Subpath analysis of each subtype of head and neck cancer based on the regulatory relationship between miRNAs and biological pathways

FENGWEI AN[1], ZHIQIANG ZHANG[2], MING XIA[3] and LIJUN XING[4]

[1]Department of Otorhinolaryngology, Jinan Military General Hospital, Jinan, Shandong 250031;
[2]Department of Neurology, The People's Hospital of Huangdao, Qingdao, Shandong 266400;
[3]Department of Otorhinolaryngology, The Second Hospital of Shandong University, Jinan, Shandong 250031;
[4]Department of Otorhinolaryngology, The People's Hospital of Rushan City, Rushan, Shandong 264500, P.R. China

Received April 9, 2015; Accepted May 25, 2015

DOI: 10.3892/or.2015.4150

Abstract. The aim of the present study was to explore the potential mechanisms involved in each subtype of head and neck squamous cell carcinoma (HNSCC) via subpath analysis and to investigate their relevance in the prevention of HNSCC. Gene expression profiles of GSE6631 and GSE39366 containing 44 and 168 HNSCC samples, respectively, were downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) from samples in GSE6631 and GSE393666 were screened using the Detection of Imbalanced Differential Signal (DIDS) method respectively. DEGs in GSE39366 were matched with the DEGs in GSE6631 and were used to classify the subtypes of HNSCC based on hierarchical clustering analysis. Furthermore, DEGs were separated into different subtypes and then the pathway information was analyzed. The regulated miRNAs for the DEGs in each subtype were analyzed to select the significant subpaths. Totally, 1,095 DEGs from GSE6631 and 2,528 DEGs from GSE39366 were screened. Samples in GSE39366 were separated into four subtypes. Specific genes in each subtype and DEGs in the common gene set involved in a variety of pathways were identified. In addition, the significant miRNA-target-pathway subpath of each subtype of HNSCC and the common gene set of HNSCC were also enriched. Our data suggest that human papillomavirus (HPV) is positively correlated with HNSCC in subtype 2. Several miRNAs (miRLet-7A, miR-1, miR-206, miR-153, miR-519A and miR-506) and their target genes (CYP46A1, BPNT1, MCM7 and COL5A1) are crucial for HNSCC prevention via different pathways and may provide further knowledge of the mechanisms involved in the progression of HNSCC.

Correspondence to: Dr Ming Xia, Department of Otorhinolaryngology, The Second Hospital of Shandong University, 247 Beiyuan Avenue, Jinan, Shandong 250031, P.R. China
E-mail: blinmin@163.com

Key words: head and neck squamous cell carcinoma, differentially expressed genes, subpath, miRNA, biological pathway

Introduction

Head and neck cancer is a broad epithelial malignancy that arises in the paranasal sinuses, nasal cavity, oral cavity, pharynx and larynx (1). Head and neck squamous cell carcinoma (HNSCC), one type of epithelial head and neck cancer, is a heterogeneous disease and the sixth most common form of cancer worldwide (2). The 5-year survival rate of HNSCC patients with stage III or IV is extremely poor (3) and various complications will be faced by HNSCC survivors during their life time (4). Although many studies have been devoted to the exploration of methods for HNSCC diagnosis, prevention and treatment, the mechanisms of HNSCC progression remain largely unknown.

Previous studies have found that a variety of factors such as tobacco, alcohol consumption and human papillomavirus (HPV) are involved in the pathogenesis of HNSCC. For example, tobacco smoking and alcohol consumption affect the prognosis of HNSCC patients (5). HPV is involved in a subgroup of HNSCC, and is crucial for the prognosis and survival of HNSCC patients (6). In addition, increasing evidence has demonstrated that miRNAs and certain pathways play crucial roles in the etiology of HNSCC. For example, the Akt pathway (7) and Snail-RKIP signaling pathway (8) are tumorigenic targets for HNSCC therapeutic intervention. miR-21 and miR-494 have been identified as possible biomarkers for HNSCC via enhancing cell growth (9). Moreover, miR-138 has been identified as a tumor suppressor and serves as a therapeutic target for HNSCC metastasis (10). Recent studies have revealed that the prognosis of HNSCC is largely determined by tumor distribution, stage and histological characteristics at presentation. In addition, the combined treatment of cetuximab and carboplatin is promising for the survival of HNSCC patients at stage III/IV (11). Chung *et al* (12) identified four subclasses of HNSCC with various clinical prognoses based on microarrays. However, due to the high heterogeneous nature of these tumors, the pathogenesis of HNSCC remains unexplained by the traditional subgroups.

Microarray analysis is an effective approach to monitor global alterations in gene expression and to identify significant

subtypes of HNSCC. Kim *et al* (13) identified the HPV status-specific significant gene set using the gene expression profile GSE39366 of HNSCC. In the present study, we used microarray analysis to screen the differentially expressed genes (DEGs) of samples in GSE3361 and GSE39366, respectively, to classify the different subtypes of heterogeneous HNSCC based on the molecular characteristics of the DEGs. Comprehensive bioinformatics was used to enrich the pathway information and miRNA-target-pathway subpath information of genes in each subtype. This study aimed to reveal the molecular mechanisms of heterogeneous HNSCC by subpath analysis and to explore several key biomarkers for the diagnosis or treatment of HNSCC in the different subtypes. This study may provide the basis for future advanced investigation into the clinical use of subtypes for HNSCC treatment.

Materials and methods

Data preprocessing and DEG screening. The gene expression profile of GSE6631 (14), containing 44 paired (from the same patient) samples of HNSCC and normal tissues, was downloaded from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database in the National Center for Biotechnology Information (NCBI) (15) based on the platform of GPL8300 Affymetrix Human Genome U95 ver. 2 array. Patients who received previous treatment (radiotherapy or chemotherapy) for the index tumor or another head and neck primary tumor within the past 5 years were excluded. Another gene expression profile of GSE39366 (16), containing 168 HNSCC samples, was downloaded from the GEO database in NCBI based on the platform of GPL9053 Agilent-UNC-custom-4X44K. The clinical characteristics of the patients included in the present study represent a broad cross-section of patients with HNSCC that is highly representative. Moreover, there was no correlation of tumor subtype with age, gender, race, alcohol use, pack years of smoking or tumor size.

All the CEL files obtained from the two gene expression profiles were preprocessed using the Python procedure (17) and then transformed into gene symbols. The mean expression value was considered as the gene expression value of each gene. The DIDS (Detection of Imbalanced Differential Signal) algorithm (18) was used to screen the DEGs of HNSCC in GSE6631 with $P<0.05$.

In addtion, Limma package in Bioconductor (19) was used to select the DEGs of the HNSCC samples in GSE6631 with $P<0.05$. DIDS algorithm was also used to screen the DEGs of HNSCC in GSE39366 with $P<0.05$.

Identification of HNSCC-associated DEGs and hierarchical clustering analysis. All selected genes from the GSE39366 profile were matched with the screened DEGs using the DIDS algorithm in GSE6631 to select the common DEGs in the two microarrays. The common DEGs were defined as the DEGs that were associated with HNSCC and to distinguish the 138 HNSCC samples into different subtypes based on their molecular characteristics.

In addition, Cluster software (20) was used to cluster the 138 HNSCC samples based on the gene expression values of the selected common DEGs. In addition, the samples were subtyped based on clustering analysis. Heat maps between the gene expression values and samples were generated using TreeView (21).

DEGs in different subtypes. In order to identify the significant DEGs among the different subtypes, we distributed the DEGs in each subtype based on their mean expression values using the following steps: g stands for one gene and score stands for the gene expression value of g in DIDS. If the score was >0, then g was upregulated in the case sample, otherwise, g was downregulated (score <0). The mean expression values for g in m subtypes are shown as $\{x_1, x_2,.....x_i.....x_m\}$. Max and min stand for the maximum and minimum expression values of g, respectively. The interval = max - min. The following formula 1 was used to identify whether g was a specific gene in subtype i or not:

$$U = \max - \min \ (U>0.02)$$

$$\begin{cases} x_i > \max - 0.25U \ (\text{if score} > 0) \\ x_i < \min + 0.25U \ (\text{if score} < 0) \end{cases}$$

Whereas U stands for the distance between the max and min of g in m subtypes, a larger U represents the significance of g among the m subtypes. One gene with this type of U was determined as the specific gene of the significant subtype. Otherwise, one gene was determined as the common gene for the m subtypes.

In addition, the U distribution curve was used to identify the specific genes and common genes (Fig. 1). The max U of the genes was 0.04, while the min U was 0. Genes with $U>0.02$ were determined to be significant specific genes in each subtype, while genes with $U<0.02$ were defined as common genes in the common gene set.

Significant pathways of the specific DEGs. The molecular signatures database (MSigDB) is a collection of annotated gene sets for use with GSEA software (22). The Kyoto Encyclopedia of Genes and Genomes (KEGG) symbols and 186 KEGG pathways were downloaded from the MSigDB database to enrich the significant pathways of the specific DEGs in each subtype using the binomial distribution (23) which were suitable for enrichment analysis of a large scale of genes. The following formula 2 is shown as:

$$P = \sum_{i=M_1}^{M} \binom{M}{i} (P_e)^i (1-P_e)^{M-i}$$

$$P_e = \frac{N1}{N}$$

If there were M genes in one KEGG pathway, then Mi genes were the specific genes. Whereas P stands for the percentage of Mi-specific genes in the total M genes, and P_e is the background frequency; N is the background gene sets (all genes in the gene expression profile), and N1 is the number of specific genes. $P<0.05$ was chosen as the threshold.

Identification of the miRNA-target-pathway subpath. The abnormal expression of genes leads to abnormal pathways in disease; however, these abnormal DEGs are regulated by various miRNAs. We analyzed the relationship between important miRNAs from the miRNA database and target genes from each subtype or common gene set to select the miRNA-target-pathway subpath (Fig. 2).

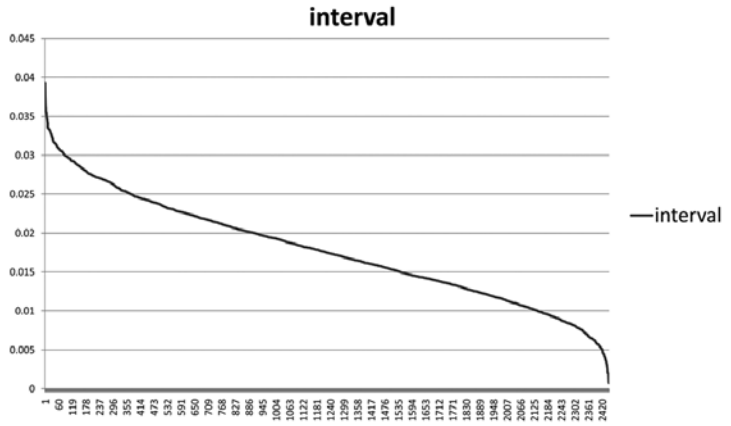

Figure 1. The U distribution of HNSCC-associated DEGs. The x-axis indicates the HNSCC-associated DEGs, the y-axis indicates the U-value of each gene. The line stands for the U distribution curve of all DEGs. HNSCC, head and neck squamous cell carcinoma; DEGs, differentially expressed genes.

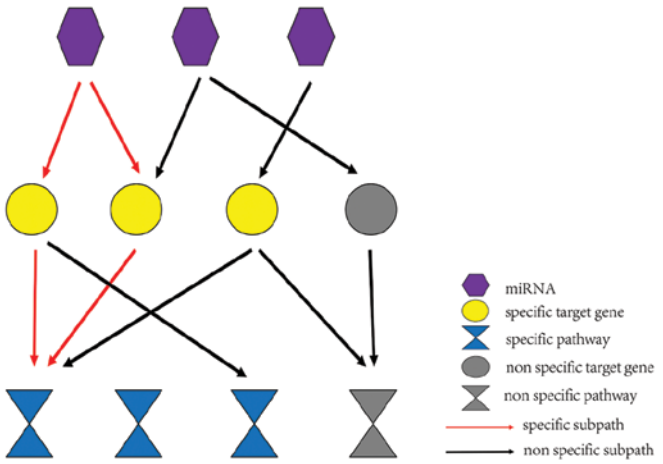

Figure 2. Identification of the significant subpath. Purple polygons represent miRNAs, yellow circles represent the target genes of the miRNAs in the subtypes, and gray circles represent the non-specific target genes. Blue funnels represent the pathways of the specific genes in the subtypes, and gray funnels represents the pathways of the non-specific genes. Red arrows represent the significant subpath (miRNA-target-pathway), and black arrows represent the non-significant subpaths.

The followed algorithm 3 was used to identify the important subpaths:

$$weight1 = 1 + \frac{G}{G^{i}}$$

$$weight2 = 1 + \frac{P}{P^{i}}$$

$$weight3 = 1 + \frac{M}{M^{i}}$$

$$weight = 1 + \frac{P|G|}{P|G^{i}|}$$

Score = log weight1 * weight2 * weight3

Whereas weight1 is the weight for miRNA, G^{i} is the total number of specific genes, G is the number of specific genes that are regulated by miRNA. If specific genes were not regulated by miRNAs, then weight1 = 1. Also, a higher weight1 stands for the close regulatory relationship between miRNA and specific genes.

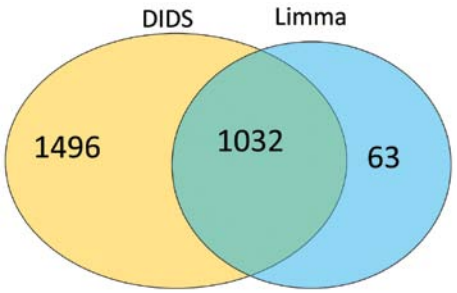

Figure 3. Venn plot of DEGs screened using the DIDS and Limma methods. DEGs, differentially expressed genes.

Weight2 is the weight for target gene, P^{i} is the total number of pathways that all the targets are involved in, and P is the number of pathways that one target is involved in. If the target genes were not involved in any pathway, then weight2 = 1. Moreover, a higher weight2 indicates numerous pathways that one target is involved in.

Weight3 stands for the weight for the pathway, M^{i} is the total number of genes that are involved in this pathway, M is the number of specific genes in this pathway. If there was no specific gene in one pathway, then weight3 = 1. Moreover, a higher weight3 represents a significant correlation between the pathway and subtype.

In addition, weight is the weight of the miRNA-target-pathway, and the score represents the weight for one subpath. A higher score stands for the significant correlation between subpath and subtype. Finally, a subpath where one miRNA regulates several genes and one gene is involved in several pathways was recognized as a significant subpath for HNSCC.

Results

DEG screening. In total, 2,528 DEGs from samples in GSE6631 were screened using DIDS algorithm with P<0.05 (Fig. 3). Moreover, 1,095 DEGs from the HNSCC samples in GSE6631 were screened compared with the normal samples using the traditional Limma in Bioconductor with P<0.05 (Fig. 3).

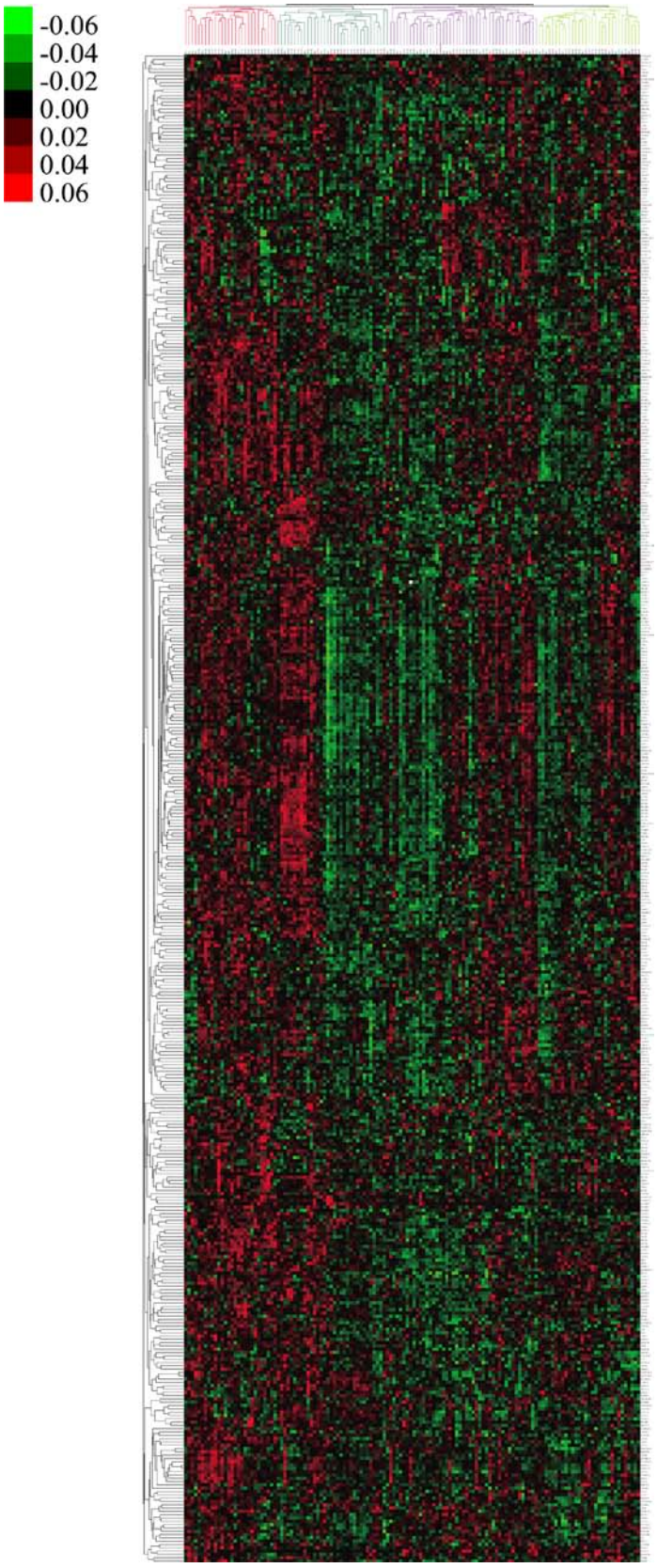

Figure 4. Hierarchical clustering heat map of the subtypes of HNSCC. Horizontal axis indicates the DEGs, vertical axis indicates the sample. Green represents downregulated genes, red represents upregulated genes. HNSCC, head and neck squamous cell carcinoma; DEGs, differentially expressed genes.

In addition, the common DEGs selected using the two methods are shown using a Venn plot. The results showed that 1,032 genes were the common DEGs selected using the two different methods. The 63 genes screened by the Limma method were recognized as DEGs while they were non-DEGs using the DIDS method. The other 1,496 genes were the DEGs

selected by DIDS method while they were non-DEGs using Limma, indicating that 94% of the DEGs in the different groups that were screened using Limma were identified with DIDS. However, the DEGs in one group screened by Limma were not able to be identified by DIDS.

Identification of HNSCC-associated DEGs and hierarchical clustering analysis. In total, 2,443 DEGs from the 138 samples in GSE39366 were screened by matching with the 2,528 DEGs selected using the DIDS algorithm from samples in GSE6631. From the heat maps, the 138 HNSCC samples were clustered into 4 subtypes based on the molecular characteristics of the selected 2,443 DEGs (Fig. 4). In addition, there were 14 samples characterized by HPV and CROI (conference on retroviruses and opportunistic infections). Thus, 10 samples were distributed in subtype 2, while the other 4 were distributed in subtype 1 and 3. Enhance, we speculated that HPV infection might be associated with HNSCC and that it may participant in the pathogenesis of subtype 2.

DEG distribution and pathway enrichment analysis. Based on formula 1, 377 specific DEGs in subtype 1, 46 specific DEGs in subtype 2, 471 specific DEGs in subtype 3 and 451 specific DEGs in subtype 4 were identified. The other 1,539 DEGs, which were not enriched in any subtype acted as the common gene set.

In addition, the enriched significant pathways of DEGs in four subtypes and in the common gene set are shown in Table I. The specific DEGs in subtype 1 were mainly enriched in the metabolism and cancer pathways, such as fatty acid metabolism, melanoma and colorectal cancer (Table IA). Specific DEGs in subtype 2 were mainly enriched in the immune related pathways, such as endocytosis, autoimmune thyroid disease, and antigen processing and presentation (Table IB). Moreover, significant pathways of specific DEGs in subtype 3 included the metabolism pathway and various cancer-related pathways, such as focal adhesion, pathways in cancer, extracellular matrix (ECM) receptor interaction, and ERBB signaling pathway (Table IC). In addiiton, the specific DEGs in subtype 4 mainly function in endocytosis, allograft rejection, cytochrome P450 and leukocyte transendothelial migration pathway (Table ID). In addition, the enriched pathways of DEGs in the common gene set were DNA replication, cell cycle and oocyte meiosis (Table IE).

Identification of significant miRNA-target-pathway subpaths. To identify the important miRNAs that are associated with the selected specific DEGs in the 4 subtypes and with DEGs in the common gene set, the miRNA-target-pathway subpaths with the top 10 scores in the 4 subtypes of HNSCC and in the common gene set were analyzed (Table II). The results showed that miRNA-target-pathway subpaths with the highest score in each subtype or in the common gene set were coincidence with the enriched significant pathways based on the binomial distribution. The top 1 miRNA-target-pathway subpath in subtype 1 was primary bile acid biosynthesis (Table IIA), the top 1 subpath of subtype 1 was sulfur metabolism (Table IIB), and the top 1 subpath of DEGs in subtype 3 was primary bile acid biosynthesis (Table IIC), while the top 1 subpath in subtype 4 was the ECM receptor interaction pathway (Table IID). In

Table I. Enrichment analysis of significant pathways of DEGs.

KEGG pathway	Count	All	P-value
A, Significant pathways of DEGs in subtype 1			
Ribosome	12	88	0.000191635
Primary bile acid biosynthesis	5	16	0.000299713
Valine leucine and isoleucine degradation	8	44	0.000305848
Melanoma	9	71	0.001984553
Fatty acid metabolism	6	42	0.006044122
Renal cell carcinoma	8	70	0.006552473
β alanine metabolism	4	22	0.010278245
T cell receptor signaling pathway	10	108	0.011052662
Colorectal cancer	7	62	0.011382675
Mismatch repair	4	23	0.01205679
Glycerolipid metabolism	6	49	0.012742397
Endometrial cancer	6	52	0.016784945
Epithelial cell signaling in *Helicobacter pylori* infection	7	68	0.018299141
Arachidonic acid metabolism	6	58	0.027344350
Nucleotide excision repair	5	44	0.029755810
Glycolysis gluconeogenesis	6	62	0.036392905
Propanoate metabolism	4	33	0.040794162
Butanoate metabolism	4	34	0.044827467
B, Significant pathways of DEGs in subtype 2			
Viral myocarditis	4	73	0.000461200
Allograft rejection	3	38	0.000847500
Graft vs. host disease	3	42	0.001136800
Type I diabetes mellitus	3	44	0.001302000
Sulfur metabolism	2	13	0.001770500
Autoimmune thyroid disease	3	53	0.002229500
Antigen processing and presentation	3	89	0.009500500
Endocytosis	4	183	0.012604900
Cell adhesion molecules cams	3	134	0.027965800
C, Significant pathways of DEGs in subtype 3			
Focal adhesion	24	201	7.02E-0500
Pathways in cancer	30	328	0.00108771
Fatty acid metabolism	7	42	0.00433402
Valine leucine and isoleucine degradation	7	44	0.00564780
Primary bile acid biosynthesis	4	16	0.00683678
Renal cell carcinoma	9	70	0.00769100
Melanoma	9	71	0.00842790
ECM receptor interaction	10	84	0.00875201
Small cell lung cancer	10	84	0.00875201
ERBB signaling pathway	10	87	0.01110054
Bladder cancer	6	42	0.01690589
Propanoate metabolism	5	33	0.02242731

Table I. Continued.

KEGG pathway	Count	All	P-value
C, Significant pathways of DEGs in subtype 3			
Pancreatic cancer	8	70	0.02253087
Regulation of actin cytoskeleton	18	216	0.02300894
Allograft rejection	5	38	0.03874522
Glioma	7	65	0.04196845
D, Significant pathways of DEGs in subtype 4			
ECM receptor interaction	21	84	2.94E-1000
Focal adhesion	28	201	3.97E-0700
Valine leucine and isoleucine degradation	10	44	3.30E-0500
Fatty acid metabolism	9	42	0.00013232
Histidine metabolism	7	29	0.00033919
Bladder cancer	8	42	0.00072084
Butanoate metabolism	6	34	0.00492207
Small cell lung cancer	10	84	0.00654187
Drug metabolism cytochrome P450	9	72	0.00703527
Glycolysis gluconeogenesis	8	62	0.00892372
Phenylalanine metabolism	4	18	0.00914416
Limonene and pinene degradation	3	10	0.01002415
Arginine and proline metabolism	7	54	0.01369338
Tyrosine metabolism	6	42	0.01391044
β alanine metabolism	4	22	0.01879874
Propanoate metabolism	5	33	0.01896635
Leukocyte transendothelial migration	11	118	0.02530274
Endocytosis	15	183	0.02914549
Allograft rejection	5	38	0.03304028
Tryptophan metabolism	5	40	0.04009753
Pyruvate metabolism	5	40	0.04009753
Steroid biosynthesis	3	17	0.04434620
Linoleic acid metabolism	4	29	0.04701262
Graft vs. host disease	5	42	0.04800130
Metabolism of xenobiotics by cytochrome P450	7	70	0.04848734
E, Significant pathways of DEGs in the common gene set			
DNA replication	16	36	6.28E-0500
Cell cycle	34	128	0.00200400
One carbon pool by folate	8	17	0.00291400
Oocyte meiosis	30	114	0.00410900
Progesterone mediated oocyte maturation	22	86	0.01757000
ERBB signaling pathway	22	87	0.02000400
Base excision repair	10	35	0.04702500

Table I. Continued.

KEGG pathway	Count	All	P-value
E, Significant pathways of DEGs in the common gene set			
Prion diseases	10	35	0.04702500
P53 signaling pathway	17	69	0.04706400

KEGG pathway, the name of the KEGG pathways; Count, the number of specific DEGs in one pathway; All, the total DEGs involved in one pathway. DEGs, differentially expressed genes.

addition, the top 1 subpath in the common gene set was DNA replication and base excision repair (Table IIE). Each miRNA in one subpath regulates several target genes, indicating that the miRNA might be a potential biomarker for HNSCC. Furthermore, target genes participate in a variety of pathways, implying that they may be therapeutic targets.

Discussion

HNSCC is a heterogeneous disease and is the sixth most common form of cancer worldwide (2). Subtype analysis of significant miRNAs and genes associated with HNSCC will be of great significance for individualized clinical treatment and prevention of HNSCC. In the present study, we classified four subtypes of HNSCC and analyzed the subpaths of each subtype of HNSCC based on the regulatory relationship between miRNAs and biological pathways.

Our results revealed that 10 samples of patients infected with HPV were distributed in subtype 2 of HNSCC, indicating that HPV may be associated with HNSCC. HPV has been identified as an etiologic agent for oropharyngeal carcinoma, which is a subset of HNSCC (24). In addition, the majority of HNSCC patients have the potential to be infected with HPV, indicating a significant correlation between HPV and HNSCC (25). Moreover, HPV-positive status affects the therapeutic response and survival of HNSCC patients (26). Based on our data, we suggest that HPV infection may be associated with HNSCC and may increase the risk of HNSCC, which may provide a strategy for HNSCC prevention.

Our study showed that the miRLet-7A-CYP46A1-primary bile acid biosynthesis pathway was the top 1 subpath in subtype 1 and subtype 3 of HNSCC (Table IIA and C). CYP46A1 (cholesterol-24S-hydroxylase) is a CYP450 superfamily enzyme that can convert cholesterol to 24S-hydroxycholesterol (27). Bile acids are derived from cholesterol and synthesis of bile acid is the predominant metabolic pathway for catabolism (28), and certain bile acid metabolism selectively increases cancer risk (29). CYP7A1 (the homologue of CYP46A1) was found to be regulated by the bile acid-activated JNK pathway in primary rat hepatocytes (30). Thus, we speculate that CYP46A1 may function in HNSCC formation via the primary bile acid biosynthesis pathway. On the other hand, Let-7A is a novel biomarker for HNSCC (31). Downregulation of Let-7A has been found in many cancers (32), and it has been reported as a tumor suppressor in head and neck cancer via eliminating

Table II. The enriched significant subpaths (miRNA-target-pathway) with the top 10 scores.

miRNA	Target	KEGG pathway	Score
A, Scores of subpath in subtype 1			
LET-7A	CYP46A1	Primary bile acid biosynthesis	1.463437
miR-98	CYP46A1	Primary bile acid biosynthesis	1.463437
miR-27A	CYP39A1	Primary bile acid biosynthesis	1.448753
miR-27B	CYP39A1	Primary bile acid biosynthesis	1.448753
miR-505	CYP46A1	Primary bile acid biosynthesis	1.399971
miR-9	AUH	Valine leucine and isoleucine degradation	1.326664
miR-29A	DBT	Valine leucine and isoleucine degradation	1.308471
miR-506	BCKDHA	Valine leucine and isoleucine degradation	1.301129
miR-524	HADHB	Valine leucine and isoleucine degradation	1.293750
miR-26A	ACADSB	Valine leucine and isoleucine degradation	1.278876
B, Scores of subpath in subtype 2			
miR-1	BPNT1	Sulfur metabolism	1.267851
miR-153	BPNT1	Sulfur metabolism	1.267851
miR-206	BPNT1	Sulfur metabolism	1.267851
miR-409-3P	BPNT1	Sulfur metabolism	1.206451
miR-144	EIF4G2	Viral myocarditis	1.197256
miR-34B	ICA1	Type I diabetes mellitus	1.186305
miR-200B	NFYA	Antigen processing and presentation	1.168123
miR-200C	NFYA	Antigen processing and presentation	1.168123
miR-429	NFYA	Antigen processing and presentation	1.168123
miR-329	EIF4G1	Viral myocarditis	1.168110
C, Score of subpath in subtype 3			
LET-7A	CYP46A1	Primary bile acid biosynthesis	1.373296
miR-98	CYP46A1	Primary bile acid biosynthesis	1.373296
miR-27A	CYP39A1	Primary bile acid biosynthesis	1.367347
miR-506	ACADVL	Fatty acid metabolism	1.329088
miR-505	CYP46A1	Primary bile acid biosynthesis	1.325001
miR-506	BCKDHA	Valine leucine and isoleucine degradation	1.319689
miR-506	DAPK1	Bladder cancer	1.299341
miR-29A	DBT	Valine leucine and isoleucine degradation	1.276188
miR-9	AUH	Valine leucine and isoleucine degradation	1.276188
miR-524	HADHB	Fatty acid metabolism	1.270789
D, Scores of subpaths in subtype 4			
miR-506	COL5A1	ECM receptor interaction	1.418664
miR-29A	COL6A3	ECM receptor interaction	1.409581
miR-506	BCKDHA	Valine leucine and isoleucine degradation	1.392192
miR-9	CD47	ECM receptor interaction	1.391241
miR-29A	DBT	Valine leucine and isoleucine degradation	1.383109
miR-30E-5P	TNXB	ECM receptor interaction	1.378884
miR-506	ACADVL	Fatty acid metabolism	1.376844
LET-7A	COL1A2	ECM receptor interaction	1.369546
miR-19B	SV2A	ECM receptor interaction	1.369546
miR-27A	RELN	ECM receptor interaction	1.369546

Table II. Continued.

miRNA	Target	KEGG pathway	Score
E, Scores of subpaths in the common gene set			
miR-519A	MCM7	DNA replication	1.580395
miR-506	TDG	Base excision repair	1.425106
miR-124A	PARP1	Base excision repair	1.424206
miR-29A	TDG	Base excision repair	1.420599
miR-30A-5P	TDG	Base excision repair	1.420599
miR-15A	WEE1	Cell cycle	1.414039
miR-16	WEE1	Cell cycle	1.414039
miR-16	CDC25A	Cell cycle	1.414039
miR-16	CHEK1	Cell cycle	1.414039
miR-195	WEE1	Cell cycle	1.414039
miR-195	CDC25A	Cell cycle	1.414039

putative tumor-initiating cells (33). Based on our results, we speculate that miRLet-7A is a tumor suppressor for HNSCC, and regulates the target CYP46A1 via the primary bile acid biosynthesis pathway in subtypes 1 and 3 of HNSCC.

In the present study, our results revealed that the miR-1/miR-153/miR-206-BPNT1-sulfur metabolism pathway was the top 1 subpath in subtype 2 of HNSCC (Table IIB). Downregulation of miR-206 was previously found to contribute to laryngeal cancer proliferation and invasion via regulation of VEGF expression (34), while VEGF was reported to be a predictor for HNSCC (35), implying that miR-206 may be an inhibitor for HNSCC. A previous study found that miR-1 inhibited cell proliferation of HNSCC by targeting TAGLN2 (36), and downregulation of miR-153 promoted tumor metastasis in human epithelial cancer by targeting ZEB2 (37). On the other hand, nucleotide biosynthesis is involved with tumor cell growth during cell proliferation (38). Moreover, nucleotide biosynthesis and protein are coupled by PRPS2 in diving tumorigenesis (39), suggesting that nucleotide biosynthesis may be involved in HNSCC progression. Bisphosphate 3'-nucleotidase 1 (BPNT1) is a member of the magnesium-dependent phosphonoesterase family (40). The roles of BPNT1 in HNSCC have not been fully elucidated. However, BPNT1 plays a role in nucleotide metabolism (41). Hence, we infer that miR-1/miR-153/miR-206 may be promoters in subtype 2 HNSC oncogenesis and metastasis by targeting BPNT1 via the sulfur metabolism pathway.

The miR-506-COL5A1-ECM receptor interaction pathway was found to be the top 1 subpath in subtype 4 HNSCC (Table IID). Studies have revealed that miR-506 affects the EMT process by regulating various EMT-related genes in cancers, such as breast and ovarian cancer (42,43). In additon, EMT is a functional property of HNSCC metastasis (44). Therefore, miR-506 may be a contributor to HNSCC metastasis. A previous study revealed that tumor cells interact with the ECM during tumor invasion or metastasis (45). COL5A1 (collagen type VI) is a collagen family protein that is involved with ECM formation (46). Overexpression of collagen type IV is found to be positively correlated with tumor size in breast cancer (47). Based on our results, we deduce that miR-506 may play crucial roles in the promotion of HNSCC metastasis by regulating COL5A1 via the ECM receptor interaction pathway.

In addition, the miR-519A-MCM7-DNA replication pathway was found to be the top 1 subpath in the common gene set (Table IIE). Studies revealed that miR-519A was downregulated by ΔNp63α in cancers (48), and overexpression of ΔNp63α was beneficial for HNSCC resistance to chemotherapy-induced cell death via inhibiting the transcription of TAp73b (49). One study found that the downregulation of ΔNp63α was a determinant of cellular response to DNA damage in HNSCC (50). Hence, miR-519A may be a tumor suppressor for HNSCC. Abnormal DNA replication promotes the instability of the genome during the early, middle or late S phase (51), while genomic instability plays crucial roles in cancer susceptibility (52), suggesting that DNA replication proteins are involved in HNSCC carcinogenesis. Minichromosome maintenance complex component 7 (MCM7) is a protein belonging to the highly conserved minichromosome maintenance family that are related to DNA replication (53). MCM2 (a member of MCM family) expression is found to be aberrant in laryngeal squamous epithelial lesions (54). In addition, MCM7 is identified as a useful biomarker for HPV-positive HNSCC (55). Our data indicate that miR-519A may act as a therapeutic biomarker for HNSCC by targeting the DNA replication associated protein MCM7.

In summary, our study revealed that there were four subtypes of HNSCC. Several miRNAs were identified as biomarkers or promoters for HNSCC in the different subtypes, followed by their relevant gene targets. miRLet-7A was found to be involved in subtype 1 and subtype 3 by targeting CYP46A1 via the primary bile acid biosynthesis pathway. miR-1/miR-153/miR-206 was found to be involved in subtype 2 with target BPNT1 via the sulfur metabolism pathway, and miR-506 in subtype 4 with COL5A1 via the ECM receptor interaction pathway. In addition, miR-519A was found in the common gene set for HNSCC with MCM7 via the DNA replication pathway. Moreover, HPV may be positively correlated with subtype 2 HNSCC and may increase the risk of HNSCC. Our study may help to explain the specific mechanisms of the HNSCC subtypes and may provide a basis for future investigation of

clinical HNSCC treatments. However, further experimental studies are still needed to confirm our results.

References

1. Argiris A, Karamouzis MV, Raben D and Ferris RL: Head and neck cancer. Lancet 371: 1695-1709, 2008.
2. Hashibe M, Brennan P, Chuang SC, Boccia S, Castellsague X, Chen C, Curado MP, Dal Maso L, Daudt AW, Fabianova E, *et al*: Interaction between tobacco and alcohol use and the risk of head and neck cancer: Pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. Cancer Epidemiol Biomarkers Prev 18: 541-550, 2009.
3. Østhus AA, Aarstad AKH, Olofsson J and Aarstad HJ: Prediction of 5 year survival from level of perceived distress in newly diagnosed head and neck squamous cell carcinoma patients. Oral Oncol 49: 964-969, 2013.
4. Datema FR, Poldermans D and Baatenburg de Jong RJ: Incidence and prediction of major cardiovascular complications in head and neck surgery. Head Neck 32: 1485-1493, 2010.
5. López RV, Zago MA, Eluf-Neto J, Curado MP, Daudt AW, da Silva-Junior WA, Zanette DL, Levi JE, de Carvalho MB, Kowalski LP, *et al*: Education, tobacco smoking, alcohol consumption, and IL-2 and IL-6 gene polymorphisms in the survival of head and neck cancer. Braz J Med Biol Res 44: 1006-1012, 2011.
6. O'Rorke MA, Ellison MV, Murray LJ, Moran M, James J and Anderson LA: Human papillomavirus related head and neck cancer survival: A systematic review and meta-analysis. Oral Oncol 48: 1191-1201, 2012.
7. Moral M and Paramio JM: Akt pathway as a target for therapeutic intervention in HNSCC. Histol Histopathol 23: 1269-1278, 2008.
8. Nieh S, Jao SW, Yang CY, Lin YS, Tseng YH, Liu CL, Lee TY, Liu TY, Chu YH and Chen SF: Regulation of tumor progression via the Snail-RKIP signaling pathway by nicotine exposure in head and neck squamous cell carcinoma. Head Neck: July 2, 2014 (Epub ahead of print).
9. Chang SS, Jiang WW, Smith I, Poeta LM, Begum S, Glazer C, Shan S, Westra W, Sidransky D and Califano JA: microRNA alterations in head and neck squamous cell carcinoma. Int J Cancer 123: 2791-2797, 2008.
10. Liu X, Jiang L, Wang A, Yu J, Shi F and Zhou X: microRNA-138 suppresses invasion and promotes apoptosis in head and neck squamous cell carcinoma cell lines. Cancer Lett 286: 217-222, 2009.
11. Wanebo HJ, Lee J, Burtness BA, Ridge JA, Ghebremichael M, Spencer SA, Psyrri D, Pectasides E, Rimm D, Rosen FR, *et al*: Induction cetuximab, paclitaxel, and carboplatin followed by chemoradiation with cetuximab, paclitaxel, and carboplatin for stage III/IV head and neck squamous cancer: a phase II ECOG-ACRIN trial (E2303). Ann Oncol 25: 2036-2041, 2014.
12. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X, *et al*: Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 5: 489-500, 2004.
13. Kim KY, Zhang X and Cha IH: Identification of human papillomavirus status specific biomarker in head and neck cancer. Head Neck: May 11, 2014 (Epub ahead of print).
14. Kuriakose MA, Chen WT, He ZM, Sikora AG, Zhang P, Zhang ZY, Qiu WL, Hsu DF, McMunn-Coffran C, Brown SM, *et al*: Selection and validation of differentially expressed genes in head and neck cancer. Cell Mol Life Sci 61: 1372-1383, 2004.
15. Pruitt KD, Tatusova T and Maglott DR: NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 33: D501-D504, 2005.
16. Walter V, Yin X, Wilkerson MD, Cabanski CR, Zhao N, Du Y, Ang MK, Hayward MC, Salazar AH, Hoadley KA, *et al*: Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. PLoS One 8: e56823, 2013.
17. Catanzaro B, Garland M and Keutzer K: Copperhead: Compiling an embedded data parallel language. In: Proc. 16th ACM Symposium on Principles and Practice of Parallel Programming, PPoPP '11, ACM, NY, pp47-56, 2011. doi: http://doi.acm.org/10.1145/1941553.1941562.
18. de Ronde JJ, Rigaill G, Rottenberg S, Rodenhuis S and Wessels LF: Identifying subgroup markers in heterogeneous populations. Nucleic Acids Res 41: e200, 2013.
19. Reimers M and Carey VJ: Bioconductor: An open source framework for bioinformatics and computational biology. Methods Enzymol 411: 119-134, 2006.
20. Yu T and Peng H: Hierarchical clustering of high-throughput expression data based on general dependences. IEEE/ACM Trans Comput Biol Bioinformatics 10: 1080-1085, 2013.
21. Guttula SV, Allam A and Gumpeny RS: Analyzing microarray data of Alzheimer's using cluster analysis to identify the biomarker genes. Int J Alzheimers Dis 2012: 649456, 2012.
22. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P and Mesirov JP: Molecular signatures database (MSigDB) 3.0. Bioinformatics 27: 1739-1740, 2011.
23. Hong G, Zhang W, Li H, Shen X and Guo Z: Separate enrichment analysis of pathways for up- and downregulated genes. J R Soc Interface 11: 20130950, 2014.
24. Klussmann JP, Mooren JJ, Lehnen M, Claessen SM, Stenner M, Huebbers CU, Weissenborn SJ, Wedemeyer I, Preuss SF, Straetmans JM, *et al*: Genetic signatures of HPV-related and unrelated oropharyngeal carcinoma and their prognostic implications. Clin Cancer Res 15: 1779-1786, 2009.
25. Michaud DS, Langevin SM, Eliot M, Nelson HH, Pawlita M, McClean MD and Kelsey KT: High-risk HPV types and head and neck cancer. Int J Cancer 135: 1653-1661, 2014.
26. Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, Forastiere A and Gillison ML: Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 100: 261-269, 2008.
27. Ohyama Y, Meaney S, Heverin M, Ekström L, Brafman A, Shafir M, Andersson U, Olin M, Eggertsen G, Diczfalusy U, *et al*: Studies on the transcriptional regulation of cholesterol 24-hydroxylase (CYP46A1): Marked insensitivity toward different regulatory axes. J Biol Chem 281: 3810-3820, 2006.
28. Chiang JY: Bile acids: Regulation of synthesis. J Lipid Res 50: 1955-1966, 2009.
29. McMichael AJ and Potter JD: Host factors in carcinogenesis: certain bile-acid metabolic profiles that selectively increase the risk of proximal colon cancer. J Natl Cancer Inst 75: 185-191, 1985.
30. Gupta S, Stravitz RT, Dent P and Hylemon PB: Down-regulation of cholesterol 7α-hydroxylase (CYP7A1) gene expression by bile acids in primary rat hepatocytes is mediated by the c-Jun N-terminal kinase pathway. J Biol Chem 276: 15816-15822, 2001.
31. Hsu CM, Lin PM, Wang YM, Chen ZJ, Lin SF and Yang MY: Circulating miRNA is a novel marker for head and neck squamous cell carcinoma. Tumour Biol 33: 1933-1942, 2012.
32. Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y, *et al*: Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res 64: 3753-3756, 2004.
33. Yu CC, Chen YW, Chiou GY, Tsai LL, Huang PI, Chang CY, Tseng LM, Chiou SH, Yen SH, Chou MY, *et al*: MicroRNA let-7a represses chemoresistance and tumourigenicity in head and neck cancer via stem-like properties ablation. Oral Oncol 47: 202-210, 2011.
34. Zhang T, Liu M, Wang C, Lin C, Sun Y and Jin D: Down-regulation of miR-206 promotes proliferation and invasion of laryngeal cancer by regulating VEGF expression. Anticancer Res 31: 3859-3863, 2011.
35. Mineta H, Miura K, Ogino T, Takebayashi S, Misawa K, Ueda Y, Suzuki I, Dictor M, Borg A and Wennerberg J: Prognostic value of vascular endothelial growth factor (VEGF) in head and neck squamous cell carcinomas. Br J Cancer 83: 775-781, 2000.
36. Nohata N, Sone Y, Hanazawa T, Fuse M, Kikkawa N, Yoshino H, Chiyomaru T, Kawakami K, Enokida H, Nakagawa M, *et al*: miR-1 as a tumor suppressive microRNA targeting TAGLN2 in head and neck squamous cell carcinoma. Oncotarget 2: 29-42, 2011.
37. Xu Q, Sun Q, Zhang J, Yu J, Chen W and Zhang Z: Downregulation of miR-153 contributes to epithelial-mesenchymal transition and tumor metastasis in human epithelial cancer. Carcinogenesis 34: 539-549, 2012.
38. Tong X, Zhao F and Thompson CB: The molecular determinants of de novo nucleotide biosynthesis in cancer cells. Curr Opin Genet Dev 19: 32-37, 2009.
39. Cunningham JT, Moreno MV, Lodi A, Ronen SM and Ruggero D: Protein and nucleotide biosynthesis are coupled by a single rate-limiting enzyme, PRPS2, to drive cancer. Cell 157: 1088-1103, 2014.

40. Quiroz JA, Gould TD and Manji HK: Molecular effects of lithium. Mol Interv 4: 259-272, 2004.
41. Spiegelberg BD, Xiong JP, Smith JJ, Gu RF and York JD: Cloning and characterization of a mammalian lithium-sensitive bisphosphate 3'-nucleotidase inhibited by inositol 1,4-bisphosphate. J Biol Chem 274: 13619-13628, 1999.
42. Arora H, Qureshi R and Park W-Y: miR-506 regulates epithelial mesenchymal transition in breast cancer cell lines. PLoS One 8: e64273, 2013.
43. Yang D, Sun Y, Hu L, Zheng H, Ji P, Pecot CV, Zhao Y, Reynolds S, Cheng H, Rupaimoole R, *et al*: Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. Cancer Cell 23: 186-199, 2013.
44. Chen C, Wei Y, Hummel M, Hoffmann TK, Gross M, Kaufmann AM and Albers AE: Evidence for epithelial-mesenchymal transition in cancer stem cells of head and neck squamous cell carcinoma. PLoS One 6: e16466, 2011.
45. Stetler-Stevenson WG, Aznavoorian S and Liotta LA: Tumor cell interactions with the extracellular matrix during invasion and metastasis. Annu Rev Cell Biol 9: 541-573, 1993.
46. Burg MA, Tillet E, Timpl R and Stallcup WB: Binding of the NG2 proteoglycan to type VI collagen and other extracellular matrix molecules. J Biol Chem 271: 26110-26116, 1996.
47. Ioachim E, Charchanti A, Briasoulis E, Karavasilis V, Tsanou H, Arvanitis DL, Agnantis NJ and Pavlidis N: Immunohistochemical expression of extracellular matrix components tenascin, fibronectin, collagen type IV and laminin in breast cancer: Their prognostic value and role in tumour invasion and progression. Eur J Cancer 38: 2362-2370, 2002.
48. Huang Y, Chuang A, Hao H, Talbot C, Sen T, Trink B, Sidransky D and Ratovitski E: Phospho-ΔNp63α is a key regulator of the cisplatin-induced microRNAome in cancer cells. Cell Death Differ 18: 1220-1230, 2011.
49. Chung J, Lau J, Cheng LS, Grant RI, Robinson F, Ketela T, Reis PP, Roche O, Kamel-Reid S, Moffat J, *et al*: SATB2 augments ΔNp63α in head and neck squamous cell carcinoma. EMBO Rep 11: 777-783, 2010.
50. Fomenkov A, Zangen R, Huang Y-P, Osada M, Guo Z, Fomenkov T, Trink B, Sidransky D and Ratovitski EA: RACK1 and stratifin target ΔNp63α for a proteasome degradation in head and neck squamous cell carcinoma cells upon DNA damage. Cell Cycle 3: 1285-1295, 2004.
51. Donley N and Thayer MJ: DNA replication timing, genome stability and cancer: Late and/or delayed DNA replication timing is associated with increased genomic instability. Semin Cancer Biol 23: 80-89, 2013.
52. Suzuki M and Takahashi T: Aberrant DNA replication in cancer. Mutat Res 743-744: 111-117, 2013.
53. Lee J-K and Hurwitz J: Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures. Proc Natl Acad Sci USA 98: 54-59, 2001.
54. Chatrath P, Scott IS, Morris LS, Davies RJ, Rushbrook SM, Bird K, Vowler SL, Grant JW, Saeed IT, Howard D, *et al*: Aberrant expression of minichromosome maintenance protein-2 and Ki67 in laryngeal squamous epithelial lesions. Br J Cancer 89: 1048-1054, 2003.
55. Strati K, Pitot HC and Lambert PF: Identification of biomarkers that distinguish human papillomavirus (HPV)-positive versus HPV-negative head and neck cancers in a mouse model. Proc Natl Acad Sci USA 103: 14152-14157, 2006.