymorphisms in exon 9 of *ARID1A* gene (*Pro912Thr* and *Gln920Ter*) may contribute significantly towards risk of GC in Kashmiri population especially in patients with advanced age, preobesity and family history of Gastrointestinal malignancy. Further larger studies in several geographic locations and multiple ethnical populations are required to verify our results.

Acknowledgements We are very much thankful to the study participants attending the tertiary care hospital for their cancer care.

Funding Information The Study was funded by the department of Biochemistry, Government Medical College Srinagar-190010.

Compliance with Ethical Standards

Conflict of Interest The authors declare that there is no conflict of interest.

Ethical Approval The study was approved by the Ethical Clearance Committee of Government Medical College and Associated Hospitals (No. 66/ETH/GMC).

Informed Consent All the samples were collected after taking written informed consent from the patients and proper ethical procedures were followed.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and Mortality worldwide for 36 cancers in 185 countries. CA: A Cancer. Journal for Clinicians 68:394–424
2. Quriesh MA, Khan SM, Masoodi MA, Qurieshi U, Ain Q, Jan YQ, Haq I, Sheikh ZA (2016) Epidemiology of cancers in Kashmir, India: an analysis of hospital data. Adv Prev Med 02:1–6
3. Iqbal QM, Ganai AM, Bhat GM, Fazili AB (2016) Pattern and magnitude of various cancers registered at regional centre of a tertiary care institute in North India. Int J Community Med Public Health 3:1672–1680
4. Wang D, Chen Y, Pan K, Wang W, Chen S, Chen J, Zhao J, Lv L, Pan Q, Li Y, Wang Q, Huang L, Ke M, He J, Xia J (2012) Decreased expression of the ARID1A gene is associated with poor prognosis in primary gastric cancer. PLoS One 7:e40364
5. Wani I, Parray F, Wani RA, Naqash SH, Wani KA, Malik AA, Choudri NA, Wani MA, Khan NA, Sheikh TA (2013) Noon chai and gastric cancer. Int J Case Rep Images 4(3):138–142
6. Wu R, Wang T, Shih L (2014) The emerging roles of ARID1A in tumor suppression. Cancer Biology therapy 15:655–664
7. Guan B, Mao TL, Panuganti PK, Kuhn E, Kurman RJ, Maeda D, Chen E, Jeng YM, Wang TL, Shih IM (2011) Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. Am J Surg Pathol 35:625–632
8. Wu JN, Roberts CWM (2013) ARID1A mutations in cancer: Another epigenetic tumor suppressor? Cancer Discov 3:35–43
9. Huang HN, Lin MC, Huang WC, Chiang YC, Kuo KT (2014) Loss of ARID1A expression and its relationship with PI3K-Akt pathway alterations and ZNF217 amplification in ovarian clear cell carcinoma. Mod Pathol 27:983–990

10. Mamo A, Cavallone L, Tuzmen S, Chabot C, Ferrario C, Hassan S, Edgren H, Kallioniemi O, Aleynikova O, Przybytkowski E, Malcolm K, Mousses S, Tonin PN, Basik M (2012) An integrated genomic approach identifies ARID1A as a candidate tumor-suppressor gene in breast cancer. Oncogene 31:2090–2100
11. Samartzis EP, Noske A, Dedes KJ, Fink D, Imesch P (2013) ARID1A Mutations and PI3K/AKT Pathway Alterations in Endometriosis and Endometriosis-Associated Ovarian Carcinomas. Int J Mol Sci 14:18824–18849
12. Jones S, Meng L, Parsons W, Zhang X, Wesseling J, Kristel P, Schmidt MK, Markowitz S, Yan H, Bigner D, Hruban RH, Eshleman JR, Iacobuzio-Donahue CA, Goggins M, Maitra A, Malek SN, Powell S, Vogelstein B, Kinzler KW, Velculescu VE, Papadopoulos N (2012) Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. Hum Mutat 33:100–103
13. Wiegand KC, Lee AF, Al-Agha OM, Chow C, Kalloger SE, Scott DW, Steidl C, Wiseman SM, Gascoyne RD, Gilks B, Huntsman DG (2011) Loss of BAF250a (ARID1A) is frequent in high-grade endometrial carcinomas. J Pathol 224:328–333
14. Cajuso T, Hänninen UA, Kondelin J, Gylfe AE, Tanskanen T, Katainen R, Pitkänen E, Ristolainen H, Kaasinen E, Taipale M, Taipale J, Böhm J, Renkonen-Sinisalo L, Mecklin J, Järvinen H, Tuupanen S, Kilpivaara O, Vahteristo P (2014) Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2, and ARID4A in microsatellite unstable colorectal cancer. Int J Cancer 135:611–623
15. Deng N, Zhou H, FanYuan YH (2017) Single nucleotide polymorphisms and cancer susceptibility. Oncotarget 8:110635–110649
16. Chu D, Wai L (2019) Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selection than expectation. BMC Cancer 19:359–365
17. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467
18. Fadare O, Renshaw IL, Liang SX (2012) Does the loss of ARID1A (baf-250a) expression in endometrial clear cell carcinomas have any clinicopathologic significance? A pilot assessment. J Cancer 3:129–136
19. Yang L, Wei S, Zhao R, Wu Y, Qiu H, Xiong H (2016) Loss of ARID1A expression predicts poor survival prognosis in gastric cancer: a systematic meta-analysis from 14 studies. Sci Rep 6:2891–2899
20. Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S, Chan TL, Kan Z, Chan ASY, Tsui WY, Lee SP, Ho SL, Chan AKW, Cheng GHW, Robert PC, Rejto PA, Gibson NW, Pocalyko DJ, Moa MXJ, Leung SY (2011) Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. Nat Genet 43:1219–1212
21. Database of Single Nucleotide Polymorphism (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: http://www.ncbi.nim.nih.gov/snp/?term=rs753300592)
22. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G (2010) Protein structure analysis of mutations causing inheritable diseases: an e-Science approach with life scientist friendly interfaces. BMC Bioinformatics 11:548–554
23. Capriotti E, Fariselli P, Calabrese R, Casadio R (2005) Predicting protein stability changes from sequences using support vector machines. Bioinformatics 21(Suppl 2):54–58
24. Cheng J, Randall A, Baldi P (2005) Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62:1125–1132
25. Osier MV, Pakstis AJ, Goldman D, Edenberg HJ, Kidd JR, Kidd KK (2002) A proline–threonine substitution in codon 351 of ADH1C is common in Native Americans. Clin Exp Res 26:1759–1763
26. Karimi P, Islami F, Anandasabapathy S, Freedmen ND, Kamangar F (2014) Gastric Cancer: Descriptive epidemiology, risk factors, screening and prevention. Cancer Epidemiol Biomarkers Prev 23:700–713
27. Lee SR, Kim HO, Yoo CH (2012) Impact of chronologic age in the elderly with gastric cancer. J Korean Surg Soc 82:211–218
28. Pak LM, Wang J (2006) The appropriate treatment for elderly gastric cancer patients. Art Surg 1:4–9
29. Bernini M, Barbi S, Roviello F, Scarpa A, Moore P, Pedrazzani C, Beghelli S, Marrelli D, Manzoni GD (2006) Family history of gastric cancer: a correlation between epidemiologic findings and clinical data. Gastric Cancer 9:9–13
30. Tavana Z, Khalili A, Namazi G, Ebrahimi A, Davoodi S, Alborizi S, Roozmeh S (2018) Prevalence of common polymorphisms of AT-rich interaction domain 1A and endothelial nitric oxide synthase in patients with endometriosis compared to control group. J Endometriosis Pelvic Pain Disord 10:26–31
31. Bidou L, Bugaud O, Belakhov V, Baasov T, Namy O (2017) Characterization of new-generation aminoglycoside promoting premature termination codon readthrough in cancer cells. RNA Biol 14:378–388
32. Ueno D, Matsumoto H, Kubota H, Higashida M, Akiyama T, Shiotani A, Hirai T (2017) Prognostic factors for gastrectomy in elderly patients with gastric cancer. World J Surg Onc 15:59–65
33. Neuhaus H (2017) Early gastric cancer in super-agers: to treator not to treat? Gastrointest Endosc 85:973–975
34. Lin Y, Ueda J, Kikuchi S, Totsuka Y, Wei WQ, Qiao YL, Inoue M (2011) Comparative epidemiology of gastric cancer between Japan and China. World J Gastroenterol 17:4421–4428
35. Wang Q (2016) Cancer predisposition genes: molecular mechanisms and clinical impact on personalized cancer care: examples of Lynch and HBOC syndromes. Acta Pharmacol Sin 37:143–149

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