Biochimica et Biophysica Acta 1852 (2015) 1059–1067

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbadis

# SNP rs1049430 in the 3′-UTR of SH3GL2 regulates its expression: Clinical and prognostic implications in head and neck squamous cell carcinoma

Guru Prasad Maiti [a,b], Amlan Ghosh [c], Pinaki Mondal [d], Aradhita Baral [e], Sayantan Datta [f,g], Sudip Samadder [a], Sandeep P. Nayak [h], Jayanta Chakrabarti [i], Jaydeep Biswas [i], Nilabja Sikdar [f], Shantanu Chowdhury [e], Bidyut Roy [f], Susanta Roychowdhury [j], Chinmay Kumar Panda [a,*]

[a] Department of Oncogene Regulation, Chittaranjan National Cancer Institute, 37, S. P. Mukherjee Road, Kolkata 700026, India
[b] Department of Molecular Biology and Biotechnology, University of Kalyani, Kalyani, Nadia, WB, India
[c] Department of Biological Science, Presidency University, Kolkata, India
[d] National Brain Research Centre, NH-8, Manesar, Gurgaon, Haryana 122051, India
[e] Proteomics and Structural Biology Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi, India
[f] Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India
[g] Department of Ophthalmology, Johns Hopkins School of Medicine, Baltimore, MD, USA
[h] Department of Surgical Oncology, Kidwai Memorial Institute of Oncology, Marigowda Road, Bangalore 560029, India
[i] Department of Surgical Oncology, Chittaranjan National Cancer Institute, 37, S. P. Mukherjee Road, Kolkata 700026, India
[j] Cancer Biology and Inflammatory Disorder Division, CSIR-Indian Institute of Chemical Biology, Kolkata 700032, India

ARTICLE INFO

*Article history:*
Received 25 January 2015
Accepted 20 February 2015
Available online 26 February 2015

*Keywords:*
3′-UTR
Allele-specific isoform
HNSCC
rs1049430
SH3GL2

ABSTRACT

Single nucleotide polymorphisms (SNPs) in the 3′-UTR region are emerging cis-regulatory factors associated with the occurrences of several human diseases. SH3GL2, which is located at chromosome 9p21-22, is associated with hyperplastic/mildly dysplastic lesions of the head and neck and has a long 3′-UTR with multiple SNPs. The aim of the present study was to determine the susceptible allele(s) in the 3′-UTR SNPs of SH3GL2 in head and neck squamous cell carcinoma (HNSCC). First, we screened the genotypes of all SNPs located in the 3′-UTR of SH3GL2 in 110 controls and 147 cases in Indian populations by sequencing. A SNP (rs1049430:>G/T) that showed only heterozygosity was further confirmed by genotyping with an Illumina GoldenGate platform in 530 controls and 764 cases. Genotype-specific survival analysis of the HNSCC patients was performed. In addition, genotype-specific mRNA stability, isoform expression and protein expression were analyzed. SNP rs1049430 was not associated with disease occurrence, but it was associated with poor patient outcome. The G allele was associated with decreased SH3GL2 mRNA stability, differential splicing and low protein expression. Thus, our data demonstrate that the presence of the susceptible G allele in SNP rs1049430 is associated with the inactivation of SH3GL2 and could be used as a prognostic marker of HNSCC.



## 1. Introduction

The Src-homology-3 domain (SH3) of intracellular proteins is an evolutionarily conserved 50- to 60-amino acid module that is involved in the transduction of signals for cell polarization, motility, enzymatic activation and transcriptional regulation [26]. SH3-domain GRB2-like-2 protein (SH3GL2, also called endophilin-1) has been found to mediate protein–protein interactions that are important for the endocytosis of the epidermal growth factor receptor (EGFR) to allow for its homeostasis [8]. Downregulation of SH3GL2 is associated with the development of different human cancers, and upregulation of the gene is associated with Alzheimer's disease [7,9,23,27,29]. In our previous study, frequent alterations (deletion/promoter methylation) and reduced expression of SH3GL2 have been found to be associated with hyperplastic/mild dysplastic lesions of the head and neck [10,17]. The prognostic importance of this gene has also been evaluated, and it has been suggested to be a candidate tumor suppressor gene for the development of head and neck squamous cell carcinoma (HNSCC).

SH3GL2 has 10 exons, with a short 5′-UTR and a long 3′-UTR (approximately 1.4 kb in length) (http://asia.ensembl.org). The 3′-UTR of SH3GL2 harbors multiple SNPs. It is evident that SNPs located in the 3′-UTR of a gene can regulate alternative polyadenylation and mRNA stability [12,13,18], and their associations with different diseases, including tumors, have been reported [1,5,6,15,21,22,24,30,31]. In contrast with deletions and methylation events, allelic variations in SNPs in the 3′-UTR of SH3GL2 have not yet been evaluated with regard to their importance in the metabolism of this gene as well as their associations with diseases, particularly HNSCC.

Therefore, the aim of this study was to determine the susceptible allele(s) of 3′-UTR SNPs of SH3GL2 in Indian patients with HNSCC and also to evaluate its importance in the progression and prognosis of

* Corresponding author. Tel.: +91 33 24743922; fax: +91 33 2475 7606.
*E-mail address:* ckpanda.cnci@gmail.com (C.K. Panda).

http://dx.doi.org/10.1016/j.bbadis.2015.02.009
0925-4439/

**Table 1**
Basic characteristics of case and control samples.

| Variables | Case (n = 764) | Control (n = 530) | p-Value |
|---|---|---|---|
| Gender, n (%) | | | |
| Male | 559 (73.2) | 407 (76.7) | 0.806 (NS) |
| Female | 169 (22.1) | 119 (22.4) | |
| Unknown | 36 (4.7) | 4 (0.7) | |
| Age, mean range | 46.5 (20–88) | 48.1 (22–85) | NS |
| Tobacco | | | |
| Yes | 539 (70.5) | 316 (59.6) | 0.00004 (S) |
| No and unknown | 225 (29.4) | 214 (40.4) | |

NS: non-significant and S: significant.

this disease. Thus, we screened SNPs in the 3′-UTR of SH3GL2 and identified a candidate SNP rs1049430 (G > T) that showed heterozygosity in the Indian population. Expression analysis (RNA/protein) of the different alleles of this SNP suggested that the G′ allele could negatively regulate the expression of this gene and that it was significantly associated with the clinical outcome of this disease.

## 2. Materials and methods

### 2.1. Patient population, tumor tissues, control population and cell lines

Tissue and blood samples from head and neck lesions were collected from patients at the Chittaranjan National Cancer Institute, Kolkata, India. The samples were frozen immediately after collection at −80 °C until use. Part of the freshly operated tissues was collected in TRIzol (Invitrogen, USA) for RNA isolation. To perform a case–control study, 764 head and neck lesions from 764 patients receiving no prior treatment and 530 blood samples from ethnically matched but unrelated control individuals with no past or present history of HNSCC were collected. Oral swabs and blood samples were collected from 20 control individuals. This study was approved by the institutional ethical committee, and written informed consent was obtained from the case and control individuals. Among the 764 head and neck lesions, 253 were leukoplakia, and 511 were invasive HNSCC. Detailed clinicopathological histories of the patients and information pertaining to the unrelated controls are presented in Table 1.

To observe genetic diversities among Indian populations, individuals were recruited from three isolated populations, including the Rabha (Tibeto-Burman, n = 32), Santhal (Austro-Asiatic, n = 19) and Madia (Dravidian, n = 32) tribal populations. Written informed consent was also obtained from the volunteers. Detailed clinicopathological histories of the patients and information pertaining to the unrelated controls are presented in Table 1.

The cell lines Hep2, KB and MCF7 were procured from the National Center for Cell Sciences, Pune, India, and UPCI:SCC084 was kindly provided by Prof. Susanne M. Gollin from the University of Pittsburgh, PA. A summary of the total samples used in the different experiments is shown in Fig. 1.

### 2.2. DNA extraction

Samples containing >60% dysplastic epithelial/tumor cells were obtained for isolation of DNA according to the standard procedure [19]. High-molecular-weight DNA from tissue samples was extracted by proteinase K digestion, followed by phenol/chloroform extraction. Genomic DNA from peripheral blood lymphocytes was isolated using QIAGEN Blood DNA Isolation Kits according to the manufacturer's protocol.

### 2.3. RNA isolation and cDNA synthesis

Total RNA was isolated from 10 primary HNSCC tumors and four cell lines using TRIzol Reagent (Invitrogen, USA) according to the manufacturer's protocol as described previously [17].

### 2.4. Genetic and epigenetic alteration analysis

In our previous study, the deletion and promoter methylation status of SH3GL2 has been published using the same set of 147 HNSCC samples [17].

### 2.5. Genotyping of SNP(s) located in the 3′-UTR of SH3GL2

The entire 3′-UTR region of SH3GL2 was amplified by PCR in 110 control samples, 147 tumor samples and 4 cell lines using two sets of primers (Table S1). PCR products were analyzed by direct sequencing in both the sense and antisense directions with a BigDye Terminator Cycle Sequencing 3.1 Kit (Applied Biosystems, USA) according to the manufacturer's instructions.

To perform a case–control study, the genotyping of informative SNP(s) was conducted in 764 tumor and 530 control samples in a batch of SNP genotyping experiments as described in our previous study [20] using an Illumina GoldenGate genotyping platform, according to the manufacturer's protocol.

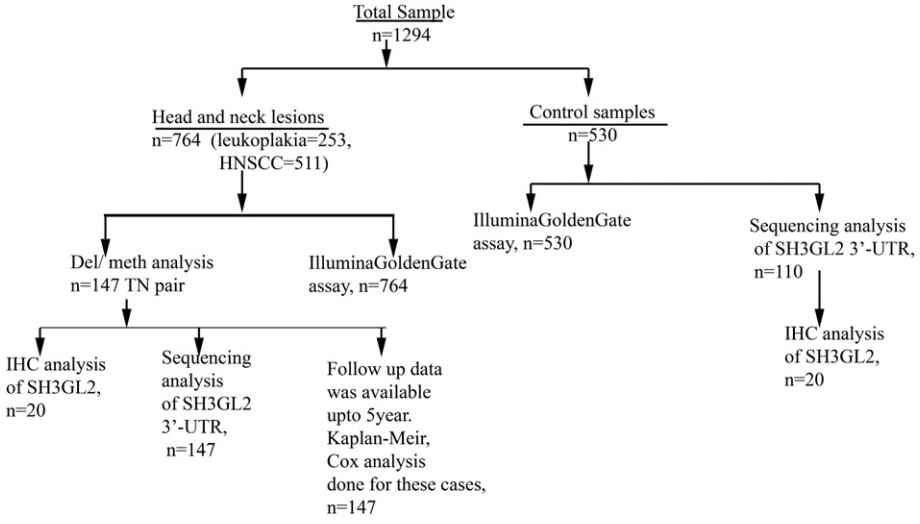


**Fig. 1.** Summary of the samples used in different experiments. 'n' indicates the sample number. Del; deletion analysis, meth; methylation analysis, TN; tumor and adjacent normal.

### 2.6. Statistical analysis

Fisher's exact test was used to evaluate case–control differences in the distribution of rs1049430 genotypes and HNSCC risk. Four different comparisons were made in the case–control study as follows: i) head and neck lesions (leukoplakia + HNSCC) versus normal tissue; ii) HNSCC versus leukoplakia; iii) HNSCC versus normal tissue; and iv) leukoplakia versus normal tissue. For each set, p-values, odds ratios (OR) and 95% CIs were determined. All statistical tests were 2-sided and were considered significant at a $p < 0.05$.

Complete follow-up data were obtained from the 147 HNSCC patients. Curves were calculated according to the Kaplan–Meier method for the 147 HNSCC samples, for which survivability was evaluated with this method with the log rank test for censored survival data. Alterations (deletion/methylation) in SH3GL2 in the 147 HNSCC samples have been published in our previous paper [17]. A Cox-proportional hazards model was used to determine the statistical significances of several potential prognostic factors. From this model, we estimated the hazard ratio (HR) for each potential prognostic factor with a 95% confidence interval (CI) in a multivariate fashion. All statistical analyses were performed using statistical programs Epi Info 6.04b and SPSS 9.0 (SPSS, Chicago, IL).

### 2.7. mRNA stability assay

Overnight subconfluent cultures of the cell lines Hep2 and MCF7 ($1 \times 10^6$ cells) were treated with actinomycin-D (Sigma; 5 μg/ml) in six 35-mm petri dishes [15,25]. A control without actinomycin-D was cultured concomitantly in the same manner. After treatment, the cells were harvested at different time points (0 h, 4 h, 8 h, 12 h, 24 h and 48 h) of incubation followed by total RNA isolation using TRIzol Reagent (Invitrogen, USA) according to the manufacturer's protocol.

Real-time quantification of SH3GL2 mRNA was performed with an ABI Prism 7500 using Power SYBR Green PCR Master Mix (Applied Biosystems, USA), as described in our previous study [16]. The primers used in this experiment are shown in Table S1.

### 2.8. Allele-specific expression assay

An allele-specific expression assay of the Hep2, KB and MCF7 cell lines was performed according to the standard procedure [2,25]. Briefly, genomic DNA and corresponding cDNA of SH3GL2 were amplified separately by PCR using primers encompassing all SNPs assessed. The sequencing reaction was performed as described in an earlier section ("Section 2.5"). For each sample, the peaks at the position of the SNP for the genomic DNA versus the cDNA were determined. The primers used in this experiment are shown in Table S1.

### 2.9. Isoform expression analysis

To determine the role of 3′-UTR genetic variations in SH3GL2 in its differential alternative splicing, semi-quantitative reverse transcription PCR was performed on Hep2, KB and MCF7 cell lines with specific primer sets (Table S1). 18S rRNA was used as a housekeeping control. Briefly, PCR was performed for 35 cycles, and the products were electrophoresed on 2% agarose gels and visualized with a Gel Documentation system (BioRad, USA).

### 2.10. Immunohistochemical (IHC)/immunocytochemistry (ICC) analysis

Assessment of the protein expression of *SH3GL2* was conducted by immunohistochemical analysis for 20 primary HNSCC samples and by immunocytochemical analysis for 20 normal exfoliated cells of oral swab and cancer cell lines (MCF7, KB, and SCC084) according to the

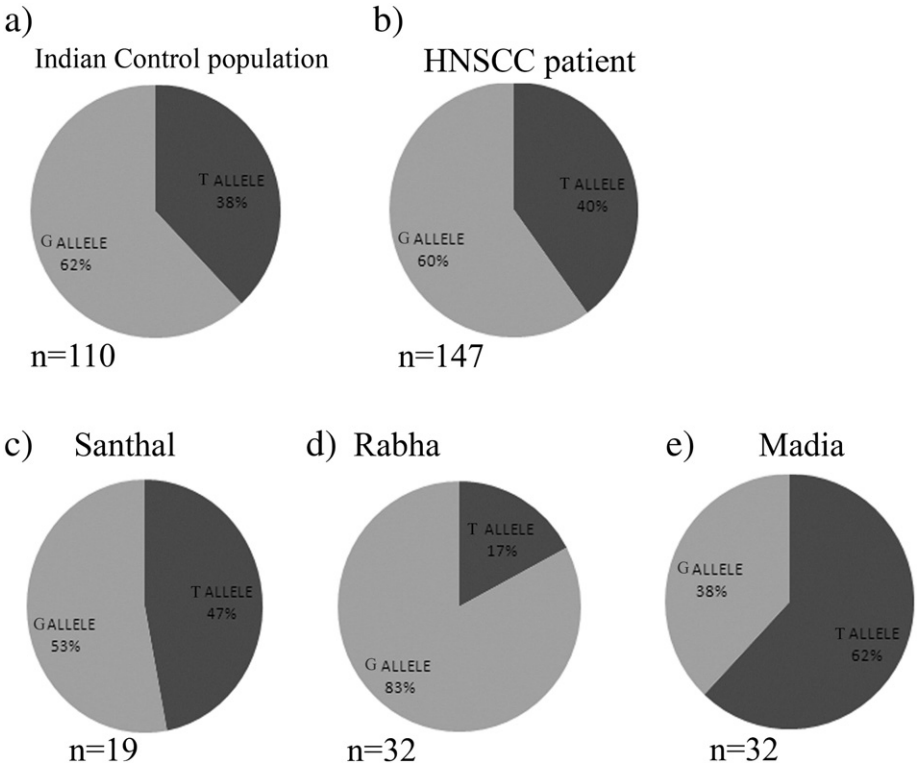


**Fig. 2.** Allele distribution of SNP rs1049430 in HNSCC patients and different isolated control populations. Pie-charts show the G and T allele frequencies of rs1049430 in a) Indian control population, b) HNSCC disease population, c) isolated Santhal population, d) isolated Rabha population and e) isolated Madia population. 'n' indicates the number of samples used in the experiment. Percent (%) of each allele in each population is shown in the chart.

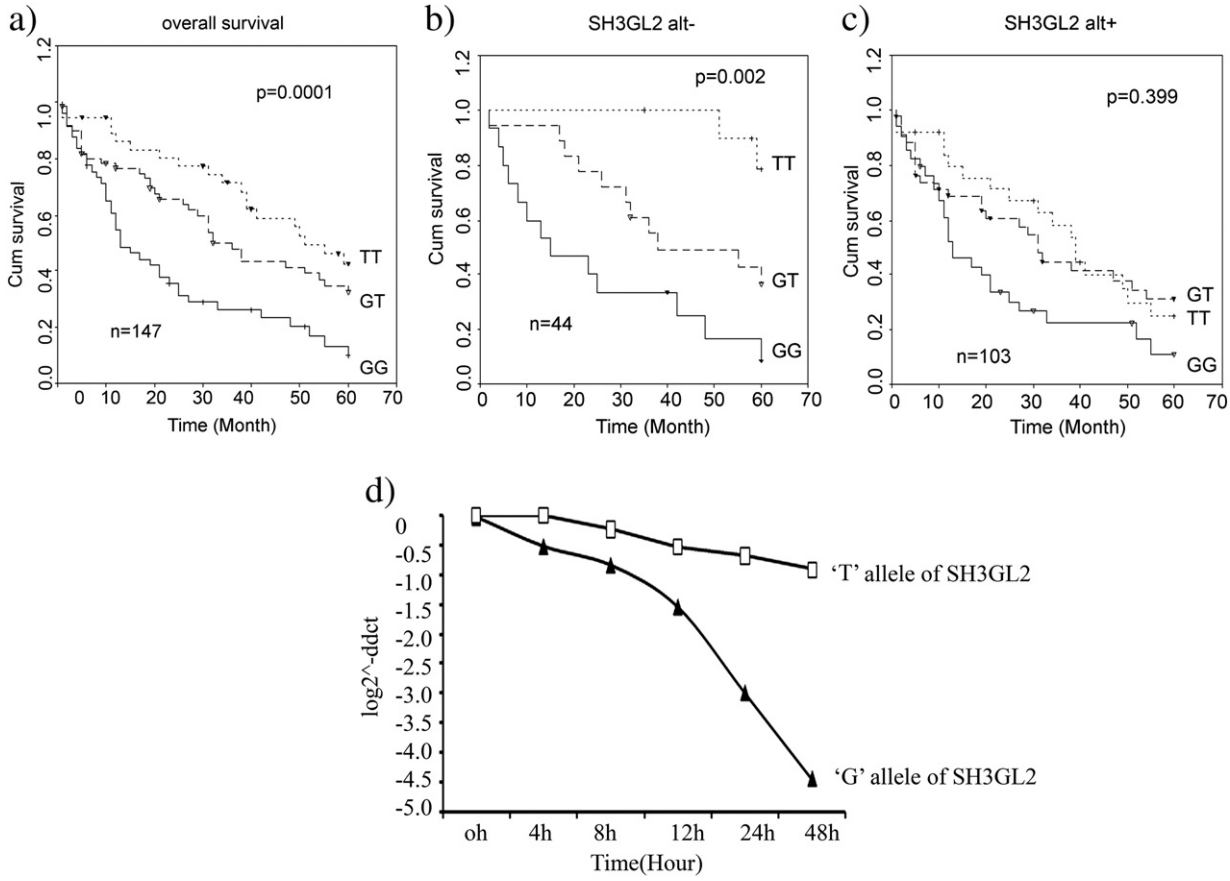


**Fig. 3.** rs1049430 allele specific survival analysis of HNSCC patients. a) Kaplan–Meier estimation on overall survival in HNSCC patients of three genotype groups (GG, GT, TT) of SNP rs1049430 irrespective of genetic alterations. b) The effect of rs1049430 on overall survival of the HNSCC patients having no genetic alterations of SH3GL2. c) The effect of rs1049430 on overall survival of the HNSCC patients with genetic alterations of SH3GL2. n; number of patients selected, alt+; genetic/epigenetic alteration present, alt−; genetic/epigenetic alteration absent. Log rank p value indicates the significance level. d) Allele specific mRNA stability analysis of SH3GL2: qRT-PCR analysis of SH3GL2 mRNA was performed after inhibition of transcription by actinomycin-D treatment for different time points. Hep2 and MCF7 cell lines were taken for G and T allele containing mRNAs respectively. Sub-confluent cultures of the cell lines were treated with actinomycin-D (5 μg/ml). The cells were harvested after 0 h, 4 h, 8 h, 12 h, 24 h and 48 h of treatment and total RNA was isolated. SH3GL2 mRNA was quantitated by SYBR green method and plotted against time.

standard procedure described earlier [16]. A primary antibody for *SH3GL2* (sc-10874), an HRP-conjugated rabbit anti-goat (sc-2020) secondary antibody, 3-3-diaminobenzidine (DAB), and a FITC-conjugated secondary antibody (rabbit anti-goat, sc-2777) were purchased from Santa Cruz Biotechnology Inc., USA. The slides were photographed with a bright-field microscope (Leica DM1000, Germany). Imaging of the cover slips was performed with a fluorescence microscope (Leica DM4000B, Germany).

Staining intensity was determined independently by 2 observers using a 4-tiered system as follows: low or negative (+), moderately positive (++), strongly positive (+++) or intensely positive (++++) [28]. The two scores were combined for the final evaluation of expression.

### 2.11. Bioinformatics analysis

To find SNPs located within the 3′-UTR of the SH3GL2 gene, the NCBI SNP database (http://www.ncbi.nlm.nih.gov/snp/) was used (Table S2). Four publicly available algorithms, PicTar (http://pictar.mdc-berlin.de/), TargetScan (http://www.targetscan.org/), Microcosm Targets (www.ebi.ac.uk-srv/microcosm/) and DIANA microT (http://diana.pcbi.upenn.edu/), were used to predict the human miRNAs (Table S3). The secondary structure and free energy change (ΔG) of the SH3GL2 mRNA with the allelic variation of rs1049430 were predicted using the mFOLD tool (http://mfold.rna.albany.edu/) [11].

## 3. Results

### 3.1. Genotyping of the 3′-UTR SNPs of SH3GL2 in Indian populations

All of the SNPs except rs1049430 were non-informative in the Indian control population (Table S2). SNP rs1049430 showed 62% G and 38% T alleles in the population (Fig. 2). However, differential allele frequencies were observed in the three ethnic Indian tribal populations (Santhal, Rabha and Madia), with significant differences in allele frequencies in Madia ($p = 0.0018$) and Rabha ($p = 0.00036$) compared with the control population (Table S4). Among the cell lines, Hep2 contained

**Table 2**
Multivariate analysis of the SNP (rs1049430) with different clinicopathological parameters.

| Univariate | Overall survival | | |
|---|---|---|---|
| Variable | p value | HR | 95% CI for HR |
| rs1049430 | <**0.0001** | 1.9753 | 1.4872–2.6235 |
| SH3GL2 ALT | **0.0085** | 1.8849 | 1.1753–3.0231 |
| Node | **0.0205** | 1.7805 | 1.0930–2.9003 |
| Stage | **0.0706** | 0.7919 | 0.6150–1.0198 |
| Histology | **0.0114** | 0.6300 | 0.7470–1.2237 |
| HPV | **0.0398** | 0.6460 | 0.4258–0.9799 |
| Tobacco | 0.9075 | 1.0298 | 0.6274–1.6903 |

ALT, genetic alteration; HPV, human papillomavirus; bold letters indicate the significant or borderline significant value.

GG alleles, KB contained GT alleles and MCF7 had TT alleles at this SNP (Table S4b). In the case–control study, comparisons of GT, GG and GT+GG with respect to the TT genotype were performed. No significant association of the SNP with the development of i) head and neck lesions (leukoplakia + HNSCC) compared with normal tissue; ii) HNSCC compared with leukoplakia; iii) HNSCC compared with normal tissue; or iv) leukoplakia compared with normal tissue was observed (Table S5).

### 3.2. SNP rs1049430 and the prognosis of HNSCC

Survival analysis of HNSCC cases (n = 147) stratified by genotypes revealed that the carriers of the GT and TT genotypes of SNP rs1049430 of SH3GL2 had better survival rates compared with the GG carriers (p = 0.0001) (Fig. 3a, b, c). A similar phenomenon was also evident in the HNSCC patients without alterations in SH3GL2. However, no significant difference in disease prognosis was detected among the

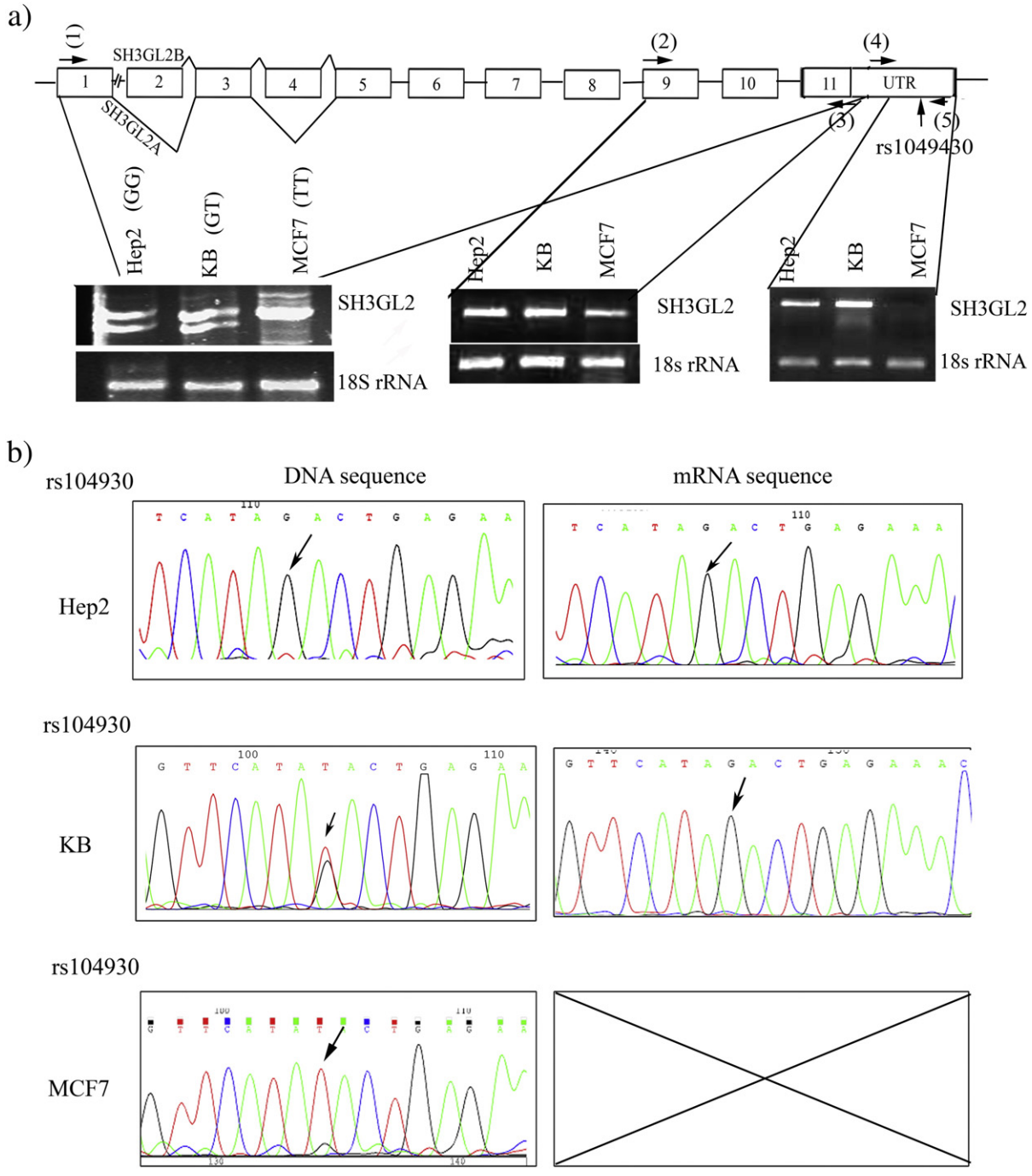


**Fig. 4.** Allele specific expression analysis of SH3GL2: Allele expression analysis was performed by RT-PCR and sequencing of genomic DNA and corresponding mRNA at the SNP rs1049430 region of SH3GL2. a) The genetic organization of SH3GL2 with different primer locations used in this study. PCR amplification of SH3GL2 mRNA using primer set (1) and (3) from exons 1 and 11 showing total two isoforms in GG containing cell line Hep2 and GT containing KB, one isoform in TT allele containing MCF7 cell line. PCR amplification of SH3GL2 mRNA using primer set (2) and (3) from exons 9 and 11 respectively in Hep2, KB and MCF7 showing single transcript. Amplification of 3-UTR region of SH3GL2 in Hep2, KB and MCF7 using primer set (4) and (5) showed the band in Hep2 and KB but no band in MCF7. b) The chromatogram showing the rs1049430 genotype GG in Hep2 cell line in both genomic DNA and mRNA. Both G and T alleles of the SNP were present in genomic DNA, but only G allele was present in mRNA of KB cell line. Genotype of the SNP in MCF7 cell line showed that only T allele was present in genomic DNA and no allele was expressed in mRNA.

genotypes with alterations (deletion and/or methylation) in SH3GL2 for the tumors ($p = 0.399$). Genetic alterations in SH3GL2 have been reported in the same set of samples ($n = 147$) in our previous study [17]. Multivariate analysis showed that patients with the GG genotype ($p = 0.0001$; HR = 1.97; CI = 1.4872–2.6235) along with SH3GL2 alterations ($p = 0.0085$; HR = 1.8849; CI = 1.1753–3.0231), a higher histological grade and nodal invasion ($p = 0.0205$; HR = 1.7805; CI = 1.0930–2.9003) in the absence of HPV infection ($p = 0.0398$; HR = 0.6460; CI = 0.6274–1.6903) in tumors had a worse prognosis (Table 2).

### 3.3. The G allele is associated with less stable secondary structure of SH3GL2 mRNA

To predict the possible mechanism of SNP rs1049430 (G/T) in the differential survival of HNSCC patients, we analyzed the RNA folding and free energy change of SH3GL2 mRNA (Fig. S1). Bioinformatics analysis revealed that the predicted minimal free energy (MFE) of the secondary structure of mRNA for the T allele was lower than that for the G allele ($\Delta G = -27.20$ kcal/mol vs. $\Delta G = -25.16$ kcal/mol). This result indicated that the mRNA structure associated with the T allele may be more stable than that associated with the G allele.

The validation of bioinformatics analysis was performed by an allele-specific SH3GL2 mRNA stability assay of the Hep2 and MCF7 cell lines (Fig. 3d). The expression of SH3GL2 was significantly reduced in the Hep2 cells after 12 h of actinomycin-D treatment. In contrast, in the MCF7 cells, the expression of SH3GL2 did not change significantly, even after 48 h of actinomycin-D treatment. This finding indicates the differential stability of SH3GL2 mRNA carrying the SNP with either the G or T allele.

### 3.4. Allele-specific SH3GL2 expression analysis

To qualitatively assess the relative abundance of the allele-specific mRNA level of SH3GL2, RT-PCR was performed using one primer set encompassing the SNP (primer set 4 & 5) (Table S1) and another primer set encompassing exons 8 and 10 (primer set 2 & 3) using cDNA from the Hep2, KB and MCF7 cell lines.

Using the SNP primer set 4 & 5, no PCR product was found for the MCF7 cells, whereas with the primer set 2 & 3, PCR products were

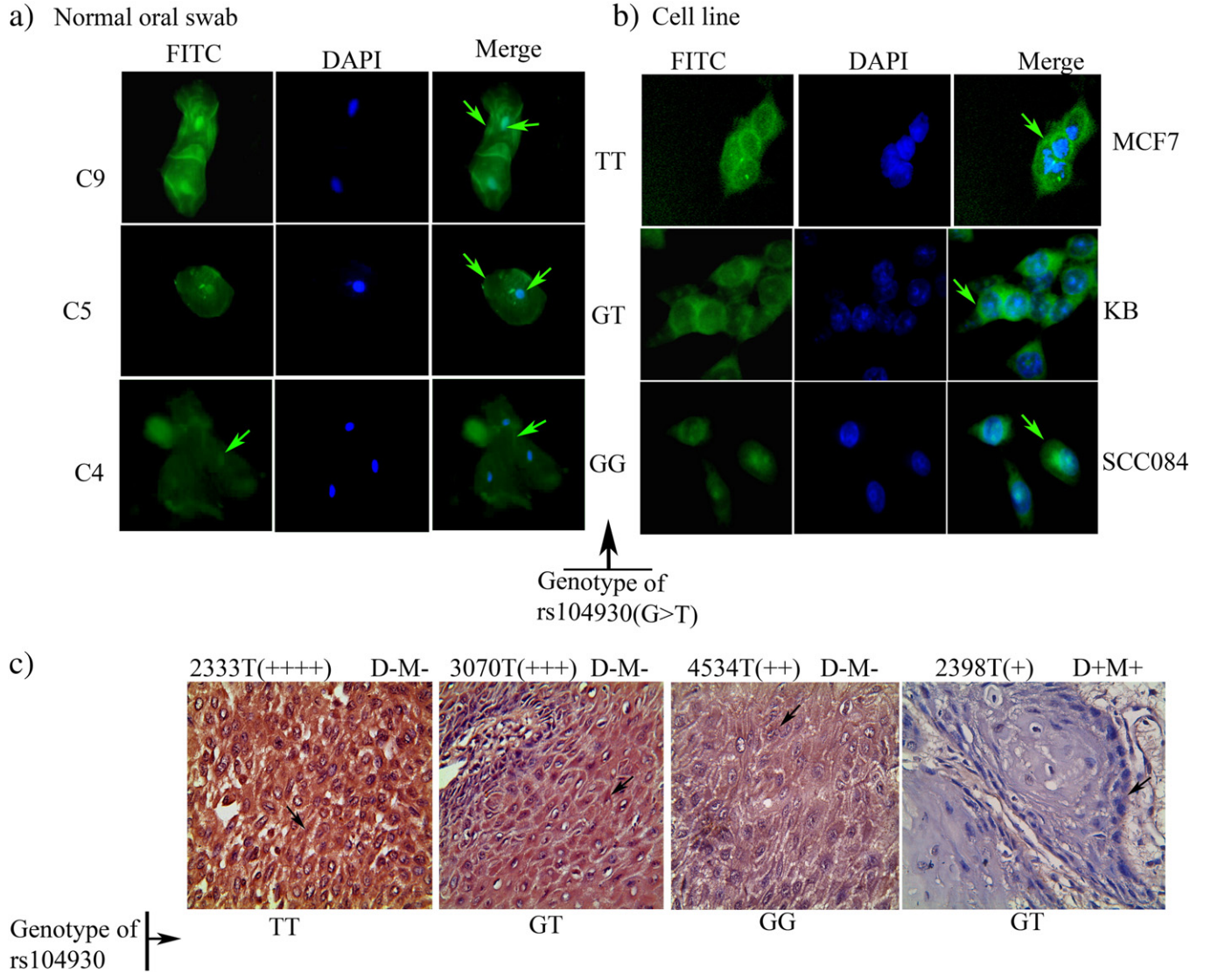


**Fig. 5.** Immunohistochemical/immunocytochemical analysis of SH3GL2 protein. a) Immunocytochemical analysis of SH3GL2 in normal oral swab showing nuclear and cytoplasmic expression. Green arrows indicate the protein expression and localization. C4, C5 and C9 indicate the number of normal swab samples.b) Immunocytochemical analysis in SCC084, KB and MCF7 showing the cytosolic expression of SH3GL2. c) Immunohistochemical analysis of SH3GL2 in primary HNSCC samples showing cytoplasmic expression. Black arrows indicate the protein expression in cell. (+) sign indicates the level of protein expression. 2333T, 3070T, 4534T and 2398T denote the registration number of HNSCC tumor samples. D+M+,gene deletion and promoter methylation present; D−M−,gene deletion and promoter methylation absent.

evident, indicating the presence of a short SH3GL2 transcript in the MCF7 cells in contrast with the Hep2 and KB cells (Fig. 4a). Sequencing analysis of the PCR products resulting from the amplification of the SNP using the primer set 4 & 5 indicated that only the G allele was evident in both the Hep2 and KB cell lines, whereas genomic DNA sequencing of the region showed that the G allele was evident in the Hep2 cells, the G and T alleles were present in the KB cells, and the T allele was evident in the MCF7 cells (Fig. 4b). Thus, it seems that the SH3GL2 mRNA with the SNP containing the T allele had a shorter 3′-UTR region.

To determine whether the SNPs containing the G and T alleles have any role on the transcript length variation of SH3GL2, expression analysis was performed using the primer set 1 & 3 on the cell lines (Fig. 4a). A larger transcript was common for all of the cell lines, whereas smaller transcripts were observed in the Hep2 and KB cell lines, which were probably due to alternative splicing.

### 3.5. Effect of SNP rs1049430 on SH3GL2 protein expression

Immunohistochemical/immunocytochemical analysis revealed the differential expression of SH3GL2 depending upon the allelic variation of the SNP in the normal oral swab, HNSCC samples and cell lines (Fig. 5). In the normal oral swab, high cytoplasmic and nuclear expression of SH3GL2 was observed in cells with the TT genotype. It was expressed at comparatively lower levels in the samples with either the GT or GG genotype. In the cell lines, the expression of SH3GL2 was greater in the MCF7 and KB cells with the TT and GT genotypes, respectively, compared with the Hep2 cells with the GG genotype. Similar results were observed in the HNSCC samples with no genetic/epigenetic alterations in *SH3GL2*. The protein expression level of *SH3GL2* was significantly correlated in the normal ($p = 0.01$) and HNSCC ($p = 0.00032$) patients with the rs1049430 genotypes (Table 3a & b).

## 4. Discussion

The aim of the present study was to identify the role of 3′-UTR SNP(s) in the regulation of SH3GL2 expression and the susceptible SH3GL2 allele(s) associated with HNSCC. To accomplish this aim, we analyzed the genotypes of all SNPs in the 3′-UTR of SH3GL2. We found that only one SNP, rs1049430, which is located in the 3′-UTR, was informative (i.e. heterozygous) in our populations. Our data provide evidence that this new class of regulatory polymorphisms is an important modifier of HNSCC-related patient deaths.

To identify the regulatory mechanism of SNP rs1049430 involved in the occurrence of HNSCC, a case–control study was performed of 764 cases and 530 controls using an Illumina GoldenGate assay platform. The genotyping of the SNP was also conducted in an isolated Indian control population with this method. No association of SNP rs1049430 with HNSCC was observed in the case–control study (Table S5). However, when we compared the HNSCC allele frequencies of the patients in the different populations with those of the isolated Indian control population, significant associations were observed for those in the Rabha and Madia populations (Table S4). These findings indicate that the Rabha and Madia populations might be differentially susceptible to HNSCC compared with the control population. However, further study is needed to explore the population-specific disease susceptibility.

To determine whether there is any correlation of the genotype with disease aggressiveness, allele-specific survival was determined for the patients, and multivariate Cox regression analysis was performed. In this study, we observed that the G allele of rs1049430-carrying patients was more detrimental to survival compared with the T allele (Table 2). Kaplan–Meier overall survival analysis revealed significant (log rank $p = 0.001$) survival advantages for the T allele carriers (Fig. 3). Moreover, survival analysis of the patients with no genetic alterations in SH3GL2 showed significant allele-specific survival advantages, but no significant survival differences were observed for the patients possessing genetic alterations. These findings indicate that in addition to genetic alterations in SH3GL2, SNP rs1049430 plays a role in HNSCC progression. In addition, these results indicate that either genetic alterations in SH3GL2 and/or the G allele of rs1049430 are predictors of poor patient survival. Similar to the findings of our study, 3′-UTR SNPs of several genes, such as MDM4 and ITGB4, have been shown to be associated with the clinical outcomes of different tumors, such as non-small cell lung cancer, ovarian cancer, and breast cancer [3,24,30]. To the best of our knowledge, this is the first report describing an association of the 3′-UTR SNP genotype of SH3GL2 with the clinical outcome of HNSCC.

In our previous study, we have observed that downregulation of SH3GL2 due to genetic alteration is associated with overexpression of EGFR and also with poor patient survival [17]. To determine the cause of the SNP rs1049430-associated differential survival of the HNSCC patient with no alterations in SH3GL2, the mRNA 3′-UTR secondary structure and minimal free energy (MFE) change were calculated using a bioinformatics tool. The predicted MFE of the secondary

**Table 3**
Association of protein expression with genotype of SNP rs1049430 in control healthy population and HNSCC patient population.

a)

| Control samples | Genotype of rs1049430 | Protein level (ICC analysis) |
|---|---|---|
| C1 | GT | ++ |
| C2 | GG | + |
| C3 | GG | + |
| C4 | GG | ++ |
| C5 | GT | + |
| C6 | TT | +++ |
| C7 | GG | + |
| C8 | GT | + |
| C9 | TT | +++ |
| C10 | GG | + |
| C11 | GG | + |
| C12 | GG | + |
| C13 | GT | ++ |
| C14 | GG | ++ |
| C15 | GG | + |
| C16 | GT | ++ |
| C17 | GG | + |
| C18 | TT | +++ |
| C19 | TT | +++ |
| C20 | GG | + |
| $p = 0.01$* | | |

b)

| HNSCC sample reg. no. | Genetic alterations | Genotype of rs1049430 | Protein level (IHC analysis) |
|---|---|---|---|
| 2398 | D +M+ | GG | + |
| 5219 | D +M− | GG | + |
| 1087 | D−M+ | GT | + |
| 2333 | D−M− | GG | ++ |
| 5497 | D−M− | GT | +++ |
| 1048 | D−M− | TT | ++++ |
| 3070 | D−M− | GT | ++ |
| 598 | D−M− | GG | ++ |
| 5398 | D−M− | GG | ++ |
| 914 | D−M− | TT | +++ |
| 3941 | D−M− | GG | ++ |
| 756 | D−M− | GT | +++ |
| 5497 | D−M− | GT | +++ |
| 5114 | D−M− | GG | ++ |
| 2035 | D−M− | GG | ++ |
| 4534 | D−M− | GG | + |
| 5540 | D−M− | GT | +++ |
| 6329 | D−M− | TT | +++ |
| 6907 | D−M− | TT | ++++ |
| 6835 | D−M− | GT | +++ |
| $p = 0.00032$* | | | |

C(1–20), control healthy sample's reg. number; D +/−, deletion of SH3GL2 present/absent; M +/−; promoter methylation of SH3GL2 present/absent.
* Significant level.

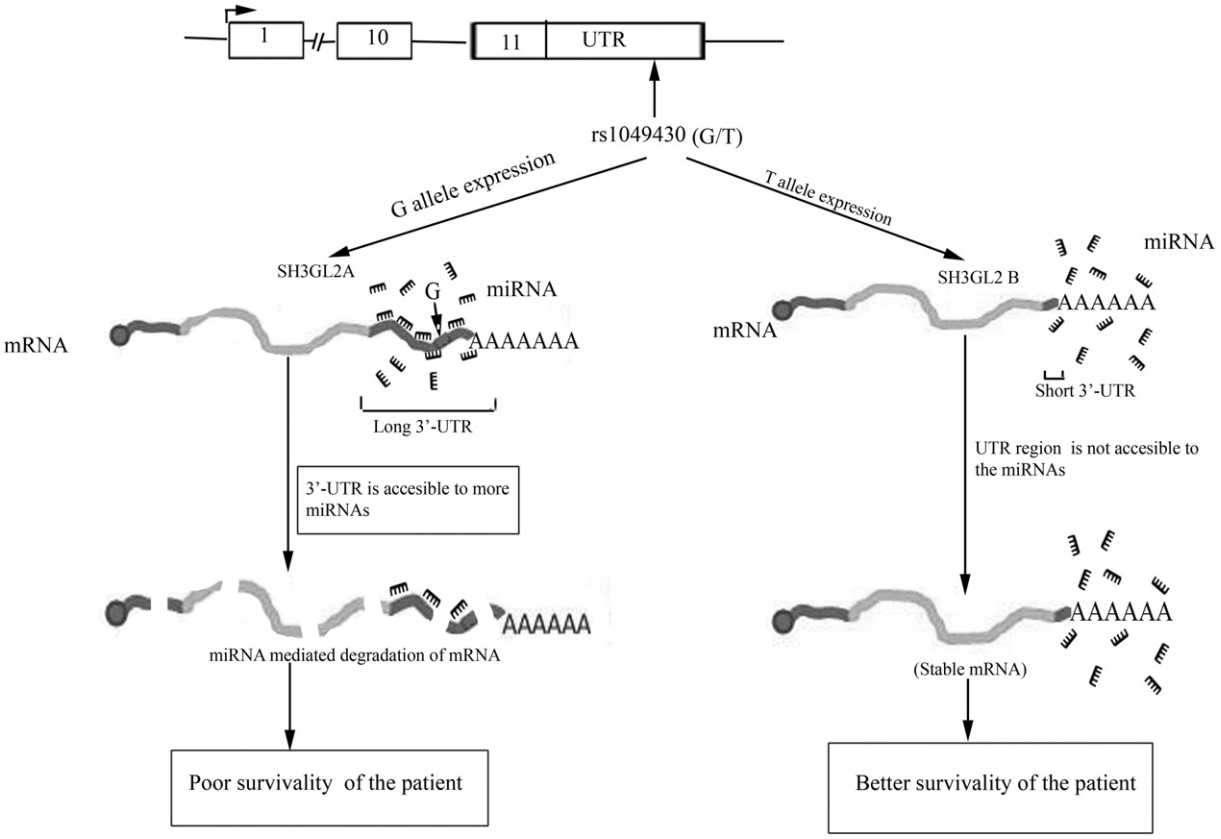


**Fig. 6.** Proposed model for the regulation of SH3GL2 and clinical outcome of the patients: G allele containing mRNA isoforms of SH3GL2 are expressed with longer 3′-UTR as a result of which greater number of miRNA get the accesses to the 3-UTR of SH3GL2. Consequently, miRNA mediated decay of SH3GL2 is enhanced. As the SH3GL2 expression is a predictor of patient survivability, the patients with G allele containing SH3GL2 have poor clinical outcome. On the contrary, T allele of SH3GL2 isoforms is expressed with shorter 3-UTR resulting in lower number of miRNA accession to 3′-UTR of SH3GL2. This occurrence leads to comparatively stable mRNA leading to better survivability of the patients.

structure of the SH3GL2 mRNA with the T allele in rs1049430 was lower than that of the mRNA with the G allele (−27.20 vs. −25.16 kcal/mol) (Fig. S1), indicating the increased stability of the mRNA with the T allele. To support the mRNA stability prediction, an mRNA stability assay was performed using cell lines with homozygous T and G rs1049430 alleles after inhibition of transcription with actinomycin-D treatment (Fig. 3d). We observed that the stability of the T allele-containing mRNA was much higher than that of the G allele-containing mRNA, indicating that a higher amount of SH3GL2 protein may be produced by the T allele-containing mRNA due to its increased stability.

To further explore the possible regulatory mechanism of rs1049430 involved in SH3GL2 expression and in the clinical outcomes of HNSCC patients, an allele expression assay was performed to assess different cell lines with homozygous and heterozygous SNP alleles (Fig. 4). We observed that the G allele contained within a UTR region encompassing the SNP was expressed in cell lines with homozygous (Hep2) or heterozygous (KB) alleles. No T allele was observed in the heterozygous (GT) or homozygous (TT) mRNA in KB or MCF7, respectively. This finding might have been due to the differential termination of SH3GL2 pre-mRNA in the 3′-UTR region. It is possible that the G allele-containing transcript could include the SNP, leading to a longer transcript. In contrast, the T allele-containing transcript did not include the SNP, leading to a shorter transcript. An allele-specific differential 3′-UTR length variation has also been reported in different genes, such as IRF5, PPIL2, and PTER [12]. In addition, the two G allele-containing transcripts and one T allele-containing transcript indicated the differential control of SH3GL2 splicing (Fig. 4a). A similar phenomenon has also been reported in different genes with SNPs in the 3′-UTR, such as L-RAP, CAST, and GPR177 [14]. This finding may have been due to the fact that splicing machinery may be influenced by RNA allele-specific secondary structure and may play an important role in exon definition for particular transcripts [4].

Our data showed that the G allele-containing transcript, i.e. the longer transcript, had a decreased stability compared with the T allele-containing transcript, i.e. the shorter transcript. This finding might have been due to the multiple miRNA binding sites in the longer 3′-UTR regions of SH3GL2, as revealed by bioinformatics analysis using different on-line algorithms (Table S3, Fig. 6). The stability of allele-specific SH3GL2 transcripts was validated in protein expression analysis (Fig. 5), which revealed lower protein expression in association with the longer transcript compared with the shorter transcript.

Thus, the association of rs1049430 with SH3GL2 mRNA processing and stability is important for determining the clinical outcome of HNSCC and the poor prognosis of patients. However, a detailed, in vitro, allele-specific expression profile of the SNP is warranted to understand its importance in the regulation of SH3GL2 expression.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbadis.2015.02.009.

## Conflict of interest

The authors have declared that no conflict of interests exists.

## Acknowledgements

We are thankful to the Director, Chittaranjan National Cancer Institute, Kolkata, India for active encouragement and support during this work. Also, we acknowledge Saimul Islam, Shreya Sarker and Anirban Roychowdhury for suggestions and help during the different experiments. Financial support for this work was provided by grants from

the Department of Biotechnology, Ministry of Science and Technology, Government of India (BT/PR/5524/Med/14/649/2004) to Dr. CK Panda and Dr. S Roychowdhury, Council of Scientific and Industrial Research Project, Government of India (IAP-001) to Dr. S Roychowdhury and CSIR-SRF Fellowship (grant no. 09/30 (0053)2k9-EMR-I) to Dr. GP Maiti.

## References

[1] D. Baralle, M. Baralle, Splicing in action: assessing disease causing sequence changes, J. Med. Genet. 42 (2005) 737–748.
[2] N. Blagitko, S. Mergenthaler, U. Schulz, H.A. Wollmann, W. Craigen, T. Eggermann, H.H. Ropers, V.M. Kalscheuer, Human GRB10 is imprinted and expressed from the paternal and maternal allele in a highly tissue- and isoform-specific fashion, Hum. Mol. Genet. 9 (2000) 1587–1595.
[3] A. Brendle, H. Lei, A. Brandt, R. Johansson, K. Enquist, R. Henriksson, K. Hemminki, P. Lenner, A. Forsti, Polymorphisms in predicted microRNA-binding sites in integrin genes and breast cancer: ITGB4 as prognostic marker, Carcinogenesis 29 (2008) 1394–1399.
[4] E. Buratti, F.E. Baralle, Influence of RNA secondary structure on the pre-mRNA splicing process, Mol. Cell. Biol. 24 (2004) 10505–10514.
[5] S. Chatterjee, J.K. Pal, Role of 5′- and 3′-untranslated regions of mRNAs in human diseases, Biol. Cell. 101 (2009) 251–262.
[6] K. Chen, F. Song, G.A. Calin, Q. Wei, X. Hao, W. Zhang, Polymorphisms in microRNA targets: a gold mine for molecular epidemiology, Carcinogenesis 29 (2008) 1306–1311.
[7] S. Dasgupta, J.S. Jang, C. Shao, N.D. Mukhopadhyay, U.K. Sokhi, S.K. Das, M. Brait, C. Talbot, R.C. Yung, S. Begum, W.H. Westra, M.O. Hoque, et al., SH3GL2 is frequently deleted in non-small cell lung cancer and downregulates tumor growth by modulating EGFR signaling, J. Mol. Med. (Berl.) 91 (2013) 381–393.
[8] I. Dikic, Mechanisms controlling EGF receptor endocytosis and degradation, Biochem. Soc. Trans. 31 (2003) 1178–1181.
[9] W.E. Farrell, D.J. Simpson, J.E. Bicknell, A.J. Talbot, A.S. Bates, R.N. Clayton, Chromosome 9p deletions in invasive and noninvasive nonfunctional pituitary adenomas: the deleted region involves markers outside of the MTS1 and MTS2 genes, Cancer Res. 57 (1997) 2703–2709.
[10] A. Ghosh, S. Ghosh, G.P. Maiti, M.G. Sabbir, N. Alam, N. Sikdar, B. Roy, S. Roychoudhury, C.K. Panda, SH3GL2 and CDKN2A/2B loci are independently altered in early dysplastic lesions of head and neck: correlation with HPV infection and tobacco habit, J. Pathol. 217 (2009) 408–419.
[11] M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, E. Segal, The role of site accessibility in microRNA target recognition, Nat. Genet. 39 (2007) 1278–1284.
[12] T. Kwan, D. Benovoy, C. Dias, S. Gurd, C. Provencher, P. Beaulieu, T.J. Hudson, R. Sladek, J. Majewski, Genome-wide analysis of transcript isoform variation in humans, Nat. Genet. 40 (2008) 225–231.
[13] T. Kwan, E. Grundberg, V. Koka, B. Ge, K.C. Lam, C. Dias, A. Kindmark, H. Mallmin, O. Ljunggren, F. Rivadeneira, K. Estrada, J.B. van Meurs, et al., Tissue effect on genetic control of transcript isoform variation, PLoS Genet. 5 (2009) e1000608.
[14] G. Li, J.H. Bahn, J.H. Lee, G. Peng, Z. Chen, S.F. Nelson, X. Xiao, Identification of allele-specific alternative mRNA processing via transcriptome sequencing, Nucleic Acids Res. 40 (2012) e104.
[15] J. Luo, Q. Cai, W. Wang, H. Huang, H. Zeng, W. He, W. Deng, H. Yu, E. Chan, C.F. Ng, J. Huang, T. Lin, A microRNA-7 binding site polymorphism in HOXB5 leads to differential gene expression in bladder cancer, PLoS One 7 (2012) e40127.
[16] G.P. Maiti, A. Ghosh, P. Mondal, S. Ghosh, J. Chakraborty, A. Roy, S. Roychowdhury, C.K. Panda, Frequent inactivation of SLIT2 and ROBO1 signaling in head and neck lesions: clinical and prognostic implications, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. 119 (2014) 202–212.
[17] G.P. Maiti, P. Mondal, N. Mukherjee, A. Ghosh, S. Ghosh, S. Dey, J. Chakrabarty, A. Roy, J. Biswas, S. Roychoudhury, C.K. Panda, Overexpression of EGFR in head and neck squamous cell carcinoma is associated with inactivation of SH3GL2 and CDC25A genes, PLoS One 8 (2013) e63440.
[18] P.J. Mishra, R. Humeniuk, P.J. Mishra, G.S. Longo-Sorbello, D. Banerjee, J.R. Bertino, A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 13513–13518.
[19] S. Mitra, D. Mazumder Indra, N. Bhattacharya, R.K. Singh, P.S. Basu, R.K. Mondal, A. Roy, E.R. Zabarovsky, S. Roychoudhury, C.K. Panda, RBSP3 is frequently altered in premalignant cervical lesions: clinical and prognostic significance, Genes Chromosom. Cancer 49 (2010) 155–170.
[20] P. Mondal, S. Datta, G.P. Maiti, A. Baral, G.N. Jha, C.K. Panda, S. Chowdhury, S. Ghosh, B. Roy, S. Roychoudhury, Comprehensive SNP scan of DNA repair and DNA damage response genes reveal multiple susceptibility loci conferring risk to tobacco associated leukoplakia and oral cancer, PLoS One 8 (2013) e56952.
[21] M.S. Nicoloso, H. Sun, R. Spizzo, H. Kim, P. Wickramasinghe, M. Shimizu, S.E. Wojcik, J. Ferdin, T. Kunej, L. Xiao, S. Manoukian, G. Secreto, et al., Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility, Cancer Res. 70 (2010) 2789–2798.
[22] G. Papagregoriou, K. Erguler, H. Dweep, K. Voskarides, P. Koupepidou, Y. Athanasiou, A. Pierides, N. Gretz, K.N. Felekkis, C. Deltas, A miR-1207-5p binding site polymorphism abolishes regulation of HBEGF and is associated with disease severity in CFHR5 nephropathy, PLoS One 7 (2012) e31021.
[23] N. Potter, A. Karakoula, K.P. Phipps, W. Harkness, R. Hayward, D.N. Thompson, T.S. Jacques, B. Harding, D.G. Thomas, R.W. Palmer, J. Rees, J. Darling, et al., Genomic deletions correlate with underexpression of novel candidate genes at six loci in pediatric pilocytic astrocytoma, Neoplasia 10 (2008) 757–772.
[24] X. Pu, J.A. Roth, M.A. Hildebrandt, Y. Ye, H. Wei, J.D. Minna, S.M. Lippman, X. Wu, MicroRNA-related genetic variants associated with clinical outcomes in early-stage non-small cell lung cancer patients, Cancer Res. 73 (2013) 1867–1875.
[25] M. Putnik, C. Zhao, J.A. Gustafsson, K. Dahlman-Wright, Effects of two common polymorphisms in the 3′ untranslated regions of estrogen receptor beta on mRNA stability and translatability, BMC Genet. 10 (2009) 55.
[26] A.R. Ramjaun, A. Angers, V. Legendre-Guillemin, X.K. Tong, P.S. McPherson, Endophilin regulates JNK activation through its interaction with the germinal center kinase-like kinase, J. Biol. Chem. 276 (2001) 28913–28919.
[27] Y. Ren, H.W. Xu, F. Davey, M. Taylor, J. Aiton, P. Coote, F. Fang, J. Yao, D. Chen, J.X. Chen, S.D. Yan, F.J. Gunn-Moore, Endophilin I expression is increased in the brains of Alzheimer disease patients, J. Biol. Chem. 283 (2008) 5685–5691.
[28] J.J. Sheu, C.H. Hua, L. Wan, Y.J. Lin, M.T. Lai, H.C. Tseng, N. Jinawath, M.H. Tsai, N.W. Chang, C.F. Lin, C.C. Lin, L.J. Hsieh, et al., Functional genomic analysis identified epidermal growth factor receptor activation as the most common genetic event in oral squamous cell carcinoma, Cancer Res. 69 (2009) 2568–2576.
[29] S. Sinha, N. Chunder, N. Mukherjee, N. Alam, A. Roy, S. Roychoudhury, Panda C. Kumar, Frequent deletion and methylation in SH3GL2 and CDKN2A loci are associated with early- and late-onset breast carcinoma, Ann. Surg. Oncol. 15 (2008) 1070–1080.
[30] J. Wynendaele, A. Bohnke, E. Leucci, S.J. Nielsen, I. Lambertz, S. Hammer, N. Sbrzesny, D. Kubitza, A. Wolf, E. Gradhand, K. Balschun, I. Braicu, et al., An illegitimate microRNA target site within the 3′ UTR of MDM4 affects ovarian cancer progression and chemosensitivity, Cancer Res. 70 (2010) 9641–9649.
[31] X. Zhang, H. Yang, J.J. Lee, E. Kim, S.M. Lippman, F.R. Khuri, M.R. Spitz, R. Lotan, W.K. Hong, X. Wu, MicroRNA-related genetic variations as predictors for risk of second primary tumor and/or recurrence in patients with early-stage head and neck cancer, Carcinogenesis 31 (2010) 2118–2123.