Pathology & Oncology Research (2020) 26:2315–2325
https://doi.org/10.1007/s12253-020-00828-w

ORIGINAL ARTICLE

Identification of Novel Candidate Biomarkers for Oral Squamous Cell Carcinoma Based on Whole Gene Expression Profiling

Sudaporn Kengkarn[1] · Songsak Petmitr[1] · Usa Boonyuen[1] · Onrapak Reamtong[1] · Sopee Poomsawat[2] · Sirima Sanguansin[3]

Received: 25 November 2019 / Revised: 23 April 2020 / Accepted: 20 May 2020 / Published online: 29 May 2020

Abstract

This study aimed to determine the whole gene expression profiles and to ascertain potential biomarkers for 22 oral squamous cell carcinoma (OSCC) among Thai patients using the Illumina Human HT-12, V4.0 Expression BeadChip array. Result indicated 2,724 differential expressed genes composed of 1,560 up-regulated and 1,164 down-regulated genes (unpaired t-test, p-value <0.05; fold change ≥2.0 and ≤2.0). The top 9 up-regulated genes were validated in 39 OSCC cases using TaqMan real-time quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) assay. Among these, the up-regulation of peptidase inhibitor 3 (PI3) and keratin 17 (*KRT17*) genes was harbored in all 39 OSCC patients (100%). Likewise, statistical analysis indicated that gene expression in 8 selective genes including keratin 16 (*KRT16*), keratin 14 (*KRT14*), keratinocyte differentiation-associated protein (*KRTDAP*), keratin 6B (*KRT6B*), *PI3*, S100 calcium binding protein A7 (*S100A7*), stratifin (*SFN*) and keratin 5 (*KRT5*) was significantly associated with well differentiated OSCC (p-value <0.05). Moreover, high level of *KRT17* protein was significantly associated with well differentiated OSCC compared to moderately OSCC (p-value = 0.041). Notably, using nested-PCR analysis indicated all OSCC cases in this study were HPV-free. Especially, *KRTDAP*, *PI3*, *SFN* mRNA expression were first reported among patients with OSCC. Conclusion, the whole transcript expression study and TaqMan real-time qRT-PCR assay were relevant regarding the increase in gene expression in OSCC. In addition, the up-regulation of *PI3* and *KRT17* might constitute potential candidate molecular biomarkers to diagnose patients with OSCC.

Keywords Oral squamous cell carcinoma · Gene expression profile · Biomarker · *PI3* · *KRT17*

Background

Oral cancer is one of the most prevalent cancers in the world with approximately 300,000 new diagnosis and 145,000 deaths annually [1]. Among them, oral squamous cell carcinoma (OSCC) represents the highest frequency, more than 90% of oral cancers [2]. OSCC is characterized by an invasive growth pattern, frequent cervical lymph-node metastasis and high recurrence rate. The 5-year survival rate of OSCC is about 63.2 and has not significantly changed over the past 20 years [3]. The development of OSCC is a multi-step process caused by the accumulation of many genetic changes. Therefore, understanding the absolute molecular alterations leading to improved molecular measurements would result in earlier diagnosis, treatment and prognosis of patients with OSCC. Determining diagnostic markers of OSCC is very important in providing diagnosis and treatment plans of individual patients. The first step to explore those candidates is to determine which genes are frequently and strongly expressed in the primary sites of OSCC by complementary DNA microarrays allowing investigation of the whole gene expression profile simultaneously. Studies of gene expression in OSCC using microarray were reported in several ethnic groups such as Americans, Japanese, Chinese and Indian [4–7], but has yet to be investigated among Thai patients.

The aim of this study was to determine the level of whole transcript expression in Thai OSCC using the Illumina Human HT-12, V4.0 Expression BeadChip array, followed by

✉ Sirima Sanguansin
sirima.san@mahidol.ac.th

[1] Department of Molecular Tropical Medicine & Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

[2] Department of Oral and Maxillofacial Pathology, Faculty of Dentistry, Mahidol University, Bangkok, Thailand

[3] Department of Oral Biology, Faculty of Dentistry, Mahidol University, Bangkok, Thailand

validation of the selective 9 up-regulated genes using Taqman real-time quantitative reverse transcriptase PCR (qRT-PCR) assay. These gene expressions may be evaluated as potential biomarkers for OSCC.

Materials and Methods

Tissue Sample Collection

Thirty-nine cases of fresh frozen tissues of OSCC and their corresponding normal oral tissues were collected from patients undergoing surgery at the Ear, Nose and Throat Unit at Rajvithi Hospital. Patients had not received previous radiotherapy or radiation before collecting the tissues. After resection, all specimens were frozen in liquid nitrogen at -80°C before DNA and RNA extraction. The available information of individual patients included sex, age at diagnosis, risk behaviors, stage of tumor, lymph node metastasis and grade of tumor was shown in Table 1. Each section of tissue sample was stained with hematoxylin and eosin to verify pathological status by a pathologist. Tumor classification was staged according to the Tumor-Node-Metastasis classification system of the American Joint Committee on Cancer (AJCC) [8].

This study was approved by the Ethics Committees of Rajvithi Hospital (053/2557), the Faculty of Dentistry and Faculty of Pharmacy, Mahidol University Institutional Review Board (MU-DT/PY-IRB 2012-064.3012) and Faculty of Tropical Medicine, Mahidol University (MUTM 2016-056-02), Thailand. All patients were informed and gave written informed consent.

RNA Extraction

Total RNA was extracted from each tissue sample using TRIzol® reagent (Invitrogen, Carlsbad, CA, USA) following manufacturer instructions. Contamination of genomic DNA in RNA samples were eliminated using DNase I, RNase-free (Thermo Fisher Scientific, IL, USA) according to manufacturer instructions. After that, DNase-treated RNA samples were cleaned using RNeasy® MinElute™ Cleanup Kit (Qiagen, Germany). The first step for quality and quantity of RNA was determined at absorbance 260 and 280 nm by NanoDropTM 1000 (Thermo Fisher Scientific, MA, USA). RNA with 260/280 ratios in the range between 1.8 and 2.1 were used in this study. To confirm the quality of RNA samples, concentration and integrity of RNA was accessed by BioAnalyzer 2100 Agilent with the Agilent RNA 6000 Nano Kit (Agilent Technologies, CA, USA). Quality was evaluated using the RNA Integrity Number (RIN) value that determined for intact 18S and 28S ribosomal peaks. The RNA samples with 18S and 28S ribosomal peaks and RIN between 7 and 10 were used in the microarray experiments.

Microarray Analysis

Whole transcript expression levels of approximately 47,000 target genes were determined using the high-throughput Illumina HumanHT-12, V4.0 Expression BeadChip array. Total RNA from 22 OSCC tissue samples and their corresponding normal oral tissues were used for microarray analysis assays. High quality total RNA was used to generate cDNA. Amplification and labeling of complementary RNA with biotin was performed using the Illumina® TotalPrep™ RNA Amplification Kit (Ambion, TX, USA). Labeled cRNA was hybridized to HumanHT-12, V4 BeadChip arrays (Illumina Inc., CA, USA) followed by staining with streptavidin-Cy3 (Amersham Biosciences, NJ, USA) to visualize. The arrays were washed using Illumina high-stringency wash buffer for 30 min at 55°C followed by scanning

Table 1 Clinicopathologic features of the 39 OSCC samples

Clinical characteristic		Number(%)
Age, years (median, range)		
	≤ 50	9 (23.1%)
	> 50	30 (76.9%)
Sex		
	Female	17 (43.6%)
	Male	22 (56.4%)
Histological grade		
	Well differentiated	28 (71.8%)
	Moderately differentiated	11 (28.2%)
TMN stage		
	II + III	5 (12.8%)
	IV	34 (87.2%)
Lymph node metastasis		
	No	9 (23.1%)
	Yes	30 (76.9%)
Smoking		
	No	12 (30.8%)
	Yes	27 (69.2%)
Alcohol consumption		
	No	18 (46.2%)
	Yes	21 (53.8%)
Betel nut chewing		
	No	29 (74.4%)
	Yes	10 (25.6%)
HPV		
	No	39 (100.0%)
	Yes	0 (0%)

according to standard Illumina protocols. The arrays were scanned using iScan systems. The Bead Array Reader was analyzed using GenomeStudio Software (Illumina Inc., CA, USA).

Determination of Selective Gene Expression Using Taqman Real-Time qRT-PCR

Total RNA from 39 OSCC tissues and their corresponding normal oral tissues were extracted using TRIzol® reagent (Invitrogen, Carlsbad, CA, USA) following manufacturer instructions. Then total RNA was reversed transcribed into cDNA using SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen, Carlsbad, CA, USA) following manufacturer protocol. To validate RNA expression and explore the potential biomarkers for Thai OSCC, the top 9 up-regulated genes including 5 genes in the keratin family: keretin 5 (*KRT5*), keretin 6B (*KRT6B*), keretin 14 (*KRT14*), keretin 16 (*KRT16*) and keretin 17 (*KRT17*)), keratinocyte differentiation-associated protein (*KRTDAP*), peptidase inhibitor 3 (*PI3*), S100 calcium binding protein A7 (*S100A7*) and stratifin (*SFN*) were further analyzed among cDNA of 39 patients with OSCC by Taqman real-time qRT-PCR assay. Taqman real-time qRT-PCR was performed on the LightCycler® 480 real-time PCR (Roche Diagnostics, Mannheim, Germany) using Roche LightCycler® 480 Probes Master (Roche Diagnostics, Mannheim, Germany) according to manufacturer instructions. Each reaction contained 1x Roche Lightcycler TaqMan master mix, specific primer and fluorescently-tagged probe for

Table 2 Primer and Taqman probe sequences of targeted genes validation

Gene symbol	Gene description	Sequence (5′→ 3′)		Product size (bp)
PI3	Elafin or Peptidase inhibitor 3	Forward	CAGTCTCCACTAAGCCTGGC	300
		Reverse	GTGGATGAGAGAGGCAGCTC	
		Probe	HEX-ATCTTGATCCGGTGCGCCATG-BBQ	
KRT17	Keratin 17	Forward	CGGAGACAGAGAACCGCTAC	251
		Reverse	CACAATGGTACGCACCTGAC	
		Probe	HEX–AGGATGCCCACCTGACTCAG TACAA–BBQ	
KRT16	Keratin 16	Forward	AGGGCCAGAGCTCCTAGAAC	153
		Reverse	GCTTTATTAGCCCACCACCA	
		Probe	HEX–CCAGCTATCTCCCCTGCTCCTCT–BBQ	
KRT5	Keratin 5	Forward	GTGAGTACCAGGAGCTCATGAA	139
		Reverse	ACTGCTTGTGACAACAGAGATG	
		Probe	YAK–ATCTCCACGTCCAGGGCCAGCTT–BBQ	
KRT14	Keratin 14	Forward	GGCCTGCTGAGATCAAAGAC	261
		Reverse	GGCTCTCAATCTGCATCTCC	
		Probe	YAK–CACAGTGGACAATGCCAATGTCCT–BBQ	
KRT6B	Keratin 6B	Forward	GGTGGAGGACCTCAAGAACA	296
		Reverse	TCTCCTCATATTGGGCCTTG	
		Probe	HEX–AGGATGTGGATGCTGCCTACATGA–BBQ	
KRTDAP	Keratinocyte differentiation-associated protein	Forward	CTTTAACACCCCGTTCCTGA	191
		Reverse	CTTCCAGTGGAGGTCATGGT	
		Probe	HEX–TCCTTTCCTCAACTGGGATGCCT–BBQ	
S100A7	S100 calcium binding protein A7	Forward	CTGCTGACGATGATGAAGGA	148
		Reverse	CTCCCAGCAAGGACAGAAAC	
		Probe	HEX-AGGGCACAAATTACCTCGCCGAT-BBQ	
SFN	Stratifin	Forward	CAGGCTACTTCTCCCCTCCT	181
		Reverse	TCAATCTCGGTCTTGCACTG	
		Probe	HEX-TGGCTGAGAACTGGACAGTGGCA-BBQ	
GAPDH [a]	Glyceraldehyde-3-phosphate dehydrogenase	Forward	GAAGGCTGGGGCTCATTT	138
		Reverse	CAGGAGGCATTGCTGATGAT	
		Probe	6FAM–CATGTTCGTCATGGGTGTGAACC–BBQ	

[a] *GAPDH* primers were added as internal controls to correct for the differences in different cells

each selective gene was shown in Table 2. The Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene was quantified as the endogenous normalization control. Each case was performed in duplicate.

Detection of HPV

DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen® GmbH, Hilden, Germany) according to the manufacturer protocol. The quality and quantity of DNA was determined at absorbance 260 and 280 nm by NanoDropTM 1000 (Thermo Fisher Scientific, MA, USA). Detection of HPV was performed using nested PCR amplification of a conserved viral L1 gene using MY09 and MY11 and HPV1003 and HPV1004 primers as previously published [9]. The amplification of *GAPDH* gene was examined for internal quality control in all samples. DNA extracted from Hela cells was used as a positive control. All amplified PCR products were resolved by 1% agarose gel electrophoresis stained with RedSafe (iNtRON Biotechnology Inc., South Korea).

Immunohistochemistry Analysis for Keratin 17 Expression

Immunohistochemistry assay was performed on a 4 μm thick section from formalin-fixed paraffin-embedded tissue. The sections were deparaffinized, rehydrated, endogenous peroxidase blocked and antigen was retrieved in microwave with 10 mM citrate buffer pH 6.0 for 15 min followed by treating with 5% bovine serum albumin (Sigma-Aldrich, MO, USA) for 30 min. Then sections were incubated with mouse monoclonal antibody (clone E3) against cytokeratin 17 (CK17) (Dako Denmark A/S, Glostrup, Denmark) at 1:100 dilution for 2 h at room temperature and visualized using labeled polymer (Dako Envision System, Dako Corporation, CA, USA). Negative controls were created by substituting PBS for primary antibody. Microscopic examination was performed by a board certified oral pathologist (SP). The CK17 or KRT17 expression was determined based on the percentage of golden brown staining in cytoplasm of cells, regarding the staining intensity. The scoring was divided into low level (1–25% tumor cell positivity) and high level (≥26% tumor cell positivity).

Microarray Data and Statistical Analysis

All microarray data were analyzed using GeneSpring GX Software (Agilent Technologies, CA, USA). The first step data were normalized and then the expression profiles of OSCC tissues and their corresponding normal oral tissues were performed using the unpaired t-test. Genes with a fold change ≥ 2.0 and ≤ 2.0 (p-value < 0.05) were considered significant. Unsupervised hierarchical clustering analysis was performed to assess correlations between the two groups of samples. The available information of individual patients was analyzed using the Statistical Package for the Social Sciences, Version 14.5 Software (SPSS Inc., IL, USA). The two-sided Pearson chi-square and Fisher's exact test and Odds Ratio were used to analyze the differences in the more than 10 fold up-regulations and clinical characteristics of patients. The relationship between immunohistochemical scoring and clinopathologic features of patients was evaluated using two-tailed Pearson chi-square. A p-value of < 0.05 was considered statistically significant.

Results

Identification of Expressed Genes Related to OSCC

Gene expression levels in OSCC cases were analyzed in comparison with their paired normal control. Approximately 2,724 genes were differently expressed in OSCC. A fold change of ≥ 2.0 (p-value < 0.05) was observed in 1,560 genes and a fold change of ≤ 2.0 (p-value < 0.05) was detected in 1,164 genes. From the unsupervised hierarchical clustering analysis with 10 fold change (p-value < 0.05), the heat map represented 41 genes up-regulated and 130 genes down-regulated in OSCC when compared with the paired normal control (Fig. 1). Among these, *KRT16* was the highest up-regulated at 26.47 fold change whereas Myosin, heavy chain 2 (*MYH2)* was the highest down-regulated at -93.36 fold change. The top 20 up- and down-regulated genes in OSCC are shown in Table 3.

Quantification of Selective Genes Using Taqman Real-Time qRT-PCR

Quantitative real-time reverse transcriptase polymerase chain reaction was performed using Taqman probes to validate the finding of top 9 up-regulated genes in microarray analysis. Among 39 OSCC cases, the frequency of up-regulation in 9 selective genes varied from 92.31 to 100% of cases (Fig. 2). Among these, all 39 cases (100%) harbored *PI3* and *KRT17* gene up-regulations.

HPV Detection

Nested PCR examination using 2 primer sets specific for L1 region of HPV showed negative among all 39 DNA samples extracted from OSCC tissues whereas that of DNA extracted from Hela cells was positive (Fig. 3a). All OSCC DNA was positive for GAPDH prior HPV detection (Fig. 3b).

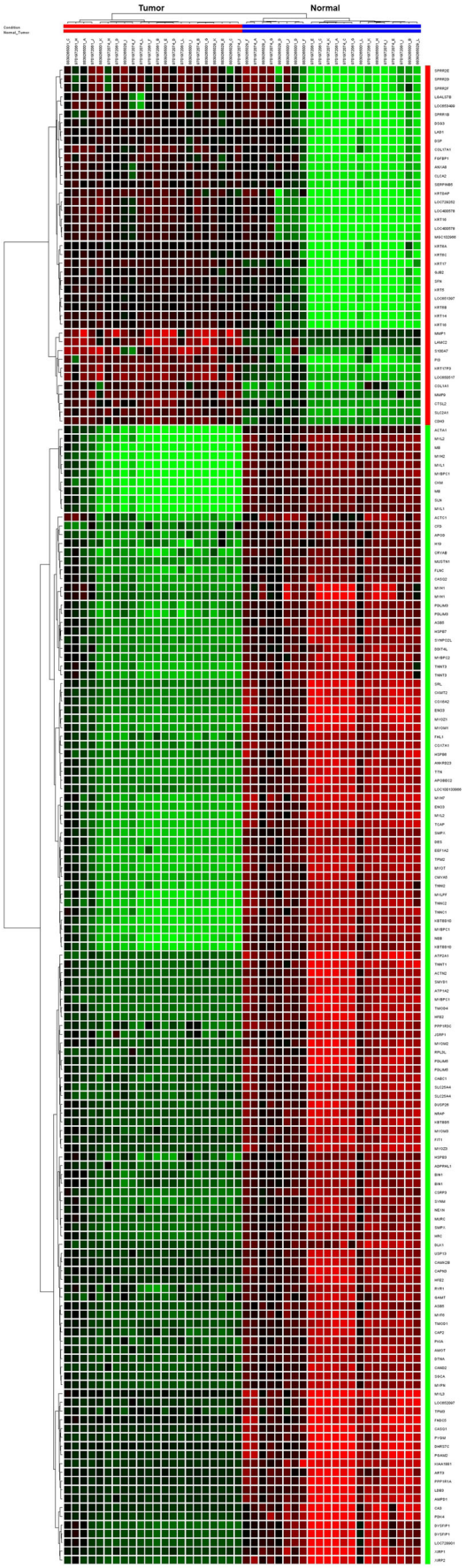

Fig. 1 Hierarchical clustering analysis. The gene expression profile of 22 OSCC cases and their corresponding normal oral tissues revealed 41 up-regulated and 130 down-regulated genes with fold change at cut off ≥ 10 (p-value < 0.05). Columns represent OSCC samples (Tumor) and their corresponding normal oral tissues (Normal), and rows represent the 171 genes that were up-regulated (red) and down-regulated (green) in OSCC

Keratin 17 Protein Expression Using Immunohistochemistry

Protein expression of *KRT17* or cytokeratin 17 was further validated in 39 OSCC tissues and their corresponding normal tissues using immunohistochemical staining. The result demonstrated positive golden brown staining in cellular cytoplasm in all OSCC specimens (100%), whereas no staining was detected in their corresponding normal oral tissues (Fig. 4). Among 28 well differentiated OSCC, 20 cases (71.4%) had high expression and 8 cases (2.86%) had low expression. The expression of KRT17 in 11 moderately differentiated OSCC was high in 4 cases (36.4%) and low in 7 cases (63.6%). The high expression level of KRT17 was significantly correlated with well differentiated OSCC (p-value = 0.041).

The Association Between the Selective Genes Expression with Clinicopathologic Data of Patients with OSCC

Pearson's chi-square, Fisher's exact test and relative risk analysis were performed to evaluate the correlation of more than 10 fold up-regulated genes with clinicopathologic data of patients with primary OSCC including patient's grade of tumor, stage of tumor and lymph node involvement status. A p-value of less than 0.05 was considered significant. The more than 10 fold up-regulations of 8 selective genes, *KRT16, KRT14, KRTDAP, KRT6B, PI3, S100A7, SFN* and *KRT5*, were significantly associated with a well differentiated type of histological grade (p-value = 0.012, 0.005, 0.002, 0.024, 0.009, 0.008, 0.002 and < 0.001 respectively) (Table 4).

Conclusion

The development of OSCC was associated with various environmental factors such as tobacco smoking, alcohol drinking, betel nut chewing and HPV infection. In this study, no HPV infection was detected in this group of patients. HPV prevalence in this study was similar to the related study conducted by Nopmaneepaisarn and colleagues [10]. Under consideration of patient's characteristics, most OSCC cases in this study comprised elderly patients with at least one of either risk factor; smoking, drinking or betel nut chewing. Currently, the microarray technique was used to analyze whole gene expressions in several human cancers such as lung cancer [11],

Table 3 Top 20 up- and down-regulated genes in OSCC (*p-value* < 0.05)

Gene ID	Reference sequence	Gene Symbol	Gene description	Fold change
3868	NM_005557.3	*KRT16*	Keratin 16	26.47
3872	NM_000422.2	*KRT17*	Keratin 17	24.99
3861	NM_000526.4	*KRT14*	Keratin 14	21.78
388533	NM_001244847.1	*KRTDAP*	Keratinocyte differentiation-associated protein	20.74
3854	NM_005555.3	*KRT6B*	Keratin 6B	19.80
5266	NM_002638.3	*PI3*	Peptidase inhibitor 3	19.54
6278	NM_002963.3	*S100A7*	S100 calcium binding protein A7	18.89
2810	NM_006142.3	*SFN*	Stratifin	17.67
3852	NM_000424.3	*KRT5*	Keratin 5	16.69
2706	NM_004004.5	*GJB2*	Gap junction protein, beta 2	15.73
3853	NM_005554.3	*KRT6A*	Keratin 6A	13.63
6513	NM_006516.2	*SLC2A1*	Solute carrier family 2 member 1	13.53
1308	NM_000494.3	*COL17A1*	Collagen type XVII alpha 1 chain	13.39
4312	NM_001145938.1	*MMP1*	Matrix metallopeptidase 1	13.15
6703	NM_002421.3	*SPRR2D*	Small proline rich protein 2D	12.62
1001	NM_001317195.1	*CDH3*	Cadherin 3	12.51
4318	NM_004994.2	*MMP9*	Matrix metallopeptidase 9	12.31
3918	NM_005562.2	*LAMC2*	Laminin subunit gamma 2	12.05
653499	NM_018891.2	*LGALS7B*	Galectin 7B	11.72
1277	NM_001042507.3	*COL1A1*	Collagen type I alpha 1 chain	11.52
4620	NM_001100112.1	*MYH2*	Myosin heavy chain 2	-93.36
4604	NM_001254718.1	*MYBPC1*	Myosin binding protein C, slow type	-84.76
4632	NM_079420.2	*MYL1*	Myosin light chain 1	-84.11
1158	NM_001824.4	*CKM*	Creatine kinase, M-type	-81.89
4625	NM_000257.3	*MYH7*	Myosin heavy chain 7	-77.22
7125	NM_003279.2	*TNNC2*	Troponin C2, fast skeletal type	-76.62
4151	NM_005368.2	*MB*	Myoglobin	-75.60
6588	NM_003063.2	*SLN*	Sarcolipin	-74.67
4633	NM_000432.3	*MYL2*	Myosin light chain 2	-73.89
8557	NM_003673.3	*TCAP*	Titin-cap	-73.53
4604	NM_001254718.1	*MYBPC1*	Myosin binding protein C, slow type	-73.18
29895	NM_001324458.1	*MYLPF*	Myosin light chain, phosphorylatable, fast skeletal muscle	-71.31
4633	NM_001324459.1	*MYL2*	Myosin light chain 2	-69.15
2027	NM_000432.3	*ENO3*	Enolase 3	-65.03
7134	NM_005368.2	*TNNC1*	Troponin C1, slow skeletal and cardiac type	-63.66
7169	NM_203377.1	*TPM2*	Tropomyosin 2	-62.31
10324	NM_203378.1	*KBTBD10*	Kelch like family member 41	-61.76
58529	NM_001193503.1	*MYOZ1*	Myozenin 1	-60.03
9499	NM_001976.4	*MYOT*	Myotilin	-58.27
10324	NM_006063.2	*KBTBD10*	Kelch like family member 41	-56.60

ovarian cancer [12], cholangiocarcinoma [13], breast cancer [14] and oral cancer [4–7]. Studies of gene expression profiles in OSCC using the microarray technique were reported in several ethnic groups. In the present study, the whole transcript expression level among Thai patients with OSCC as compared with paired normal tissues exhibited 1,560 up-regulated and 1,164 down-regulated genes, which was higher than those reported among American [4], Japanese [5], Chinese [6] and Indian [7] populations. This may have been due to improvements in the design technology of the exon array with a greater number of probe sets as part of the new database resulting in higher resolution. Notably, some of the top 20 up-regulated genes in this study were found in related studies in OSCC for several ethnic groups, including *KRT17*

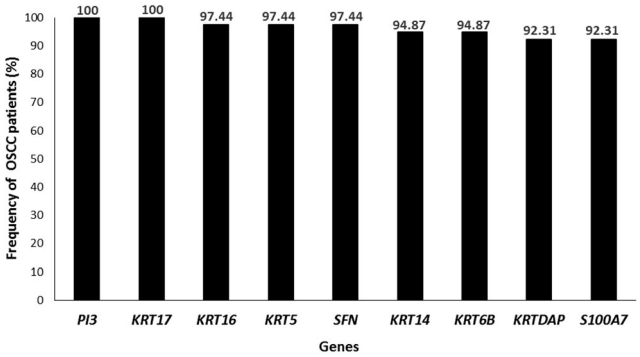

Fig. 2 Frequency of gene expression in OSCC. The mRNAs expression (*KRT17, KRT16, KRT5, KRT14, KRT6B, KRTDAP, PI3, S100A7*, and *SFN*) in 39 patients with OSCC were analyzed by Taqman real-time qRT-PCR using *GAPDH* mRNA as a control

in American and Chinese [4, 6], *KRT16* and *KRT6B* in American [4], matrix metallopeptidase 1 gene (*MMP1*) in Chinese [6] and cadherin 3 gene (*CDH3*) in Japanese [5] populations. Our data agree with the result from microarray analysis on keratin subtypes, in which mRNA expression levels of *KRT5*, *KRT6B*, *KRT14*, *KRT16* and *KRT17* in OSCC as well as oral epithelial dysplasia were higher when compared with normal oral mucosa [15]. While most studies in human cancers aimed to find prospective oncogenic biomarkers and confirmed the differential expression of a subset of selective genes using qRT-PCR, our study used the Taqman probe for real-time qRT-PCR to increase specificity during the hybridization step in PCR as the primer dimer would be compromised. In the present study, the top 9 up-regulated tumor-related genes, including *KRT16, KRT14, KRTDAP, KRT6B, PI3, S100A7, SFN* and *KRT5* from microarray analysis were validated by Taqman real-time qRT-PCR. The results of Taqman qRT-PCR were in total concordance with those of the microarray. Interestingly, elevated *PI3* and *KRT17* expression revealed 100% in all OSCC (39 of 39 cases), while up-regulation in 8 validated genes showed significant association with well differentiated OSCC with the highest association in *KRT5* ($p < 0.001$). Additionally, more than 10 fold elevated expression in these genes exhibited no significant association with any risk factors including, smoking, drinking, and betel nut chewing. Among these, *KRTDAP, PI3, SFN* mRNA expression had not been reported in OSCC cases.

The *PI3* (peptidase inhibitor 3 or elafin) gene encoding an elastase-specific inhibitor, plays a direct role in suppressing tumors by inhibiting elastase and thus serves as a prognostic indicator for patients with breast cancer [16]. Moreover, the overexpression of elafin protein is associated with poor survival in breast cancer [17]. The previous study using immunohistochemistry found the presence of elafin in head and neck squamous cell carcinomas and significantly higher amounts in well differentiated and moderately differentiated tumors than in poorly differentiated tumors [18]. In this study, up-regulation of *PI3* mRNA expression also significantly

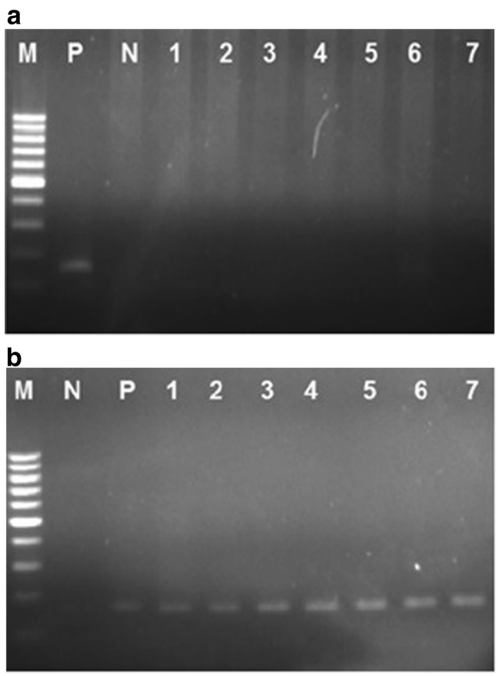

Fig. 3 PCR pattern in 1% agarose gel electrophoresis for HPV detection. **a** Nested PCR products of HPV DNA from OSCC patients. **b** PCR products of *GAPDH* gene. Lane M, 100 bp molecular weight marker; Lane N, Negative control; Lane P, Hela DNA (positive control); Lane 1–7, Representative OSCC samples

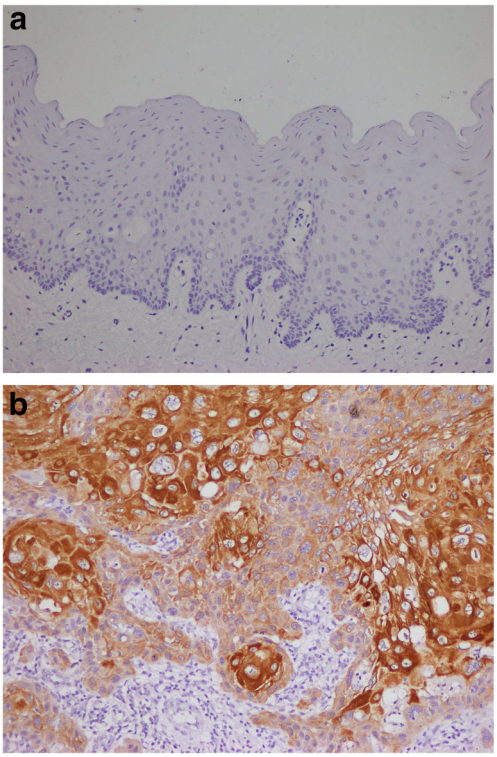

Fig. 4 Immunohistochemical staining of KRT17 (magnification x200). **a** Negative staining in corresponding normal tissue. **b** Positive cytoplasmic staining score 4 in OSCC tissue

Table 4 Association between more than 10 fold up-regulation among 9 up-regulated genes and clinicopathologic data

Variables	n	*PI3*		*KRT17*		*KRT16*		*KRT5*	
		Odds ratio (95% CI)	*p-value*	Odds ratio (95% CI)	*p-value*	Odds ratio (95% CI)	*p-value*	Odds ratio (95% CI)	*p-value*
TMN stage									
II+III	5								
IV	34	0.40 (0.02-8.15)	0.550	8.25 (0.44-159.16)	0.162	0.33 (0.02-6.73)	0.474	0.81 (0.08-8.35)	0.861
Node metastasis									
No metastasis	9								
Metastasis	30	0.20 (0.01-3.88)	0.286	3.63 (0.20-64.59)	0.381	0.17 (0.01-3.18)	0.233	0.83 (0.60-1.13)	0.233
Histological grade									
Moderately differentiated	11								
Well differentiated	28	22.50 (2.21-229.45)	**0.009**	2.70 (0.15-47.39)	0.497	10.83 (1.68-69.92)	**0.012**	22.75 (3.43-150.82)	**<0.001**

Variables	*KRT14*		*KRT6B*		*KRTDAP*		*S100A7*		*SFN*	
	Odds ratio (95% CI)	*p-value*	Odds ratio (95% CI)	*p-value*	Odds ratio (95% CI)	*p-value*	Odds ratio (95% CI)	*p-value*	Odds ratio (95% CI)	*p-value*
TMN stage										
II+III										
IV	1.85 (0.27-12.95)	0.535	0.28 (0.01-5.67)	0.410	0.21 (0.01-4.19)	0.308	0.96 (0.09-10.05)	0.976	0.69 (0.07-7.07)	0.758
Node metastasis										
No metastasis										
Metastasis	0.67 (0.12-3.86)	0.651	0.14 (0.01-2.67)	0.191	0.10 (0.01-1.94)	0.129	0.41 (0.04-3.88)	0.437	0.29 (0.03-2.69)	0.277
Histological grade										
Moderately differentiated										
Well differentiated	10.50 (2.08-53.14)	**0.005**	2.25 (1.29-37.49)	**0.024**	14.58 (2.62-81.09)	**0.002**	13.00 (1.95-86.81)	**0.008**	14.58 (2.62-81.09)	**0.002**

Bold value indicates a statistically significant difference with *p-value* less than 0.05

correlated with patients with well differentiated OSCC and elevated expression was exhibited in 100% of studied cases suggesting that it may serve as a diagnostic biomarker for OSCC. From our knowledge, we here first report the novel *PI3* mRNA expression in all OSCC cases.

Another group of genes with high elevated expression in the present study was *KRTDAP*, *S100A7*, and *SFN*, which are involved in regulating various cell processes. The *KRTDAP* gene encodes keratinocyte differentiation-associated protein (KRTDAP) functions in regulating keratinocyte differentiation and maintaining stratified epithelia [19]. The expression of the *KRTDAP* gene is limited to epithelial tissue at a high level. One related study found that *KRTDAP* mRNA expression was up-regulated among patients with human papilloma virus-positive head and neck squamous cell carcinoma who did not respond to radiation with or without chemotherapeutic agents [20]. The *S100A7* gene encodes for psoriasin, a calcium-binding protein, involved in regulating many cell processes, for example, cell cycle progression and differentiation. Gene expression levels of *S100A7* kept increasing during different stages of malignancy from healthy gingiva, irritation fibroma and leukoplakia, through OSCC [21]. Elevated mRNA levels as well as protein expression of *S100A7* were significantly associated with well differentiated OSCC, uterine cervical SCC and lingual SCC [22, 23]. The *SFN* gene encodes for stratifin, a tumor suppressor protein 14-3-3 sigma [24]. Stratifin serves as a negative regulator of cell cycle progression and reduces cell growth by causing G2 arrest. Down-regulation of *SFN* expression was reported in esophageal squamous cell carcinoma and breast cancer with poor prognosis [25, 26]. In contrast, increased expression of the stratifin protein was observed in lung invasive adenocarcinoma with poor outcome [27]. The present study revealed that up-regulation in mRNA expression of *KRTDAP*, *S100A7*, and *SFN* were significantly associated with well-differentiated tumors.

Keratin is a large family of intermediate filament proteins. Various keratin compositions are expressed depending on epithelial cell type and differentiated status [15]. The present study detected mRNA up-regulation in 5 keratin subtypes; *KRT5*, *KRT6B*, *KRT14*, *KRT16*, *KRT17*. Interestingly, both *KRT17* mRNA and KRT17 protein expression was detected in all OSCC cases. Likewise, high level of KRT17 protein expression was significantly correlated with well differentiated OSCC compared to moderately OSCC. Our results was similar to previous reports indicating that *KRT17* mRNA expression was significantly high among patients with OSCC at early stages, among negative neck nodes metastasis and among well differentiated tumors [28] and KRT17 protein significantly expressed in well differentiated OSCC when compared with moderately and poorly differentiated OSCC [29]. Moreover, the high expression of *KRT17* mRNA and its protein have been described as related to poor clinical outcomes in lung adenocarcinoma [30] and epithelial ovarian cancer [31]. Furthermore, from our knowledge, this may be the first report of *KRT17* mRNA and KRT17 protein expression in all OSCC cases. Several lines of evidence indicated up-regulated expression among the other keratin subtypes in various tumors but their significance in clinicopathology revealed conflicts. The increased *KRT5* mRNA expression and keratin 5 protein overexpression were significantly correlated with poor prognosis and disease recurrence of serous ovarian cancer [32] while increased expression of *KRT5* mRNA was associated with a low prevalence of lymphovascular invasion and lymph node metastasis in muscle-invasive urothelial bladder cancer [33]. High mRNA expression levels of *KRT6B* was determined in infiltrating bladder urothelial carcinoma tissues as well in lung metastasis tissues and correlated with poor prognosis [34]. The up-regulation of *KRT14* mRNA and overexpression of this protein were presented in OSCC when compared with non-tumorous oral epithelial cells [35], while up-regulations of *KRT16* were reported in oral epithelial dysplasia and were significantly associated with OSCC cases [15].

In summary, we first analyzed the genome-wide gene expression profile of Thai patients with OSCC, using the Illumina Human HT-12 v4.0 Expression BeadChip array and validated the 9 up-regulated gene expression using Taqman real-time qRT-PCR. The result revealed elevated expression of *PI3* and *KRT17* might serve as potential biomarkers to diagnose OSCC among Thai patients. Likewise, up-regulation of *KRTDAP, PI3, SFN* mRNAs among patients with OSCC was first investigated here. This study focused on investigating whole transcription profile and validation of the candidate gene expression that might serve as molecular biomarkers for patients with OSCC. Consequently, the role of these candidate proteins in OSCC development will be further studied.

Acknowledgments We thank Dr. Porn-ake Apipant and Dr. Witamol Rumpungsuk (the ENT Unit, Rajvithi Hospital) for providing tissue samples and clinical data, Associate Professor Siribang-on Piboonniyom Khovidhunkit for technical support on HPV analysis and Mr. Thomas McManamon for editing this manuscript.

Funding Information This study was financially supported by the Mahidol University Research Grant 2557–2559 (SS).

Compliance with Ethical Standards

Ethics Approval All the procedures were conducted after receiving approval by the Ethics Committees of Rajavithi Hospital (053/2557), the Faculty of Dentistry and Faculty of Pharmacy, Mahidol University Institutional Review Board (MU-DT/PY-IRB 2012-064.3012) and Faculty of Tropical Medicine, Mahidol University (MUTM 2016-056-02), Thailand in compliance with the ethics standards of the Helsinki Declaration. All patients were informed and freely provided written informed consent.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest All authors declare no conflict of interest.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2015) Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136(5):E359–E386. https://doi.org/10.1002/ijc.29210
2. Brinkman BMN, Wong DT (2006) Disease mechanism and biomarkers of oral squamous cell carcinoma. Curr Opin Oncol 18(3): 228–233. https://doi.org/10.1097/01.cco.0000219250.15041
3. National Cancer Institute (2015) SEER Cancer Statistics Review, 1975–2012 http://seer.cancer.gov/statfacts/html/oralcav.html. Accessed 8 March 2016
4. Méndez E, Cheng C, Farwell DG, Ricks S, Agoff SN, Futran ND, Weymuller EA Jr, Maronian NC, Zhao LP, Chen C (2002) Transcriptional expression profiles of oral squamous cell carcinomas. Cancer 95(7):1482–1494. https://doi.org/10.1002/cncr.10875
5. Tomioka H, Morita K, Hasegawa S, Omura K (2006) Gene expression analysis by cDNA microarray in oral squamous cell carcinoma. J Oral Pathol Med 35(4):206–211. https://doi.org/10.1111/j.1600-0714.2006.00410x
6. Wang YH, Tian X, Liu OS, Fang XD, Quan HZ, Xie S, Gao S, Tang ZG (2014) Gene profiling analysis for patients with oral verrucous carcinoma and oral squamous cell carcinoma. Int J Clin Exp Med 7(7):1845–1852
7. Chakrabarti S, Multani S, Dabholkar J, Saranath D (2015) Whole genome expression profiling in chewing-tobacco-associated oral cancers: a pilot study. Med Oncol 32(3):60. https://doi.org/10.1007/s12032-015-0483-4
8. Amin MB, Edge S, Greene F et al (2017) AJCC Cancer Staging Manual, 8th edn. Springer, New York
9. Khovidhunkit SP, Buajeeb W, Sanguansin S, Poomsawat S, Weerapradist W (2008) Detection of Human Papillomavirus in Oral Squamous Cell Carcinoma, Leukoplakia and Lichen Planus in Thai Patients. Asian Pac J Cancer Prev 9(4):771–775
10. Nopmaneepaisarn T, Tangjaturonrasme N, Rawangban W, Vinayanuwattikun C, Keelawat S, Bychkov A (2019) Low Prevalence of p16-positive HPV-related Head-Neck Cancers in Thailand: Tertiary Referral Center Experience. BMC Cancer 19(1):1050. https://doi.org/10.1186/s12885-019-6266-0
11. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, Meyerson M (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci USA 98(24):13790–13795. https://doi.org/10.1073/pnas.191502998
12. Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockhart DJ, Burger RA, Hampton GM (2001) Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci USA 98(3):1176–1181. https://doi.org/10.1073/pnas.98.3.1176
13. Subrungruang I, Thawornkuno C, Chawalitchewinkoon-Petmitr P, Pairojkul C, Wongkham S, Petmitr S (2013) Gene expression profiling of intrahepatic cholangiocarcinoma. Asian Pac J Cancer Prev 14(1):557–563. https://doi.org/10.7314/Apjcp.2013.14.1.557
14. Arnutti P, Kotepui M, Asanprakit W, Punyarit P, Chavalitshewinkoon-Petmitr P, Harnroongroj T, Petmitr S (2013) Determination of whole transcription profiles and specific pathways in invasive ductal breast carcinoma. Int J Clin Exp Patho 6(6):1112–1120
15. Sakamoto K, Aragaki T, Morita K, Kawachi H, Kayamori K, Nakanishi S, Omura K, Miki Y, Okada N, Katsube K, Takizawa T, Yamaguchi A (2011) Down-regulation of keratin 4 and keratin 13 expression in oral squamous cell carcinoma and epithelial dysplasia: a clue for histopathogenesis. Histopathol 58(4):531–542. https://doi.org/10.1111/j.1365-2559.2011.03759.x
16. Hunt KK, Wingate H, Yokota T, Liu Y, Mills GB, Zhang F, Fang B, Su CH, Zhang M, Yi M, Keyomarsi K (2013) Elafin, an inhibitor of elastase, is a prognostic indicator in breast cancer. Breast Cancer Res 15(1):R3. https://doi.org/10.1186/bcr3374
17. Labidi-Galy SI, Clauss A, Ng V, Duraisamy S, Elias KM, Piao HY, Bilal E, Davidowitz RA, Lu Y, Badalian-Very G, Gyorffy B, Kang UB, Ficarro S, Ganesan S, Mills GB, Marto JA, Drapkin R (2015) Elafin drives poor outcome in high-grade serous ovarian cancers and basal-like breast tumors. Oncogene 34(3):373–383. https://doi.org/10.1038/onc.2013.562
18. Westin U, Nystrom M, Ljungcrantz I, Eriksson B, Ohlsson K (2002) The presence of elafin, SLPI, IL1-RA and STNFalpha RI in head and neck squamous cell carcinomas and their relation to the degree of tumour differentiation. Mediators Inflamm 11(1):7–12. https://doi.org/10.1080/09629350210304
19. Tsuchida S, Bonkobara M, McMillan JR, Akiyama M, Yudate T, Aragane Y, Tezuka T, Shimizu H, Cruz PD, Ariizumi K (2004) Characterization of Kdap, a protein secreted by keratinocytes. J Invest Dermatol 122(5):1225–1234. https://doi.org/10.1111/j.0022-202X.2004.22511x
20. Thibodeau BJ, Geddes TJ, Fortier LE, Ahmed S, Pruetz BL, Wobb J, Chen P, Wilson GD, Akervall JA (2015) Gene Expression Characterization of HPV Positive Head and Neck Cancer to Predict Response to Chemoradiation. Head Neck Pathol 9(3): 345–353. https://doi.org/10.1007/s12105-014-0597-6
21. Probstmeier R, Kraus D, Wenghoefer M, Winter J (2019) S100 Proteins as Biomarkers in Risk Estimations for Malignant Transformation in Oral Lesions. Methods Mol Biol 1929:763 – 71. https://doi.org/10.1007/978-1-4939-9030-6_48
22. Kesting MR, Sudhoff H, Hasler RJ, Nieberler M, Pautke C, Wolff KD, Wagenpfeil S, Al-Benna S, Jacobsen F, Steinstraesser L (2009) Psoriasin (S100A7) up-regulation in oral squamous cell carcinoma and its relation to clinicopathologic features. Oral Oncol 45(8):731–6. https://doi.org/10.1016/j.oraloncology.2008.11.012
23. Kong F, Li Y, Hu E, Wang R, Wang J, Liu J, Zhang J, He D, Xiao X (2016) The Characteristic of S100A7 Induction by the Hippo-YAP Pathway in Cervical and Glossopharyngeal Squamous Cell Carcinoma. PLoS One 11(12):e0167080. https://doi.org/10.1371/journal.pone.0167080
24. Chen J, He QY, Yuen AP, Chiu JF (2004) Proteomics of buccal squamous cell carcinoma: the involvement of multiple pathways in tumorigenesis. Proteomics 4(8):2465–2475. https://doi.org/10.1002/pmic.200300762
25. Ren HZ, Pan GQ, Wang JS, Wen JF, Wang KS, Luo GQ, Shan XZ (2010) Reduced stratifin expression can serve as an independent prognostic factor for poor survival in patients with esophageal squamous cell carcinoma. Dig Dis Sci 55(9):2552–2560. https://doi.org/10.1007/s10620-009-1065-0
26. Mirza S, Sharma G, Parshad R, Srivastava A, Gupta SD, Ralhan R (2010) Clinical significance of Stratifin, ERalpha and PR promoter methylation in tumor and serum DNA in Indian breast cancer patients. Clin Biochem 43(4–5):380–386. https://doi.org/10.1016/j.clinbiochem.2009.11.016
27. Husni RE, Shiba-Ishii A, Nakagawa T, Dai T, Kim Y, Hong J, Sakashita S, Sakamoto N, Sato Y, Noguchi M (2019) DNA

hypomethylation-related overexpression of SFN, GORASP2 and ZYG11A is a novel prognostic biomarker for early stage lung adenocarcinoma. Oncotarget 10(17):1625–1636. https://doi.org/10.18632/oncotarget.26676

28. Toyoshima T, Koch F, Kaemmerer P, Vairaktaris E, Al-Nawas B, Wagner W (2009) Expression of cytokeratin 17 mRNA in oral squamous cell carcinoma cells obtained by brush biopsy: preliminary results. J Oral Pathol Med 38(6):530–534. https://doi.org/10.1111/j.1600-0714.2009.00748x
29. Kitamura R, Toyoshima T, Tanaka H, Kawano S, Kiyosue T, Matsubara R, Goto Y, Hirano M, Oobu K, Nakamura S (2012) Association of cytokeratin 17 expression with differentiation in oral squamous cell carcinoma. J Cancer Res Clin 138(8):1299–1310. https://doi.org/10.1007/s00432-012-1202-6
30. Liu J, Liu L, Cao L, Wen Q (2018) Keratin 17 promote lung adenocarcinoma progression by enhancing cell proliferation and invasion. Med Sci Monit 24:4782–4790. https://doi.org/10.12659/MSM.909350
31. Wang YF, Lang HY, Yuan J, Wang J, Wang R, Zhang XH, Zhang J, Zhao T, Li YR, Liu JY, Zeng LH, Guo GZ (2013) Overexpression of keratin 17 is associated with poor prognosis in epithelial ovarian cancer. Tumour Biol 34(3):1685–1689. https://doi.org/10.1007/s13277-013-0703-5
32. Ricciardelli C, Lokman NA, Pyragius CE, Ween MP, Macpherson AM, Ruszkiewicz A, Hoffmann, Oehler MK (2017) Keratin 5 overexpression is associated with serous ovarian cancer recurrence and chemotherapy resistance. Oncotarget 8(11):17819–17832. https://doi.org/10.18632/oncotarget.14867
33. Eckstein M, Wirtz MR, Gross-Weege M, Breyer J, Otto W, Stoehr R, Sikic D, Keck B, Eidt S, Burger M, Bolenz C, Nitschke K, Porubsky S, Hartmann A, Erben P (2018) mRNA –expression of KRT5 and KRT10 defines distinct prognostic subgroups of muscle-invasive urothelial bladder cancer correlating with histological variants. Int J Mol Sci 19(11):3396. https://doi.org/10.3390/ijms19113396
34. Liu YR, Yin PN, Silvers CR, Lee YF (2019) Enhanced metastatic potential in the MB49 urothelial carcinoma model. Sci Rep 9(1):7425. https://doi.org/10.1038/s41598-019-43641-5
35. Ohkura S, Kondoh N, Hada A, Arai M, Yamazaki Y, Sindoh M, Takahashi M, Matsumoto I, Yamamoto M (2005) Differential expression of the keratin-4, -13, -14, -17 and transglutaminase 3 genes during the development of oral squamous cell carcinoma from leukoplakia. Oral Oncol 41(6):607–613. https://doi.org/10.1016/j.oraloncology.2005.01.011