



Oncoproteomic and gene expression analyses identify prognostic biomarkers for second primary malignancy in patients with head and neck squamous cell carcinoma

Sacarin Bunbanjerdasuk^{1,2} · Nutchavadee Vorasan³ · Thammakorn Saethang⁴ · Tanjitti Pongrujikorn¹ · Duangjai Pangpunyakulchai⁵ · Narongsak Mongkonsiri⁵ · Lalida Arsa⁶ · Nintita Thokanit⁷ · Warut Pongsapich⁸ · Tauangtham Anekpuritanang⁹ · Nuttapong Ngamphaiboon¹⁰ · Artit Jinawath⁶ · Somkiat Sunpaweravong¹¹ · Trairak Pisitkun⁴ · Bhoom Suktitipat^{3,12,13} · Natini Jinawath^{1,7,13}

Received: 16 September 2018 / Revised: 21 December 2018 / Accepted: 6 January 2019 / Published online: 8 February 2019
© United States & Canadian Academy of Pathology 2019

Abstract

Patients with head and neck squamous cell carcinoma are at increased risk of developing a second primary malignancy, which is associated with poor prognosis and early death. To help improve clinical outcome, we aimed to identify biomarkers for second primary malignancy risk prediction using the routinely obtained formalin-fixed paraffin-embedded tissues of the index head and neck cancer. Liquid chromatography-tandem mass spectrometry was initially performed for candidate biomarker discovery in 16 pairs of primary cancer tissues and their matched normal mucosal epithelia from head and neck squamous cell carcinoma patients with or without second primary malignancy. The 32 candidate proteins differentially expressed between head and neck cancers with and without second primary malignancy were identified. Among these, 30 selected candidates and seven more from literature review were further studied using NanoString nCounter gene expression assay in an independent cohort of 49 head and neck cancer patients. Focusing on the p16-negative cases, we showed that a multivariate logistic regression model comprising the expression levels of *ITPR3*, *KMT2D*, *EMILIN1*, and the patient's age can accurately predict second primary malignancy occurrence with 88% sensitivity and 75% specificity. Furthermore, using Cox proportional hazards regression analysis and survival analysis, high expression levels of *ITPR3* and *DSG3* were found to be significantly associated with shorter time to second primary malignancy development (log-rank test $P = 0.017$). In summary, we identified a set of genes whose expressions may serve as the prognostic biomarkers for second primary malignancy occurrence in head and neck squamous cell carcinomas. In combination with the histopathologic examination of index tumor, these biomarkers can be used to guide the optimum frequency of second primary malignancy surveillance, which may lead to early diagnosis and better survival outcome.

Introduction

Head and neck cancer is the seventh most common cancer worldwide, with over 600,000 new cases in 2012 [1]. About 90% of all head and neck cancers are squamous cell carcinomas. Head and neck squamous cell carcinoma is

associated with an elevated likelihood of developing second primary malignancy, which is defined as a second malignancy that presents either simultaneously or after the diagnosis of an index tumor. The overall incidence of second primary malignancies in head and neck squamous cell carcinoma patients has been reported to range from 5.6 to 35.9% [2, 3], with an annual incidence ranging from 3.2 to 4% [4]. The common sites of second primary malignancy are the head and neck, lung, and esophagus [5]. Several risk factors for second primary malignancy in head and neck squamous cell carcinoma patients have been reported including the location of index tumor, patient age, cigarette smoking, alcohol consumption, and betel nut chewing [4].

Despite the progress in cancer molecular biology, the exact underlying molecular mechanism of second primary

Supplementary information The online version of this article (<https://doi.org/10.1038/s41379-019-0211-2>) contains supplementary material, which is available to authorized users.

✉ Natini Jinawath
natini.jin@mahidol.ac.th
jnatini@hotmail.com

Extended author information available on the last page of the article

malignancy is still poorly understood. Field cancerization is one of the widely accepted concepts that explains second primary malignancy tumorigenesis [6]. According to this concept, an area of the upper aerodigestive tract is considered as a field that is continuously exposed to a diversity of carcinogens resulting in multiple precancerous genetic changes, even though the mucosal epithelia still retain their normal histologic appearances. These precancerous fields may eventually become malignant leading to multiple second primary malignancy occurrences [7]. Recently, Curtius et al. has reviewed the evolutionary process that results in field creation [8]. In this context, a cancerized field is both enabled by and causes alterations in the tissue microenvironment. Measurements of the cancerized field evolution hold considerable promise as a new class of biomarker for cancer risk.

Second primary malignancies are known to have a major negative impact on head and neck cancer patients. The prognoses of head and neck cancer patients who develop second primary malignancy are worse than those with index malignancies alone [9], and it is also a leading long-term cause of mortality in head and neck cancer patients [10]. Early diagnosis of second primary malignancies often allows less invasive or curative treatments. However, current clinical screening and surveillance methods including panendoscopy and positron emission tomography/computed tomography (PET/CT) scan still have clinical limitations [11, 12]. Panendoscopy is a relatively invasive procedure and can result in serious complications such as esophageal perforation [11]. While PET/CT scan demonstrates high sensitivity for detecting synchronous second primary malignancies, its findings can be false-positive, and may miss small and/or superficial synchronous second primary malignancies [13]. Furthermore, the attempts to use chemoprevention for second primary malignancy risk reduction have so far been unsuccessful [14]. For these reasons, novel screening approaches such as molecular biomarkers for second primary malignancy risk prediction are very much needed.

Advances in omics technology have resulted in a better understanding of head and neck cancer carcinogenesis, which in turn lead to the development of novel molecular biomarkers for this type of cancer that can be successfully translated into routine clinical practice. Human papillomavirus (HPV) has emerged as an important etiologic factor of head and neck cancers, particularly in oropharyngeal cancers. HPV status is currently used as a prognostic biomarker for head and neck cancer; HPV-positive head and neck squamous cell carcinomas have a favorable prognosis and may benefit from less aggressive treatment regimens [15]. The recently released 8th edition of the American Joint Committee on Cancer (AJCC) Staging Manual, Head and Neck Section, has acknowledged the importance of using

p16 immunostaining as a surrogate marker of HPV status, and issued a distinct staging classification for HPV-associated oropharyngeal cancer, separating it from cancer of the oropharynx related to other causes [16]. Despite these advances, to date, very few studies have focused on second primary malignancy, particularly on a biomarker discovery aspect [17, 18], and so far none has been successfully translated into clinics.

In this study, we hypothesized that the cancerized fields in head and neck squamous cell carcinoma patients with or without second primary malignancy are diverse, leading to distinct protein or gene expression patterns that can be exploited as biomarkers. We then aimed to identify molecular biomarkers for second primary malignancy risk prediction using the routinely collected formalin-fixed paraffin-embedded tissue biopsies of the index head and neck squamous cell carcinomas. Using liquid chromatography-tandem mass spectrometry, we identified a set of candidate biomarkers differentially expressed in the primary tumors of head and neck squamous cell carcinoma patients with or without second primary malignancy after a minimum follow-up period of 3 years. Next, we utilized a NanoString nCounter gene expression assay to study the ability of selected biomarkers to predict second primary malignancy development and the time to second primary malignancy occurrence in another independent cohort of head and neck squamous cell carcinoma patients. Our study permitted the development of the first tissue-based gene expression biomarker panel for second primary malignancy that can easily be implemented in the routine pathology practice.

Materials and methods

Patient selection and sample description

All patients were clinically diagnosed with head and neck cancer by the otolaryngologists or oncologists and had a histopathological diagnosis of squamous cell carcinoma. The patients with only a single primary tumor after a minimum follow-up period of approximately 3 years were defined as “P_{index} group”. The head and neck squamous cell carcinoma patients who developed second primary malignancy during the same follow-up period were defined as “P_{SPM} group”. The clinical diagnosis of second primary malignancy was made based on the applied recommendation from Warren and Gates’ criteria [19] as follows: (i) both index and secondary tumors had histologic confirmation of squamous cell carcinoma; (ii) the two malignancies were anatomically separated by more than 2 cm of normal mucosa; (iii) the possibility of the second primary malignancy being a metastasis from the index tumor was

excluded. Second primary malignancy was further classified as synchronous second primary malignancy when the diagnosis of both index tumor and second tumor was made simultaneously or within 6 months after the index tumor, or as metachronous second primary malignancy when the diagnosis of second tumor was made longer than 6 months after the index tumor [20].

Cohort 1 consisted of 16 head and neck squamous cell carcinoma patients from Songklanagarind Hospital (Prince of Songkla University, Songkhla, Thailand). A pair of formalin-fixed paraffin-embedded tissue blocks containing the index tumor and normal mucosal epithelia was collected for each patient, resulting in a total of 32 samples for proteomic analysis. In addition, all second primary malignancy samples were previously confirmed to be second primary malignancies, not metastatic tumors, by comparing the loss of heterozygosity patterns identified by single nucleotide polymorphism array between each index head and neck squamous cell carcinoma and its matched second tumor [21]. Cohort 2 was an independent multi-center cohort comprising 49 head and neck squamous cell carcinoma patients from Ramathibodi Hospital (Mahidol University, Bangkok, Thailand), Siriraj Hospital (Mahidol University, Bangkok, Thailand), and Songklanagarind Hospital. Only formalin-fixed paraffin-embedded tissue blocks containing the index tumor were collected for each patient. This study was approved by the Institutional Review Board of all participating hospitals.

Sample preparation

Formalin-fixed paraffin-embedded tissue sections were prepared. Previously archived hematoxylin and eosin stained tissue slides were evaluated for tumor content and density by pathologists. The areas with at least 70% tumor cells and normal mucosal epithelial cells were marked for manual macrodissection using a needle tip or scalpel. For liquid chromatography-tandem mass spectrometry, dissected tissue was prepared for protein digestion using the filter-assisted sample preparation method [22] and physically disrupted by sonication. For NanoString gene expression assay, total RNA was isolated using High Pure FFPE RNA Isolation Kit (Roche) according to the manufacturer's instructions.

Liquid chromatography-tandem mass spectrometry

Liquid chromatography-tandem mass spectrometry analysis was performed as a service by Bioproximity, LLC (Chantilly, VA, USA). In brief, proteins were digested with trypsin, and digested peptides were desalted using C18 stop-and-go extraction (STAGE) tips [23]. Peptides were then fractionated by strong anion exchange STAGE tip

chromatography [24]. Liquid chromatography was performed on an Easy nanoLC II HPLC system (Thermo Fisher Scientific). The liquid chromatography was interfaced to a dual pressure linear ion trap mass spectrometer (LTQ Velos, Thermo Fisher Scientific) via nano-electrospray ionization. Data from mass spectrometry were processed and the Mascot generic format (MGF) files were searched using X!Tandem, *k*-score scoring algorithms and OMSSA (open mass spectrometry search algorithm). The common Repository of Adventitious Proteins and the Ensembl release 69 were used as protein databases for the searches. Proteins were required to have two or more unique peptides across the analyzed samples with *E*-value scores of 0.01 or less and protein *E*-value scores of 0.0001 or less [25]. The spectral count of each identified protein was initially scaled by the total and/or maximum peptide counts of the same patient. All scaled spectral counts were added with 0.01 to avoid division by zero, and each scaled spectral count of tumor tissue was then divided by the scaled spectral count of the corresponding normal tissue from the same patient to normalize interindividual variations.

Immunohistochemistry

A Leica BOND-MAX™ automated staining system (Leica Biosystems) was utilized according to the manufacturer's protocol for anti-CALML3 (1:1000, PA5-30232, Thermo Fisher Scientific), and anti-CKMT2 (1:200, PA5-28591, Thermo Fisher Scientific). A Ventana BenchMark XT automated staining system (Ventana Medical Systems) was used according to the manufacturer's instructions for anti-DSG3 (1:50, MAB1720, R&D Systems), anti-PLOD1 (1:50, NBP2-31885, Novus Biologicals), and anti-p16 (CINtec® p16 Histology, 705-4713, Ventana Medical Systems). Blinded scoring was independently performed by two pathologists using the same scoring criteria (Supplementary Table S1).

Digital gene expression quantification using NanoString nCounter®

A NanoString nCounter® Gene Expression Assay (NanoString Technologies) with the custom-designed CodeSets containing 40 genes was performed following the standard protocol. Data processing was conducted using nSolver Analysis Software v3.0 (NanoString Technologies). Background correction was made by subtracting the "mean + 2 standard deviation" value of the negative controls from the raw counts. Adjusted raw counts were then normalized with the geometric mean of the positive controls in each sample. The resulting data was normalized again using the geometric mean of three housekeeping genes. Normalized data was standardized using mean centering then divided by

Table 1 Clinical characteristics of patients

	Cohort 1				Cohort 2			
	P _{index}	P _{SPM}	Total	P-value	P _{index}	P _{SPM}	Total	P-value
No. of patients	5	11	16		21	28	49	
Age								
Mean (years)	61	57	58	0.387 ^a	64	57	60	0.027^{a*}
< 65	4	9	13	1.000 ^c	11	22	33	0.053 ^b
≥ 65	1	2	3		10	6	16	
Gender								
Male	5	10	15	1.000 ^c	18	28	46	0.072 ^c
Female	0	1	1		3	0	3	
Smoking status								
Never	0	1	1	1.000 ^c	2	2	4	1.000 ^c
Former ^d	1	2	3		7	9	16	
Active	2	5	7		11	16	27	
Unknown	2	3	5		1	1	2	
Alcohol consumption								
Never	2	1	3	0.500 ^c	5	1	6	0.128 ^c
Former ^d	0	2	2		4	5	9	
Active	1	3	4		10	18	28	
Unknown	2	5	7		2	4	6	
Site of primary (Index) tumor								
Oral cavity	0	1	1	1.000 ^c	6	4	10	0.172 ^c
Oropharynx	0	2	2		3	8	11	
Hypopharynx	4	7	11		5	12	17	
Larynx	1	1	2		6	4	10	
Nose & Paranasal sinuses	0	0	0		1	0	1	
Stage ^e of primary (Index) tumor								
I	0	3	3	0.646 ^c	2	3	5	0.105 ^c
II	0	0	0		2	7	9	
III	3	4	7		7	2	9	
IV	2	4	6		10	16	26	
HPV status								
p16 (+)	1	2	3	1.000 ^c	1	2	3	1.000 ^c
p16 (-)	4	7	11		20	26	46	
Unknown	0	2	2		0	0	0	
Median time from index tumor diagnosis to last follow-up or death (months)	31 (30–86)				54 (35–95)			
Classification of SPM								
Synchronous SPM		4				12		
Metachronous SPM		7				16		
Site of SPM								
Head and Neck		0				3		
Esophagus		11				21		
Lung		0				4		
Median time from index tumor diagnosis to SPM occurrence (months)		14 (0–119)				10 (0–65)		

SPM second primary malignancy, *Synchronous SPM* second primary malignancy diagnosed simultaneously or within 6 months after the index tumor, *Metachronous SPM* second primary malignancy diagnosed more than 6 months after the index tumor

^at test

^bChi-Square test

^cFisher's exact test

^dQuit smoking at least 1 year before this study began

^eThe American Joint Committee on Cancer staging (7th edition)

*Statistically significant ($P < 0.05$)

the standard deviation to obtain the Z scores for subsequent statistical analyses.

Bioinformatics and statistical analysis

Hierarchical cluster analysis using average linkage method was performed using Cluster 3.0 (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>), and visualized by Java TreeView version 1.1.6r4 (<http://jtreeview.sourceforge.net>). Functional annotation of candidate proteins was performed using DAVID 6.8 (<https://david.ncifcrf.gov/>). Statistical analyses were performed using PASW Statistics version 18.0 for Windows (SPSS Inc.), GraphPad Prism version 6.0 for Windows (GraphPad Software), R-statistical software version 3.3.1 (R Foundation, <http://www.r-project.org>), and ROCR package [26].

Results

Subject characteristics

Two independent cohorts of head and neck squamous cell carcinoma patients were included in this study. Among the 16 patients in cohort 1, 5 (31%) and 11 (69%) patients belonged to the P_{index} and P_{SPM} groups, respectively. The median duration of follow-up to ensure no occurrence of second primary malignancy was 31 (30–86) months in the P_{index} group. All second primary malignancies were esophageal squamous cell carcinomas. No statistically significant difference was detected between the clinical characteristics of patients in the P_{index} and P_{SPM} groups.

In cohort 2, of the 49 head and neck squamous cell carcinoma patients examined, 21 (43%) and 28 (57%) patients belonged to the P_{index} and P_{SPM} groups, respectively. The median duration of follow-up was 54 (35–95) months in the P_{index} group. All second primary malignancies were squamous cell carcinomas located in the esophagus, head and neck, or lung. The patients' age was the only clinical characteristic with statistically significant difference between the P_{index} and P_{SPM} groups ($P = 0.027$). The detailed clinicopathological and demographical characteristics of patients in both cohorts are summarized in Table 1.

The HPV status of head and neck tumors was also evaluated using p16 immunohistochemistry. The difference in HPV status based on p16-positivity between the P_{index} and P_{SPM} groups from the two cohorts was not statistically significant (Fisher's exact test $P = 1.000$). Of note, only three cases out of the 49 head and neck squamous cell carcinomas in cohort 2 were p16-positive. The p16 immunostaining results are shown in Supplementary Table S2.

Liquid chromatography-tandem mass spectrometry analysis and validation by immunohistochemistry

The detailed experimental outline of this study is summarized in Fig. 1. We first performed liquid chromatography-tandem mass spectrometry in 16 pairs of tumor-normal formalin-fixed paraffin-embedded tissues from P_{index} and P_{SPM} groups (cohort 1) as a means to discover candidate biomarkers for second primary malignancy. A total of 2816 proteins were initially identified. Proteins detected in <10% of all samples were then filtered out, leaving a total of 2101 proteins for unsupervised hierarchical cluster analysis. The resulted dendrogram showed that these protein profiles could correctly classify the tumor and normal tissue samples (Fig. 2a). We further investigated the protein profiles within the tumor or normal tissue subgroups separately whether they could correctly define each sample as P_{index} or P_{SPM} by performing hierarchical cluster analysis of differentially expressed proteins (t test $P < 0.05$). The results interestingly demonstrated that P_{index} and P_{SPM} samples could be accurately clustered in both tumor and normal tissue subgroups (Supplementary Figure S1).

In addition, immunohistochemistry analysis of representative proteins was performed to validate the accuracy of liquid chromatography-tandem mass spectrometry results. Four proteins that were at least two-fold up- (PLOD1) or downregulated (CKMT2, DSG3, CALML3) in tumors as compared to normal tissues, and expressed in more than half of the samples, were selected. The results demonstrated that immunohistochemistry analysis showed the same up- or downregulated trend between normal and tumor tissues as in the proteomics data in three of four proteins (DSG3, CALML3, PLOD1) (Fig. 2b–c).

Identification of candidate proteins biomarkers for second primary malignancy

To screen for the differentially expressed proteins between head and neck squamous cell carcinoma patients with and without second primary malignancy, the protein expression ratio of tumor to matched normal tissue of each patient was analyzed. The inclusion criteria for candidate proteins included: (i) proteins expressed in more than 50% of the samples, (ii) proteins showing at least three-fold difference in expression level between the P_{index} and P_{SPM} groups, and (iii) proteins exhibiting statistically significant differential expression between the two groups (t test $P < 0.05$). Based on these criteria, 32 candidate proteins were identified (Supplementary Table S3). Hierarchical cluster analysis using expression ratio of these proteins demonstrated that all cohort 1 samples could be correctly divided into P_{index} and P_{SPM} groups (Fig. 3).

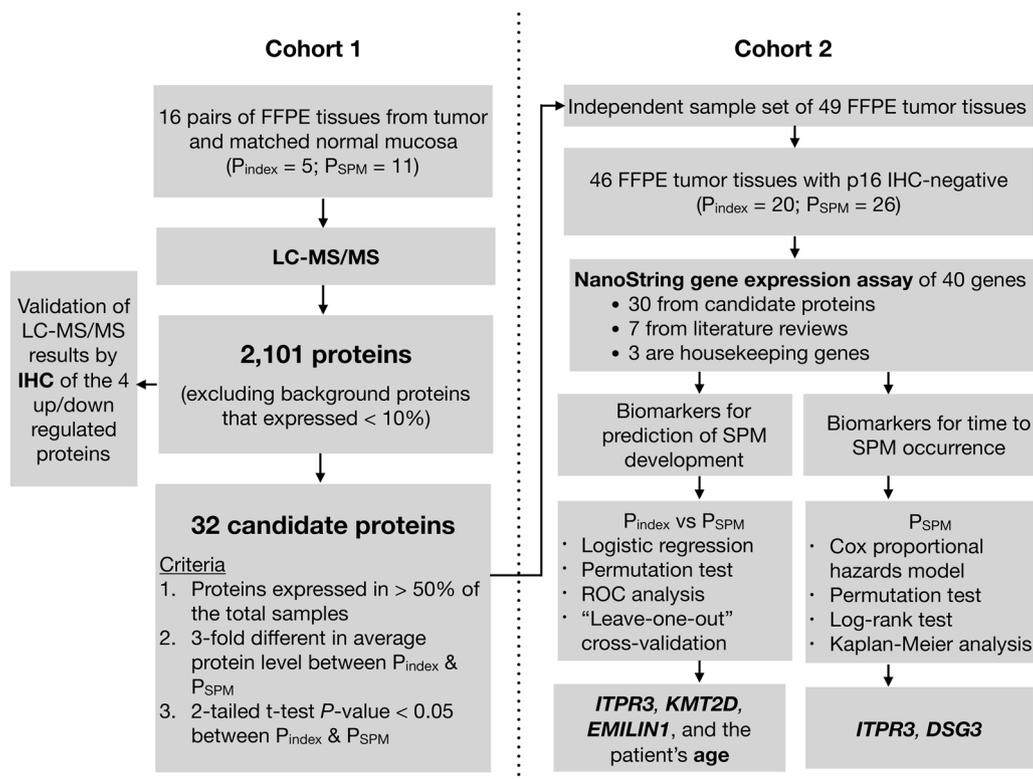


Fig. 1 Overview and experimental design of the study. The biomarker identification steps were sequentially conducted using liquid chromatography-tandem mass spectrometry in cohort 1 (left), and NanoString gene expression assay in cohort 2 (right). Multiple statistical analyses were performed to identify prognostic biomarkers for second primary malignancy occurrence and time to second primary

malignancy development. SPM second primary malignancy, P_{index} index head and neck squamous cell carcinoma without second primary malignancy, P_{SPM} index head and neck squamous cell carcinoma with second primary malignancy, FFPE formalin-fixed paraffin-embedded, LC-MS/MS liquid chromatography-tandem mass spectrometry, IHC immunohistochemistry, ROC receiver operating characteristic

We further carried out gene-annotation enrichment analysis using DAVID 6.8 to identify enriched biological functions of the candidate proteins. Among the upregulated proteins in P_{SPM} samples, the statistically significant gene ontology term ($P < 0.001$) with highest enrichment score (3.88) was “intermediate filament” (KRT37, KRT12, KRT86, KRT24). The other significant gene ontology terms ($P < 0.05$) were “cytoplasm” (UBE2N, PA2G4, S100A7, ARPC2, SERPINB5, TRIM29, CRABP2, EIF5A, RPS20, AHNAK, KRT24), and “poly(A) RNA binding” (UBE2N, PA2G4, EIF5A, RPS20, AHNAK). In contrast, the significant terms with highest enrichment score (1.31) among the downregulated proteins were “extracellular exosome” and “extracellular matrix” (PTGES3, PLOD1, FBLN2, UGDH, CSRP1, ECM1, EMILIN1) ($P < 0.01$) (Fig. 3 and Supplementary Table S4A-B).

Development of a predictive model for second primary malignancy occurrence by NanoString nCounter gene expression analysis

Considering the difficulties in translating tissue-based quantitative protein biomarkers into clinical practice, we

set out to alternatively utilize a NanoString nCounter gene expression assay to optimize the biomarker selection in an independent cohort (cohort 2). The 40-gene custom panel consisted of the 30 candidate proteins (of the 32 candidates, two did not have specific CodeSets available), three housekeeping genes, and seven frequently mutated genes in head and neck squamous cell carcinomas with synchronous nodal metastasis or metachronous recurrence from literature review [27, 28] (Supplementary Table S5). HPV-positive and HPV-negative head and neck cancers are known to exhibit different clinical and molecular characteristics [16]. Since the number of p16-positive cases in our cohort was limited (3 out of 49 cases), we focused on the 46 p16-negative head and neck squamous cell carcinoma patients for further analyses.

To identify biomarkers that could accurately determine the development of second primary malignancy, univariate logistic regression analysis of the standardized NanoString gene expression levels and selected clinical variables of the 46 p16-negative head and neck squamous cell carcinoma patients was carried out. The expression levels of three genes (*ITPR3*, *FAT1*, *KMT2D*) and the patient's age at diagnosis were statistically

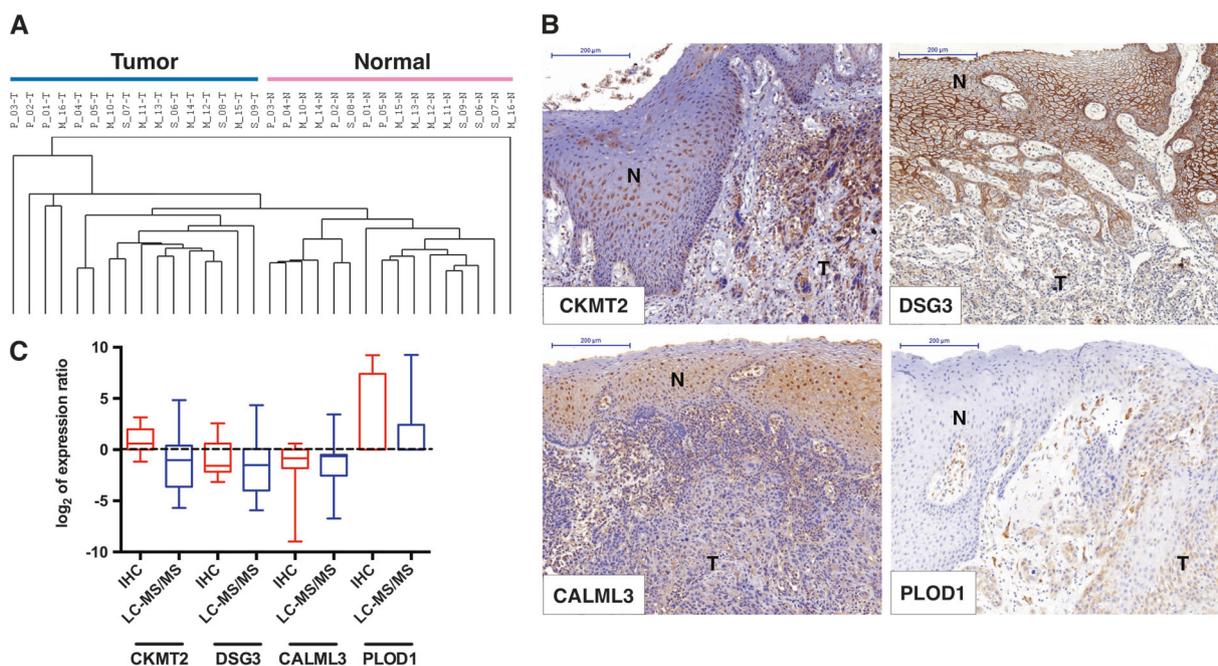


Fig. 2 Protein expression profiling by liquid chromatography-tandem mass spectrometry and validation by immunohistochemistry. **a** Dendrogram of unsupervised hierarchical cluster analysis of the 2101 protein profiles across the tumor samples and their matched normal mucosal tissues in cohort 1. **b** Representative immunohistochemistry images ($\times 10$ magnification) of the four selected proteins (CKMT2,

DSG3, CALML3, and PLOD1). N: normal mucosal epithelial cells, T: tumor cells. **c** Box plot showing the comparison between normalized expression ratios (T/N) of the four selected proteins analyzed by immunohistochemistry and liquid chromatography-tandem mass spectrometry. IHC immunohistochemistry, LC-MS/MS liquid chromatography-tandem mass spectrometry

significantly associated with the development of second primary malignancy ($P < 0.05$) (Fig. 4a). The strongest risk factor for the development of second primary malignancy was high *ITPR3* level. An increase in one standard deviation of *ITPR3* level was associated with 3.27 times higher risk of second primary malignancy development (odds ratio = 3.27; 95% confidence interval (CI) [1.36–11.04]; $P = 0.025$). Other significant risk factors included high *FAT1* (odds ratio = 2.25; 95% CI [1.13–5.49]; $P = 0.038$) and *KMT2D* levels (odds ratio = 2.20; 95% CI [1.10–5.18]; $P = 0.043$). To minimize the effect of small sample size, we performed 5000 permutation tests to correct for non-asymptotic properties of P -values by reshuffling the observed data. The resulting P -values were reported as empirical P -values. The analysis showed that the expression levels of two more genes (*EMILIN1* and *ECM1*) were also statistically significantly associated with second primary malignancy development (empirical $P < 0.05$). The results of univariate logistic regression of all genes are shown in Supplementary Table S6. Comparing between P_{index} and P_{SPM} groups, the expression levels of *ITPR3*, *FAT1*, *KMT2D*, and *ECM1* were higher in P_{SPM} samples, whereas the level of *EMILIN1* was lower (Fig. 4b).

We next aimed to identify the best combination of biomarkers for second primary malignancy development

prediction by analyzing the six statistically significant variables ($P < 0.05$ or empirical $P < 0.05$) from univariate logistic regression (*ITPR3*, *FAT1*, *KMT2D*, *EMILIN1*, *ECM1*, and patient's age) using multivariate logistic regression analysis. The most parsimonious final model based on Akaike's information criterion, Bayesian information criterion, and deviance test was the combination of *ITPR3*, *KMT2D*, *EMILIN1*, and the patient's age. The Akaike's information criterion, and Bayesian information criterion values of this model were 50.82 and 59.96, respectively. The detailed results of all multivariate logistic regression models and the final selection are shown in Supplementary Table S7 and S8.

To evaluate the performance of the selected multivariate model in predicting second primary malignancy occurrence, a receiver operating characteristic analysis was performed and an area under the curve value was calculated as 0.86 (95% CI [0.75–0.97]). Moreover, this model has a sensitivity of 88.46%, a specificity of 75.00% and an accuracy of 82.61% for predicting the occurrence of second primary malignancy. We further performed a leave-one-out cross-validation over the same 46 samples (cohort 2) to assess the robustness of our model. The resulting area under the curve, sensitivity, specificity and accuracy values were 0.80 (95% CI [0.66–0.94]), 76.92%, 70.00%, and 73.91%, respectively (Fig. 4c). These new values were not significantly different

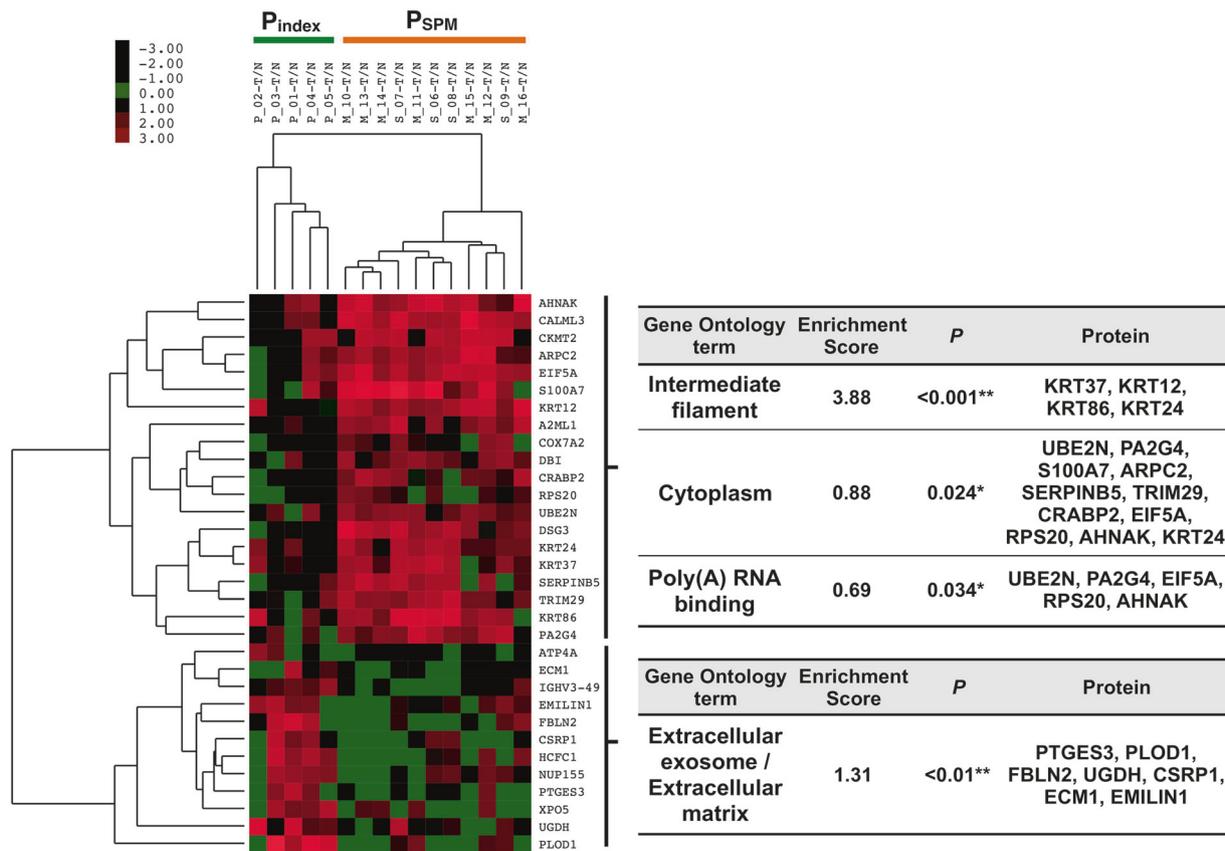


Fig. 3 Hierarchical cluster analysis of the candidate proteins. A heatmap showing the 32 candidate proteins differentially expressed between P_{index} and P_{SPM} subgroups with statistical significance (t test $P < 0.05$) across cohort 1 (left), and the summary of gene-annotation enrichment analysis of up- and downregulated proteins of each subgroup using DAVID bioinformatics resources 6.8 (right). The color

scale is shown at the upper left corner. P_{index} : index head and neck squamous cell carcinoma without second primary malignancy, P_{SPM} : index head and neck squamous cell carcinoma with second primary malignancy. *Statistically significant ($P < 0.05$), **statistically significant ($P < 0.01$)

from the original results, confirming the validity of our predictive model.

Identification of biomarkers for predicting time to second primary malignancy development

Next, we focused on head and neck squamous cell carcinoma patients who eventually developed second primary malignancy (P_{SPM} group) in order to identify the biomarkers associated with time to subsequent second primary malignancy occurrence. Univariate Cox regression analysis using gene expression levels of the 26 p16-negative head and neck squamous cell carcinomas from cohort 2 was performed. The levels of *ITPR3*, *DBI*, *AHNAK*, *IGHV3-49*, *CALML3*, *ARPC2*, *DSG3*, and *KRT37* were significantly associated with a shorter time to second primary malignancy development ($P < 0.05$) (Fig. 5a). The strongest association was with *ITPR3* level. An increase in one standard deviation of *ITPR3* level was associated with 2.68 times higher risk of second primary

malignancy development at any given time after the index tumor diagnosis (hazard ratio = 2.68; 95% CI [1.53–4.72]; $P = 0.001$). The complete results of univariate Cox regression analysis including the empirical P -values generated by 5000 permutation tests are shown in Supplementary Table S9. Using multivariate Cox proportional hazards regression analysis, the best model was the combination of *ITPR3* and *DSG3* (Supplementary Table S10). Moreover, by dividing the standardized *ITPR3* and *DSG3* level by the sample mean of the 26 P_{SPM} patients, survival analysis showed that the patients with p16-negative head and neck squamous cell carcinoma whose index tumors exhibited high *ITPR3* and *DSG3* expression levels had the shortest time interval between the diagnosis of an index head and neck squamous cell carcinoma to subsequent second primary malignancy development (log-rank test $P = 0.017$) (median time difference between the high and low risk groups based on four combinations of *ITPR3* and *DSG3* levels = 394 days) (Fig. 5b and Supplementary Table S11).

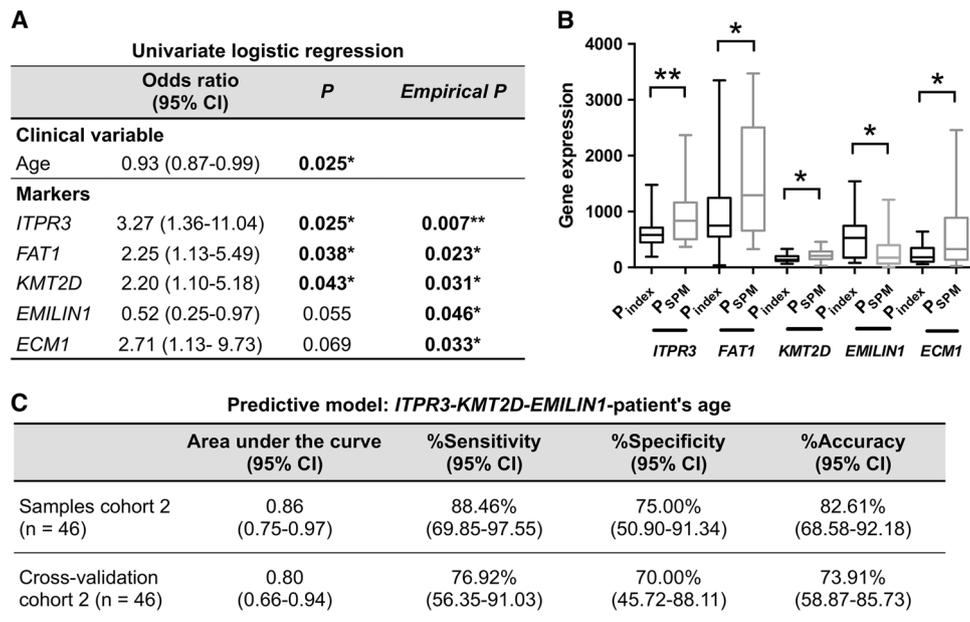


Fig. 4 Identification of biomarkers for second primary malignancy risk prediction by NanoString gene expression analysis. **a** The statistically significant variables identified by univariate logistic regression across P_{index} and P_{SPM} subgroups of the p16-negative tumors in cohort 2 ($n = 46$). The empirical P -values were generated by 5000 permutation tests. **b** Box plot depicting expression levels of the five candidate genes identified by univariate logistic regression across P_{index} and P_{SPM} subgroups of the p16-negative tumors in cohort 2. All pairwise comparisons show statistically significant differential expressions. **c** The

predictive performance of the selected multivariate logistic regression model (*ITPR3-KMT2D-EMILIN1*-patient's age) in classifying second primary malignancy risk. The robustness of the model was assessed by performing a leave-one-out cross-validation for each sample in the same cohort. The unit of patient's age is year. P_{index} : index head and neck squamous cell carcinoma without second primary malignancy, P_{SPM} : index head and neck squamous cell carcinoma with second primary malignancy. *Statistically significant ($P < 0.05$), **statistically significant ($P < 0.01$), CI confidence interval

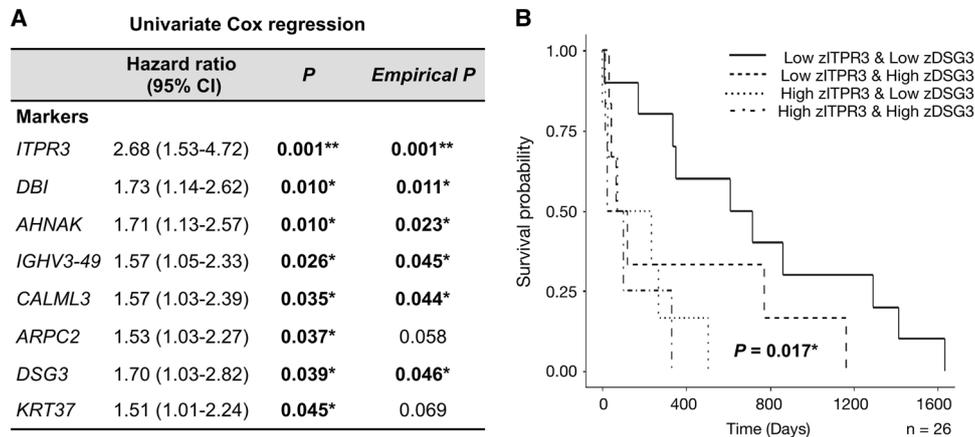


Fig. 5 Identification of biomarkers for predicting time to second primary malignancy development. **a** The statistically significant genes identified by univariate Cox regression across P_{SPM} patients with p16-negative tumor in cohort 2 ($n = 26$). The empirical P -values were generated by 5000 permutation tests. **b** Kaplan-Meier curves showing the proportions of P_{SPM} patients with p16-negative tumor stratified by standardized *ITPR3* and *DSG3* expression. High and low levels of

ITPR3 and *DSG3* expression were defined as standardized gene levels above or below mean, respectively. Patients with high *ITPR3* and high *DSG3* expression levels had the shortest time to second primary malignancy development (log-rank test $P = 0.017$). CI confidence interval, *statistically significant ($P < 0.05$), **statistically significant ($P < 0.01$)

Comparison of the expression pattern of candidate biomarkers between protein and mRNA levels

It is known that protein and mRNA expression levels often do not directly correlate [29]. To investigate whether our candidate genes originally discovered by proteomics study shared the same differential expression patterns as their protein equivalents, we compared the standardized gene expressions of the top 10 statistically significant genes identified by logistic regression and Cox regression analyses to their standardized protein levels obtained from liquid chromatography-tandem mass spectrometry. Seven of the 10 biomarkers (*AHNAK*, *ARPC2*, *CALML3*, *DBI*, *DSG3*, *EMILIN1*, *KRT12*) demonstrated a similar up- or downregulation trend between mRNA and protein levels in the tumor samples across the two patient cohorts (Fig. 6).

Discussion

In this study, we identified a set of tissue-based biomarkers for predicting second primary malignancy occurrence and time to second primary malignancy development in head and neck squamous cell carcinoma patients using a combination of high-throughput shotgun proteomics and targeted gene expression analysis. These second primary malignancy risk prediction biomarkers can help guide clinical management of head and neck squamous cell carcinoma patients, particularly in the frequency of second primary malignancy surveillance after the diagnosis of an index tumor, and the choice of treatments. The expected long-term benefit is an improvement in overall survival of head and neck cancer patients, especially those who are eligible for curative or less invasive therapy. Moreover, one of the clinical strengths of this study is the use of index head and neck squamous cell carcinoma formalin-fixed paraffin-embedded tissue as the preferred material for biomarker

discovery. This specimen is routinely acquired for histopathological diagnosis, which makes it easier to incorporate our biomarker panel into the current clinical practice. We believe that our study is the first to identify prognostic biomarkers that can accurately predict second primary malignancy occurrence, thus opening the door to the possibility of clinical application of tissue-based biomarkers for second primary malignancy in head and neck squamous cell carcinoma patients.

Mass spectrometry has been extensively used to discover novel protein biomarkers [30]. Proteins are the functional molecules in the cell, and thus are the key players that represent actual cellular physiology. For tissue-based protein biomarkers, several laboratory methods such as immunohistochemistry and targeted mass spectrometry can be used to develop the validated markers into quantitative clinical assays but not without limitations. To date, immunohistochemistry is routinely performed in clinics to determine both the qualitative and semi-quantitative aspects of a protein biomarker, however, it relies heavily on the quality of antibody and still lacks interpretation standardization, resulting in poor reproducibility [31]. While targeted mass spectrometry can do marker multiplexing and does not rely on antibody, it is technically complex and carry a high developing cost [32]. The development of tissue-based high-throughput gene expression assays has greatly improved the accuracy and reproducibility of quantitative measurement in tissue biomarker studies. Currently, several such assays have been integrated into clinical practice. For example, the Prosigna breast cancer prognostic gene signature assay based on NanoString nCounter analysis system was approved by U.S. Food and Drug Administration in 2013 as a prognostic assay for distant recurrence risk in breast cancer patients [33]. In this study, we sequentially utilized liquid chromatography-tandem mass spectrometry and NanoString nCounter system in two independent cohorts of head and neck squamous cell carcinoma patients in order to identify the best set of second primary malignancy prediction biomarkers. In light of its clinical translatability, technical reproducibility, and compatibility with small biopsied formalin-fixed paraffin-embedded tissues [34], the NanoString digital gene expression platform was chosen for biomarker panel development.

In cohort 1, the protein profiles were markedly different between head and neck squamous cell carcinomas and their matched normal squamous mucosa, which is in line with previous studies [17, 35]. Interestingly, we also observed distinct protein expression profiles between the P_{index} and P_{SPM} subgroups when analyzing the tumor and normal tissue samples separately. These results suggest that tissue microenvironments of both tumor and surrounding normal mucosal cells of the P_{index} and P_{SPM} patients may exhibit varying levels of genetic diversity, resulting in distinct

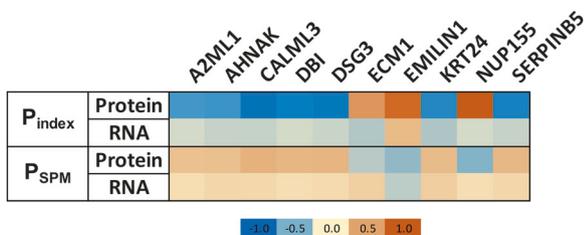


Fig. 6 Comparison of the expression patterns between protein and mRNA levels of the candidate biomarkers across the P_{index} and P_{SPM} subgroups. The standardized NanoString gene expressions of the top 10 statistically significant genes from cohort 2 were compared to their standardized protein levels in the tumor samples of cohort 1 as quantified by liquid chromatography-tandem mass spectrometry. Color codes represent the standardized levels of protein and mRNA expression from low (blue) to high (red)

protein and/or gene expression signatures. These findings strongly support our hypothesis that the aberrant field effect in the biopsied index tumors can be used to predict whether the patients are at risk of subsequent second primary malignancy development. It is proposed that genetic diversity can be utilized as a marker for field evolvability [8]. In a prospective pilot study by Roesch-Ely et al., by analyzing the protein profiles of mucosal biopsies from the oropharynx, hypopharynx, and three different regions of esophagus in head and neck squamous cell carcinoma patients and controls, tumor relapse was correctly predicted [17]. Taken together, these findings firmly established an altered field as a promising cancer risk prediction marker.

Of note, among the candidate proteins differentially expressed between P_{index} and P_{SPM} samples, upregulated proteins in P_{SPM} group were significantly enriched with proteins associated with intermediate filament, whereas those involved with extracellular matrix were downregulated. Overexpression of keratins, the intermediate filament-forming proteins of epithelial cells, is associated with enhanced tumor cell migration and invasion through interactions with extracellular environment [36]. Extracellular matrix proteins are associated with both tumor suppression and tumor promotion. Interestingly, the majority of downregulated extracellular matrix proteins in P_{SPM} samples including ECM1, EMILIN1, and FBLN2 have demonstrated tumor suppressive roles in various cancers [37–39]. Based on our findings, aberrant cytoskeletal activity and altered cell motility as a result of reduced expression of extracellular matrix-associated tumor suppressors are among the major molecular mechanisms underlying second primary malignancy development.

Recently, a large head and neck squamous cell carcinoma genomics study from The Cancer Genome Atlas (TCGA) consisting of mostly Caucasian patients revealed that 64% and 6% of oropharyngeal and non-oropharyngeal tumors, respectively, are HPV-positive [28]. In contrast, studies from Thailand reported that HPV status was positive in only 26.09% of oropharyngeal cancers and in none of the 80 non-oropharyngeal cancers tested [40, 41]. This fact is also reflected in our largely p16-negative cohorts. In addition, multiple studies have reported that patients with HPV-negative oropharyngeal squamous cell carcinoma have a higher risk of second primary malignancy development than HPV-positive oropharyngeal squamous cell carcinoma patients [42, 43]. Since most of our patients were p16-negative and HPV-negative head and neck squamous cell carcinoma patients are more likely to have second primary malignancy, we decided to focus on the p16-negative cases in cohort 2 for our gene expression analyses.

For cohort 2, in addition to the candidate genes from our proteomics study, seven genes from recent genomics literatures [27, 28] were included in the gene expression

panel. *TP53* and *NOTCH1* are among the most commonly mutated genes in head and neck squamous cell carcinoma [28]. A recent whole exome sequencing study of head and neck squamous cell carcinomas has reported that *C17orf104* and *ITPR3* are specifically mutated in synchronous nodal metastases but not in the primary cancers, while *DDR2* is exclusively mutated in metachronous recurrent tumors [27]. Additionally, *FAT1* and *KMT2D* are found to be mutated in both primary head and neck squamous cell carcinomas and their nodal metastases [27]. Hence, it is of interest to investigate whether their expressions are associated with second primary malignancy risk. In this study, the combination of three genes (*ITPR3*, *KMT2D*, *EMILIN1*) and patient's age allowed for the most accurate discrimination between P_{index} and P_{SPM} groups with a sensitivity of 88.46% and a specificity of 75.00%. Of note, while these numbers will require validation in the future studies using larger sample size, our panel showed a comparable level of performance to that of the currently used clinical gene expression tests in breast cancer including Oncotype Dx (sensitivity 71–85%, specificity 55–66% for high and intermediate risk groups) and MammaPrint (sensitivity 83–92%, specificity 41–59%) [44].

In addition to its implications in head and neck squamous cell carcinoma [27], increased expression of *ITPR3*, a major intracellular Ca^{2+} release channel, is associated with enhanced tumor proliferation and invasion in breast and colorectal cancers [45, 46]. *KMT2D*, which encodes a histone H3 lysine 4 methyltransferase, is one of the most commonly mutated genes in cancers [47]. High *KMT2D* expression has recently been reported to promote tumor progression by inducing epithelial-mesenchymal transition (EMT), and is a predictor of poor prognosis in esophageal squamous cell carcinoma [48]. In contrast, *EMILIN1*, which encodes an extracellular matrix glycoprotein associated with the development of elastic tissues, plays a suppressive role in tumor growth, tumor lymphatic vessel formation, and metastatic spread to lymph nodes [38]. A recent report has suggested that $\alpha 9\beta 1$ integrin, of which *EMILIN1* is a ligand, is the major integrin involved in regulating head and neck squamous cell carcinoma cell migration on extracellular matrix [49]. In our study, increased *ITPR3* and *KMT2D* levels, as well as reduced *EMILIN1* are significantly associated with second primary malignancy in head and neck squamous cell carcinoma. Their exact roles in second primary malignancy tumorigenesis remain to be explored.

Second primary malignancy can be classified into either a synchronous or metachronous tumor. Since survival outcomes between synchronous and metachronous second primary malignancies are different [5, 9], it is clinically useful to be able to predict the time to second tumor development. The combination of two genes, *ITPR3* and

DSG3, was statistically significantly associated with the time to second primary malignancy occurrence in our study. *DSG3*, a component of cell–cell junctions, is overexpressed in head and neck squamous cell carcinomas, and its inhibition significantly suppresses tumor growth [50]. *DSG3* has also been proposed as a predictive biomarker for cervical lymph node micrometastasis in oral cancer [51]. In line with their roles in promoting tumor proliferation and progression, high expression of both *ITPR3* and *DSG3* contributed to the shortest time to second primary malignancy development in patients with p16-negative head and neck squamous cell carcinoma.

It is generally known that transcript levels by themselves are by no means an accurate predicting factor for protein levels in many scenarios [29]. In this study, the majority of top candidate biomarkers remarkably showed a similar pattern of differential protein and mRNA expressions in the tumor samples across the P_{index} and P_{SPM} subgroups of both cohorts. This finding partly explains a smooth transition from protein to transcript level of this particular set of markers in regard to second primary malignancy risk prediction, and may also render future biomarker assay development the flexibility and feasibility to use either gene expression or targeted proteomic platforms.

According to the current follow-up recommendation by the National Comprehensive Cancer Network (NCCN) Guidelines for Head and Neck Cancers (version 2.2018) [52], history and physical exams including a complete head and neck exam; and mirror and fiberoptic examination should be performed at least every 3 months during the first year, and at least every 6 months in the second year. Based on our results, all patients with high *ITPR3* and *DSG3* levels developed second primary malignancy within ~13 months (400 days) after the diagnosis of an index tumor. Therefore, it may be beneficial to provide closer monitoring of second primary malignancy for this group of patients during their first years. Moreover, even though some synchronous second primary malignancies develop around the same time as the index tumor, they often cannot be clinically detected until their sizes are big enough. The use of our biomarkers should help increase early diagnosis of both subclinical synchronous second primary malignancies and subsequent metachronous second primary malignancies through a more rigorous surveillance protocol.

Limitations of this study include the relatively small sample size and limited follow-up time to second primary malignancy occurrence. Others include the use of a cohort of convenience where adequate formalin-fixed paraffin-embedded tissues was available, and the use of one tissue sample per tumor, which may not represent intratumor heterogeneity.

In summary, our biomarkers demonstrate great potential as a companion prognostic test for second primary

malignancy risk prediction in routine clinical practice. The expected long-term benefit of early detection of second primary malignancy is an improvement in overall survival of head and neck cancer patients, especially those who are eligible for curative or less invasive therapy. The results are compelling and warrant future validation studies in larger head and neck squamous cell carcinoma cohorts. The protein signatures also hint at tumor-extracellular matrix interactions as a major player in second primary malignancy tumorigenesis. Further functional studies of these biomarkers may better clarify the clinical utility and their roles in second primary malignancy development.

Acknowledgements We would like to thank Mr. Jeffrey Makin, Dr. Premyot Ngaoteprutaram, Dr. Tanasarun Watcharadilokkul, and Dr. Meng-Shin Shiao for their valuable comments. S. Bunbanjerdusuk is a recipient of research assistant scholarship from Faculty of Medicine Ramathibodi Hospital and Faculty of Graduate Studies, Mahidol University. N. Jinawath is a recipient of TRF Research Scholar Fund (RSA5780065), the research grants from Mahidol University-National Research Council of Thailand (NRCT), CEMB-PERDO, and the Ramathibodi Comprehensive Cancer Center. The NJ laboratory is partially supported by the Crown Property Bureau Foundation through Integrative Computational BioScience (ICBS) Center, Mahidol University.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136:E359–E86.
2. González-García R, Naval-Gías L, Román-Romero L, Sastre-Pérez J, Rodríguez-Campo FJ. Local recurrences and second primary tumors from squamous cell carcinoma of the oral cavity: a retrospective analytic study of 500 patients. *Head Neck*. 2009;31:1168–80.
3. Yamashita T, Araki K, Tomifuji M, Tanaka Y, Harada E, Suzuki T, et al. Clinical features and treatment outcomes of Japanese head and neck cancer patients with a second primary cancer. *Asia Pac J Clin Oncol*. 2017;13:172–8.
4. Lee DH, Roh J-L, Baek S, Jung JH, Choi S-H, Nam SY, et al. Second cancer incidence, risk factor, and specific mortality in head and neck squamous cell carcinoma. *Otolaryngol Head Neck Surg*. 2013;149:579–86.
5. Vaamonde P, Martin C, Labella T. Second primary malignancies in patients with cancer of the head and neck. *Otolaryngol Head Neck Surg*. 2003;129:65–70.
6. Slaughter DP, Southwick HW, Smejkal W. “Field cancerization” in oral stratified squamous epithelium. *Clin Implic multicentric Orig Cancer*. 1953;6:963–8.
7. Braakhuis BJ, Tabor MP, René Leemans C, van der Waal I, Snow GB, Brakenhoff RH. Second primary tumors and field

- cancerization in oral and oropharyngeal cancer: molecular techniques provide new insights and definitions. *Head Neck*. 2002;24:198–206.
8. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nat Rev Cancer*. 2018;18:19–32.
 9. Di Martino E, Sellhaus B, Hausmann R, Minkenberg R, Lohmann M, Esthofen MW. Survival in second primary malignancies of patients with head and neck cancer. *J Laryngol Otol*. 2002;116:831–8.
 10. Baxi SS, Pinheiro LC, Patil SM, Pfister DG, Oeffinger KC, Elkin EB. Causes of death in long-term survivors of head and neck cancer. *Cancer*. 2014;120:1507–13.
 11. Clayburgh DR, Brickman D. Is esophagoscopy necessary during panendoscopy? *Laryngoscope*. 2017;127:2–3.
 12. Haerle SK, Strobel K, Hany TF, Sidler D, Stoeckli SJ. 18F-FDG-PET/CT versus panendoscopy for the detection of synchronous second primary tumors in patients with head and neck squamous cell carcinoma. *Head Neck*. 2010;32:319–25.
 13. Yabuki K, Kubota A, Horiuchi C, Taguchi T, Nishimura G, Inamori M. Limitations of PET and PET/CT in detecting upper gastrointestinal synchronous cancer in patients with head and neck carcinoma. *Eur Arch Otorhinolaryngol*. 2013;270:727–33.
 14. Khuri FR, Lee JJ, Lippman SM, Kim ES, Cooper JS, Benner SE, et al. Randomized phase III trial of low-dose isotretinoin for prevention of second primary tumors in stage I and II head and neck cancer patients. *J Natl Cancer Inst*. 2006;98:441–50.
 15. Kang H, Kiess A, Chung CH. Emerging biomarkers in head and neck cancer in the era of genomics. *Nat Rev Clin Oncol*. 2015;12:11–26.
 16. Lydiatt WM, Patel SG, O'Sullivan B, Brandwein MS, Ridge JA, Migliacci JC, et al. Head and neck cancers—major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA Cancer J Clin*. 2017;67:122–37.
 17. Roesch-Ely M, Leopold A, Nees M, Holzinger D, Dietz A, Flechtenmacher C, et al. Proteomic analysis of field cancerization in pharynx and oesophagus: a prospective pilot study. *J Pathol*. 2010;221:462–70.
 18. Hildebrandt MA, Lippman SM, Etzel CJ, Kim E, Lee JJ, Khuri FR, et al. Genetic variants in the PI3K/PTEN/AKT/MTOR pathway predict head and neck cancer patient second primary tumor/recurrence risk and response to retinoid chemoprevention. *Clin Cancer Res*. 2012;18:3705–13.
 19. Warren S, Gates O. Multiple primary malignant tumors: a survey of the literature and a statistical study. *Am J Cancer*. 1932;16:1358–414.
 20. Hong WK, Lippman SM, Itri LM, Karp DD, Lee JS, Byers RM, et al. Prevention of second primary tumors with isotretinoin in squamous-cell carcinoma of the head and neck. *N Engl J Med*. 1990;323:795–801.
 21. Sunpaweravong S, Sunpaweravong P, Dechaphunkul T, Pongrujikorn T, Bunbanjerdasuk S, Jinawath N. Clinico-molecular study of synchronous head and neck squamous cell carcinoma (HNSCC) and esophageal squamous cell carcinoma (ESCC). *J Clin Oncol*. 2015;33:e17080.
 22. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods*. 2009;6:359–62.
 23. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem*. 2003;75:663–70.
 24. Wiśniewski JR, Zougman A, Mann M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res*. 2009;8:5674–8.
 25. Skeie JM, Mahajan VB. Proteomic interactions in the mouse vitreous-retina complex. *PLoS ONE*. 2013;8:e82140.
 26. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinformatics*. 2005;21:3940–1.
 27. Hedberg ML, Goh G, Chiosea SI, Bauman JE, Freilino ML, Zeng Y, et al. Genetic landscape of metastatic and recurrent head and neck squamous cell carcinoma. *J Clin Invest*. 2016;126:169–80.
 28. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517:576–82.
 29. Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mRNA abundance. *Cell*. 2016;165:535–50.
 30. Crutchfield CA, Thomas SN, Sokoll LJ, Chan DW. Advances in mass spectrometry-based clinical biomarker discovery. *Clin Proteomics*. 2016;13:1.
 31. Lin F, Chen Z. Standardization of diagnostic immunohistochemistry: literature review and geisinger experience. *Arch Pathol Lab Med*. 2014;138:1564–77.
 32. Meng Z, Veenstra TD. Targeted mass spectrometry approaches for protein biomarker verification. *J Proteom*. 2011;74:2650–9.
 33. Györfi B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L. Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res*. 2015;17:11.
 34. Veldman-Jones MH, Brant R, Rooney C, Geh C, Emery H, Harbron CG, et al. Evaluating robustness and sensitivity of the NanoString technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. *Cancer Res*. 2015;75:2587–93.
 35. Roesch-Ely M, Nees M, Karsai S, Ruess A, Bogumil R, Warnken U, et al. Proteomic analysis reveals successive aberrations in protein expression from healthy mucosa to invasive head and neck cancer. *Oncogene*. 2007;26:54–64.
 36. Karantza V. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene*. 2011;30:127–38.
 37. Gao F, Xia Y, Wang J, Lin Z, Ou Y, Liu X, et al. Integrated analyses of DNA methylation and hydroxymethylation reveal tumor suppressive roles of ECM1, ATF5, and EOMES in human hepatocellular carcinoma. *Genome Biol*. 2014;15:533.
 38. Danussi C, Petrucco A, Wassermann B, Modica TME, Pivetta E, Belluz LDB, et al. An EMILIN1-negative microenvironment promotes tumor cell proliferation and lymph node invasion. *Cancer Prev Res*. 2012;5:1131–43.
 39. Law E, Cheung A, Kashuba V, Pavlova T, Zabarovsky E, Lung H, et al. Anti-angiogenic and tumor-suppressive roles of candidate tumor-suppressor gene, Fibulin-2, in nasopharyngeal carcinoma. *Oncogene*. 2012;31:728–38.
 40. Pongsapich W, Jotikaprasardhna P, Lianbanchong C, Siritantikorn S, Chongkolwatana C. Human Papillomavirus Infection in Oral Cavity and Oropharyngeal Cancers: Are They the Same Story? *J Med Assoc Thai*. 2016;99:684–9.
 41. Pongsapich W, Eakkasem N, Siritantikorn S, Pithuksurachai P, Bongsabhikul K, Chongkolwatana C. Prevalence of HPV infection in hypopharyngeal and laryngeal squamous cell carcinoma at Thailand's largest tertiary referral center. *Infect Agent Cancer*. 2017;12:58.
 42. Martel M, Alemany L, Taberna M, Mena M, Tous S, Bagué S, et al. The role of HPV on the risk of second primary neoplasia in patients with oropharyngeal carcinoma. *Oral Oncol*. 2017;64:37–43.
 43. Gan SJ, Dahlstrom KR, Peck BW, Caywood W, Li G, Wei Q, et al. Incidence and pattern of second primary malignancies in patients with index oropharyngeal cancers versus index non-oropharyngeal head and neck cancers. *Cancer*. 2013;119:2593–601.
 44. Berg AO, Armstrong K, Botkin J, Calonge N, Haddow J, Hayes M, et al. Recommendations from the EGAPP Working Group: can

- tumor gene expression profiling improve outcomes in patients with breast cancer? *Genet Med.* 2009;11:66–73.
45. Mound A, Rodat-Despoix L, Bougarn S, Ouadid-Ahidouch H, Matifat F. Molecular interaction and functional coupling between type 3 inositol 1, 4, 5-trisphosphate receptor and BK Ca channel stimulate breast cancer cell proliferation. *Eur J Cancer.* 2013;49:3738–51.
 46. Shibao K, Fiedler MJ, Nagata J, Minagawa N, Hirata K, Nakayama Y, et al. The type III inositol 1, 4, 5-trisphosphate receptor is associated with aggressiveness of colorectal carcinoma. *Cell Calcium.* 2010;48:315–23.
 47. Rao RC, Dou Y. Hijacked in cancer: the MLL/KMT2 family of methyltransferases. *Nat Rev Cancer.* 2015;15:334–46.
 48. Abudurehman A, Ainiwaer J, Hou Z, Niyaz M, Turghun A, Hasim A, et al. High MLL2 expression predicts poor prognosis and promotes tumor progression by inducing EMT in esophageal squamous cell carcinoma. *J Cancer Res Clin Oncol.* 2018;144:1025–35.
 49. Gopal S, Veracini L, Grall D, Butori C, Schaub S, Audebert S, et al. Fibronectin-guided migration of carcinoma collectives. *Nat Commun.* 2017;8:14105.
 50. Chen Y, Chang J, Lee L, Wang H, Liao C, Chiu C, et al. DSG3 is overexpressed in head neck cancer and is a potential molecular target for inhibition of oncogenesis. *Oncogene.* 2007;26:467–76.
 51. Patel V, Martin D, Malhotra R, Marsh C, Doçi C, Veenstra T, et al. DSG3 as a biomarker for the ultrasensitive detection of occult lymph node metastasis in oral cancer using nanostructured immunoarrays. *Oral Oncol.* 2013;49:93–101.
 52. National Comprehensive Cancer Network. Head and neck cancers (Version 2.2018 - June 20, 2018) 2018. https://www.nccn.org/professionals/physician_gls/PDF/head-and-neck.pdf.

Affiliations

Sacarin Bunbanjersuk^{1,2} · Nutchavadee Vorasan³ · Thammakorn Saethang⁴ · Tanjitti Pongrujikhorn¹ · Duangjai Pangpunyakulchai⁵ · Narongsak Mongkonsiri⁵ · Lalida Arsa⁶ · Nintita Thokanit⁷ · Warut Pongsapich⁸ · Tauangtham Anekpuritanang⁹ · Nuttapong Ngamphaiboon¹⁰ · Artit Jinawath⁶ · Somkiat Sunpaweravong¹¹ · Trairak Pisitkun⁴ · Bhoom Suktitipat^{3,12,13} · Natini Jinawath^{1,7,13}

¹ Program in Translational Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

² Medical Genetics Center, Medical Life Sciences Institute, Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand

³ Siriraj Center of Research Excellence in Bioinformatics and Clinical Data Management, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

⁴ Center of Excellence in Systems Biology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

⁵ Immunohistopathology and Special Laboratory, Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

⁶ Molecular Histopathology Laboratory, Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

⁷ Ramathibodi Comprehensive Cancer Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

⁸ Department of Otorhinolaryngology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

⁹ Department of Pathology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

¹⁰ Medical Oncology Unit, Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

¹¹ Department of Surgery, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand

¹² Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

¹³ Integrative Computational BioScience Center (ICBS), Mahidol University, Nakhon Pathom, Thailand