Proteomics-based Predictive Model for the Early Detection of Metastasis and Recurrence in Head and Neck Cancer

ILDA PATRÍCIA RIBEIRO[1,2], LUÍSA ESTEVES[1], SANDRA ISABEL ANJO[3],
FRANCISCO MARQUES[2,4,5], LEONOR BARROSO[6], BRUNO MANADAS[3],
ISABEL MARQUES CARREIRA[1,2,7] and JOANA BARBOSA MELO[1,2,7]

[1]Cytogenetics and Genomics Laboratory, Faculty of Medicine, University of Coimbra, Coimbra, Portugal;
[2]iCBR-CIMAGO - Center of Investigation on Environment, Genetics and Oncobiology -
Faculty of Medicine, University of Coimbra, Coimbra, Portugal;
[3]CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal;
[4]Stomatology Unit, Coimbra Hospital and University Centre, CHUC, EPE, Coimbra, Portugal;
[5]Department of Dentistry, Faculty of Medicine, University of Coimbra, Coimbra, Portugal;
[6]Maxillofacial Surgery Department, Coimbra Hospital and University Centre, CHUC, EPE, Coimbra, Portugal;
[7]CNC.IBILI, Group of Aging and Brain Diseases: Advanced Diagnosis and Biomarkers, Coimbra, Portugal

Abstract. *Background/Aim: Head and neck squamous cell carcinoma (HNSCC) presents high morbidity, an overall poor prognosis and survival, and a compromised quality of life of the survivors. Early tumor detection, prediction of its behavior and prognosis as well as the development of novel therapeutic strategies are urgently needed for a more successful HNSCC management. Materials and Methods: In this study, a proteomics analysis of HNSCC tumor and non-tumor samples was performed and a model to predict the risk of recurrence and metastasis development was built. Results: This predictive model presented good accuracy (>80%) and comprises as variables the tumor staging along with DHB12, HMGB3 and COBA1 proteins. Differences at the intensity levels of these proteins were correlated with the development of metastasis and recurrence as well as with patient's survival. Conclusion: The translation of proteomic predictive models to routine clinical practice may contribute to a more precise and individualized clinical management of the HNSCC patients, reducing recurrences and improving patients' quality of life. The capability of generalization of this proteomic model to predict the recurrence and metastases development should be evaluated and validated in other HNSCC populations.*

This article is freely accessible online.

Correspondence to: Joana Barbosa de Melo, Cytogenetics and Genomics Laboratory, Faculty of Medicine, University of Coimbra, Polo Ciências da Saúde, 3000-354 Coimbra, Portugal. Tel: +351 239480027, Fax: +351 239480035, e-mail: citogenetica@fmed.uc.pt; mmelo@fmed.uc.pt

Head and neck squamous cell carcinomas (HNSCCs) represent a clinical and molecular heterogeneous group of tumors that frequently arise in the oral cavity, oropharynx, hypopharynx, and larynx (1, 2). The oral cavity, the tongue and the floor of the mouth are the most common locations of these tumors, which account for almost one-third of all oral cavity cancers (3). HNSCC tumors are frequently detected in advanced stage, being common the development of tumor recurrence and metastasis in these patients, which are responsible for poor prognosis and poor quality of life (4-6). HNSCC patients present a recurrence rate of about 50% during the first 2 years after the diagnosis of the primary tumor (7). These recurrences could be local, regional or at a distant site and encompass high rates of mortality (8). Although treatment strategies have evolved, since 1970 there have been relatively few improvements in the 5-year survival of these patients (9, 10).

Nowadays, there is an urgent need for innovative technologies and specific biomarkers to predict tumor aggressiveness, the potential for metastasis and response to treatment (11, 12). These new biomarkers could revolutionize clinical practice through the combination of clinical cancer diagnosis, screening based on proteomic analysis and histopathology (13). Proteomics represent a promising approach for identifying proteins that could help understand the underlying molecular mechanisms and signaling pathways of HNSCC. Several efforts have been made to use proteomics as a reliable approach to guide

treatment (13, 14). In this study, that aimed to establish a proteomic signature between patients without relapses or metastases and those that develop metastases or relapses, a quantitative proteomics-based approach was used to visualize major changes in protein expression in these two group of samples. Moreover, a predictive model for the early detection of metastasis and recurrence was developed based on some of these proteins.

Materials and Methods

Study population. The study protocol was approved by the Committee on Ethics in Research of the Faculty of Medicine of the University of Coimbra. All patients provided their written consent to participate in the study after being informed about the research purposes, following the regulations in the Declaration of Helsinki.

The analyzed cohort contains tumor tissue specimens from 40 HNSCC patients who underwent tumor resection. Tissue adjacent to the tumor was also collected from 32 of these patients. They were recruited between January 2011 and October 2016 from the Maxillofacial Surgery and Stomatology Units, of the Coimbra Hospital and University Centre, CHUC, EPE, Portugal. Diagnosis and staging were performed in accordance with the American Joint Committee on Cancer's TNM (tumor-node-metastasis) staging system. The follow-up periods ranged from 19 to 84 months. The detailed characterization of the cohort in the study is represented in Table I. Tissue samples were preserved in RNAlater and stored at –20˚C until use.

Proteomics sample preparation. RNAlater preserved tumor and non-tumor (adjacent to the tumor) samples were homogenized according to (15) aided by 2 steps of ultrasonication (10 min, 60% intensity, 01 s cycles; Ultrasonic processor 750W) interspersed with a 10 min incubation step at 95˚C under agitation (Eppendorf, Madrid, Spain) in the presence of glass beads (to promote the protein extraction). Samples were centrifuged to remove insoluble material and the supernatants were collected to new tubes. Samples were quantified using the 2-D Quant kit and 80 μg of total protein was used for quantitative analysis using SWATH-MS (16). Additionally, the tumor samples were pooled according to tumor location, tumor stage and the presence of metastasis (in a total of 10 pools); and the respective non-tumor samples were pooled according to the presence of metastasis (in a total of 3 pools) (17). These pools were used for protein identification. After denaturation, the samples were alkylated with acrylamide and subjected to gel digestion by using the short-GeLC approach (18). Prior to mass spectrometric analysis, the formed peptides were subjected to SPE using OMIX tips with C18 stationary phase (Agilent Technologies) as recommended by the manufacturer (17). Before protein digestion, the samples were spiked with the same amount of recombinant green fluorescent protein (MBP-GFP) to monitor samples loss during the procedure (19).

SWATH-MS analysis. Samples were analyzed on a TripleTOF™ 5600 System (AB Sciex®, Madrid, Spain) using information-dependent acquisition (IDA) of pooled samples for protein identification and SWATH-MS acquisition of each individual sample for protein quantification as presented in (18). Peptides were resolved by liquid chromatography (nanoLC Ultra 2D, Eksigent®, Madrid, Spain) on a MicroLC column ChromXP™ C18CL (300 μm

ID × 15 cm length, 3 μm particles, 120 Å pore size, Eksigent®) at 5 μL/min with a multistep gradient: 0-2 min linear gradient from 2 to 5%, 2-45 min linear gradient from 5% to 30% and, 45-46 min to 35 % of acetonitrile in 0.1% FA and 5% DMSO. Peptides were eluted into the mass spectrometer using an electrospray ionization source (DuoSpray™ Source, AB Sciex®, Madrid, Spain) with a 50 μm internal diameter (ID) stainless steel emitter (NewObjective, Woburn, MA, USA).

For IDA experiments, the mass spectrometer was set to scan full spectra (m/z 350-1250) for 250 ms, followed by up to 100 MS/MS scans (m/z 100-1500 from a dynamic accumulation time – minimum 30 ms for precursor above the intensity threshold of 1000 – in order to maintain a cycle time of 3.3 s). Candidate ions with a charge state between +2 and +5 and counts above a minimum threshold of 10 counts per second were isolated for fragmentation and one MS/MS spectrum was collected before adding those ions to the exclusion list for 25 s (mass spectrometer operated by Analyst® TF 1.7, ABSciex®). The rolling collision was used with a collision energy spread of 5.

For the SWATH-MS-based experiments, the mass spectrometer was operated in a looped product ion mode (20) and the same chromatographic conditions used as in the IDA run described above. A set of 60 windows of variable width (containing an m/z of 1 for the window overlap) was constructed covering the precursor mass range of m/z 350-1250. A 250 ms survey scan (m/z 350-1250) was acquired at the beginning of each cycle for instrument calibration and SWATH-MS/MS spectra were collected from m/z 100-1500 for 50 ms resulting in a cycle time of 3.25 s from the precursors ranging from m/z 350 to 1250. The collision energy (CE) applied to each m/z window was determined considering the appropriate CE for a +2 ion centered upon this window and the collision energy spread (CES) was also adapted to each m/z window (Table II).

A specific library of precursor masses and fragment ions was created by combining all files from the IDA experiments, and used for subsequent SWATH processing. Peptide identification and library generation were performed with ProteinPilot™ software (v5.1, AB Sciex®), using the following parameters: i) search against a database composed by Homo Sapiens from SwissProt (released June 2017) and the sequence of the recombinant protein MBP-GFP ii) acrylamide alkylated cysteines as fixed modification; iii) trypsin as digestion type. An independent False Discovery Rate (FDR) analysis, using the target-decoy approach provided by ProteinPilot™, was used to assess the quality of the identifications and confident identifications were considered when identified proteins reached a 5% local FDR (21, 22).

Data processing was performed using SWATH™ processing plug-in for PeakView™ (v2.0.01, AB Sciex®). After retention time adjustment using the MBP-GFP peptides, up to 20 peptides, with up to 5 fragments each, were chosen per protein, and quantitation was attempted for all proteins in library file that were identified from ProteinPilot™ searches. Peptides' confidence threshold was determined based on an FDR analysis using the target-decoy approach and those that met the 1% FDR threshold in at least four biological replicates were retained. The peak areas of the target fragment ions of those peptides were extracted across the experiments using an extracted-ion chromatogram (XIC) window of 4 minutes with 100 ppm XIC width. The levels of the proteins were estimated by summing all the filtered transitions from all the filtered peptides for a given protein and normalized to the total intensity obtained for the samples (19).

Table I. *Clinicopathological characteristics of the study population.*

Patients (n=40)			
	n (%)		n (%)
Gender		Differentiation	
Male	40 (100)	Well	29 (73)
Female	0	Moderate	8 (20)
Age at diagnosis (years)		Poor	3 (8)
<60	23 (58)	Margin status	
≥60	17 (43)	R0	28 (70)
Tobacco		R1	6 (15)
Yes	31 (78)	NA	6 (15)
No	6 (15)	Anatomic subsite	
NA	3 (8)	Tongue	26 (65)
Alcohol		Floor of mouth	14 (35)
Yes	29 (73)	Metastasis/	
No	5 (13)	Relapse in follow-up	
NA	6 (15)	No	22 (55)
TNM stage		Yes	18 (45)
I	6 (15)	Vital status	
II	14 (35)	Alive	24 (60)
III	9 (23)	Dead –HNSCC	10 (25)
IV	11 (28)	NA	6 (15)
Invasion peri(neural)			
Yes	20 (50)		
No	14 (35)		
NA	6 (15)		
Treatment			
Surgery alone	10 (25)		
Surgery+QT+RT	2 (5)		
Surgery+QT	0		
Surgery+RT	18 (45)		
RT+QT	4 (10)		
T alone	0		
NA	6 (15)		

NA: Not available; RT: radiotherapy; QT: chemotherapy.

Statistical analysis. The normalized levels of the proteins were analyzed using IBM SPSS Statistics 24 software for Windows (Statistical Package for the Social Sciences, IBM, USA). Initially, protein levels were normalized to the total intensity and transformed using log2. A Mann-Whitney *U*-test was implemented for the resulting proteins, taking into account the metastasis/relapse status of the patients as the grouping variable. This analysis was performed only for tumor tissue and the proteins deemed suitable were those with a minimum of 2 detected peptides. The proteins were selected based on significance, namely, those with *p*-value under 0.05 were used to construct a binary logistic regression classifier along with two categorical variables: age at initial diagnosis (<60 years and ≥60 years of age) and tumor staging (Stages I and II or III and IV combined). The variables were selected in a stepwise manner, in order to distinguish between recessive/metastatic and non-recessive/metastatic patients. This model was also used on the non-tumor samples collected from tumor-adjacent tissue. The two different tissue types were compared using a Mann-Whitney *U*-test taking into account the levels of the proteins included in the binomial logistic regression model. Survival analysis was performed using the Kaplan-Meier method.

Table II. *SWATH-MS method.*

	m/z range	Width (Da)	CES
Window 1	349.5-364.5	15	5
Window 2	363.5-379.3	15.8	5
Window 3	378.3-391.9	13.6	5
Window 4	390.9-402.2	11.3	5
Window 5	401.2-412.6	11.4	5
Window 6	411.6-422	10.4	5
Window 7	421-431.5	10.5	5
Window 8	430.5-440.5	10	5
Window 9	439.5-449.5	10	5
Window 10	448.5-457.6	9.1	5
Window 11	456.6-466.1	9.5	5
Window 12	465.1-474.2	9.1	5
Window 13	473.2-481.9	8.7	5
Window 14	480.9-489.5	8.6	5
Window 15	488.5-497.2	8.7	5
Window 16	496.2-504.4	8.2	5
Window 17	503.4-512	8.6	5
Window 18	511-519.2	8.2	5
Window 19	518.2-526.4	8.2	5
Window 20	525.4-533.6	8.2	5
Window 21	532.6-540.8	8.2	5
Window 22	539.8-548	8.2	5
Window 23	547-555.3	8.3	5
Window 24	554.3-562.5	8.2	5
Window 25	561.5-569.7	8.2	5
Window 26	568.7-576.4	7.7	5
Window 27	575.4-583.6	8.2	5
Window 28	582.6-591.3	8.7	5
Window 29	590.3-598.9	8.6	5
Window 30	597.9-607	9.1	5
Window 31	606-614.7	8.7	5
Window 32	613.7-622.3	8.6	5
Window 33	621.3-630	8.7	5
Window 34	629-637.6	8.6	5
Window 35	636.6-645.7	9.1	5
Window 36	644.7-653.4	8.7	5
Window 37	652.4-660.6	8.2	5
Window 38	659.6-668.2	8.6	5
Window 39	667.2-676.3	9.1	5
Window 40	675.3-684.9	9.6	5
Window 41	683.9-693.9	10	5
Window 42	692.9-703.3	10.4	5
Window 43	702.3-713.2	10.9	5
Window 44	712.2-724	11.8	5
Window 45	723-735.3	12.3	5
Window 46	734.3-745.6	11.3	5
Window 47	744.6-756.9	12.3	5
Window 48	755.9-768.6	12.7	5
Window 49	767.6-781.6	14	5
Window 50	780.6-796.5	15.9	5
Window 51	795.5-814	18.5	8
Window 52	813-834.7	21.7	8
Window 53	833.7-858.6	24.9	8
Window 54	857.6-884.2	26.6	8
Window 55	883.2-910.8	27.6	8
Window 56	909.8-940.9	31.1	8
Window 57	939.9-978.3	38.4	8
Window 58	977.3-1026.4	49.1	10
Window 59	1025.4-1084.9	59.5	10
Window 60	1083.9-1249.6	165.7	10

Results

Protein levels obtained by SWATH_MS were normalized, and, after initial filtering, 1,363 proteins remained in the dataset to be studied. A Mann-Whitney *U*-test was performed to distinguish the patients that developed recurrence/metastasis from those that did not, taking into account only the tumor tissue. Twenty-seven proteins with at least 2 quantified peptides were selected based on their significance ($p<0.05$).

Along with two categorical variables –the age at initial diagnosis and tumor staging– the resulting proteins were used to build a binomial logistic regression model to distinguish between recessive/metastatic and non-recessive/metastatic patients. The variables that were selected to create the classification model are the categorical variable tumor staging along with three proteins: very-long-chain 3-oxoacyl-CoA reductase (DHB12), high mobility group protein 3 (HMGB3) and collagen alpha-1(XI) chain (COBA1). The generated model could predict correctly the outcome of the patients regarding their metastatic status with 87.5% overall accuracy. The null model's accuracy was 55%.

According to the Hosmer-Lemeshow goodness of fit test (Table III), the model has a good fit for the data with a *p*-value greater than 0.05. By calculating the Nagelkerke R^2 we can verify that our model explains 51.8% of the variance in the tumors tested, meaning that the four independent variables in the logistic model, when taken together, account for 51.8% of the reason why a patient develops a relapse/metastasis or not.

The area under the ROC curve was 0.864, with a 95% CI of [0.766; 0.962], defining the ability that the model has to distinguish the patients according to their relapse/metastasis status. This means that there is an 86.4% probability that the model will correctly classify the patients into their respective class, rendering it a model with good class separation capability.

When taking the variables included in the model individually (Table IV), we can verify that none of them reached a *p*-value consistent with statistical significance ($p<0.05$). However, looking at the odds-ratio (e^B - the exponential of the regression coefficients) calculated for each of the variables there seems to be an association between them and the development of relapse/metastasis. For the tumor stage variable, an increase in tumor stage, in order words, having a stage III or IV tumor, increases in average the odds of developing a relapse or metastasis in 5.112 times. The same can be said for COBA1 (OR=1.421) – an increase on the detected level of intensity of the protein seems to lead to an increased risk for the patients to develop a relapse or metastasis. On the cases of DBH12 (OR=0.428) and HMGB3 (OR=0.528), since the odds ratios are below 1, an increase on the proteins' levels suggests a decrease on the risk of developing a relapse or metastasis.

Table III. *Pseudo Nagelkerke R squared value calculated for the model as well as the goodness of fit Hosmer and Lemeshow tests and significance.*

Nagelkerke R^2	Hosmer and Lemeshow test	
	χ^2	Significance
0.518	13.115	0.108

Table IV. *Summary of the variables in the binomial logistic regression model.*

Variable	e^B (O.R.)	Significance	95% CI for e^B	
			Lower	Upper
Tumor stage	5.112	0.055	0.964	27.120
DHB12	0.428	0.108	0.167	1.096
HMGB3	0.528	0.077	0.242	1.150
COBA1	1.421	0.089	0.948	2.130

e^B (O.R.): Odds ratio; CI: confidence interval.

It should be noted that the odds ratio is not statistically significant, since the 95% confidence intervals include 1. This would probably be solved increasing the size of the cohort in study. This model was also evaluated on the samples collected from tissue adjacent to the tumor, obtaining an 81.3% overall accuracy. Figure 1 represents the profile of protein levels across our cohort, for the three proteins included in the logistic regression model, for both types of tissue studied.

We were also interested in finding if there were differences between the type of tissue (tumor or non-tumor tissue) regarding the levels of intensity for the three proteins included in the model (Figure 2). To that purpose, we performed another Mann-Whitney *U*-test taking into account four groups: tumor-adjacent *versus* tumor tissue with or without relapse/metastasis and patients with or without relapse/metastasis within the same tissue type.

According to the Mann-Whitney *U*-test, the difference between the median protein levels of HMGB3 considering the metastatic status reached statistical significance within the tumor ($p=0.008$) and the non-tumor tissue groups ($p=0.006$), but when comparing the metastatic status across tissue types the difference was non-significant (Figure 2A). Similar results were obtained for DHB12, where we found that within the tumor tissue the difference between metastatic status was significant ($p=0.026$) as well as within the non-tumor tissue type ($p=0.002$) (Figure 2C). For COBA1, the only significant difference amongst median protein levels
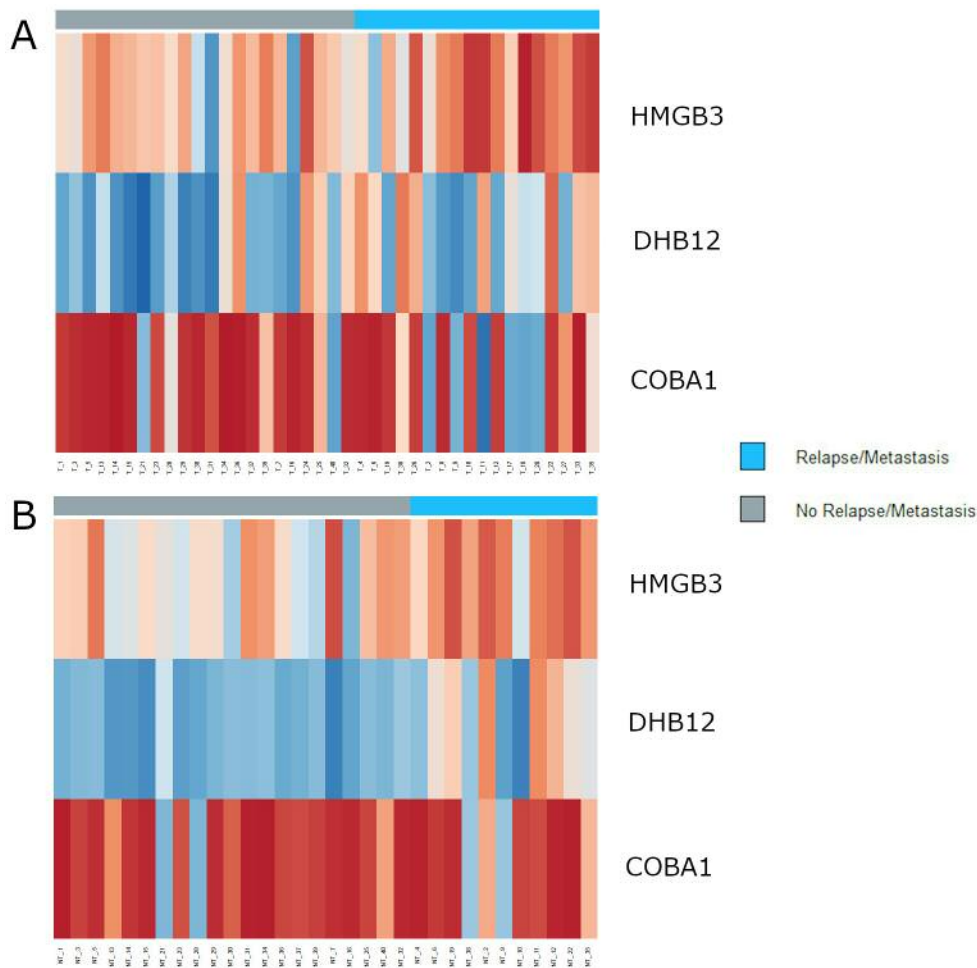
Figure 1. *Heatmaps showing the levels of intensity for the three proteins included in the logistic regression model, with the distinction between patients with relapse/metastasis (sidebar in light blue) and those without (sidebar in grey). (A) Levels of proteins HMGB3, DHB12, and COBA1 on the tumor tissues. (B) Levels of proteins HMGB3, DHB12, and COBA1 on the non-tumor tissues. Lower levels of protein intensity are represented in dark red and higher levels are represented in dark blue.*

was found between metastatic status within the tumor tissue type (*p*=0.032) (Figure 2B). These results show that, at a statistical level, regarding the profiles of HMGB3, DHB12, and COBA1, there are no differences between tissue types. However, there are differences between patients that developed a relapse or metastasis and those that did not, when considering the same tissue type.

We also wanted to evaluate the possible relevance of each of the variables included in the model on the survival of the patients, to that purpose Kaplan-Meier survival curves were obtained. For each of the proteins, two groups were established: one included the levels below the median and the other included the levels above the median protein level. For the tumor stage variable, the two groups tested were the same used to build the model (Stages I and II or III and IV combined) (Figure 3).

The mean estimated times of survival in months are represented in Table V. Since some of the curves did not reach the median estimated survival time, the mean was assumed in order to compare between groups. However, it should be noticed that, as such, these assumptions are only valid for the duration of the study.

We verified that patients that were on stages III and IV of the disease survived, on average, approximately two years and four months less than those on stages I and II. For HMGB3, the patients that had a level of protein above the median level survived, on average, one year more than the other group. In the case of DHB12, a level above the median means a mean survival time approximately 10 months higher when compared to the group that had protein levels below the median. For COBA1, levels below the median protein level mean a higher average time of survival in about one year.
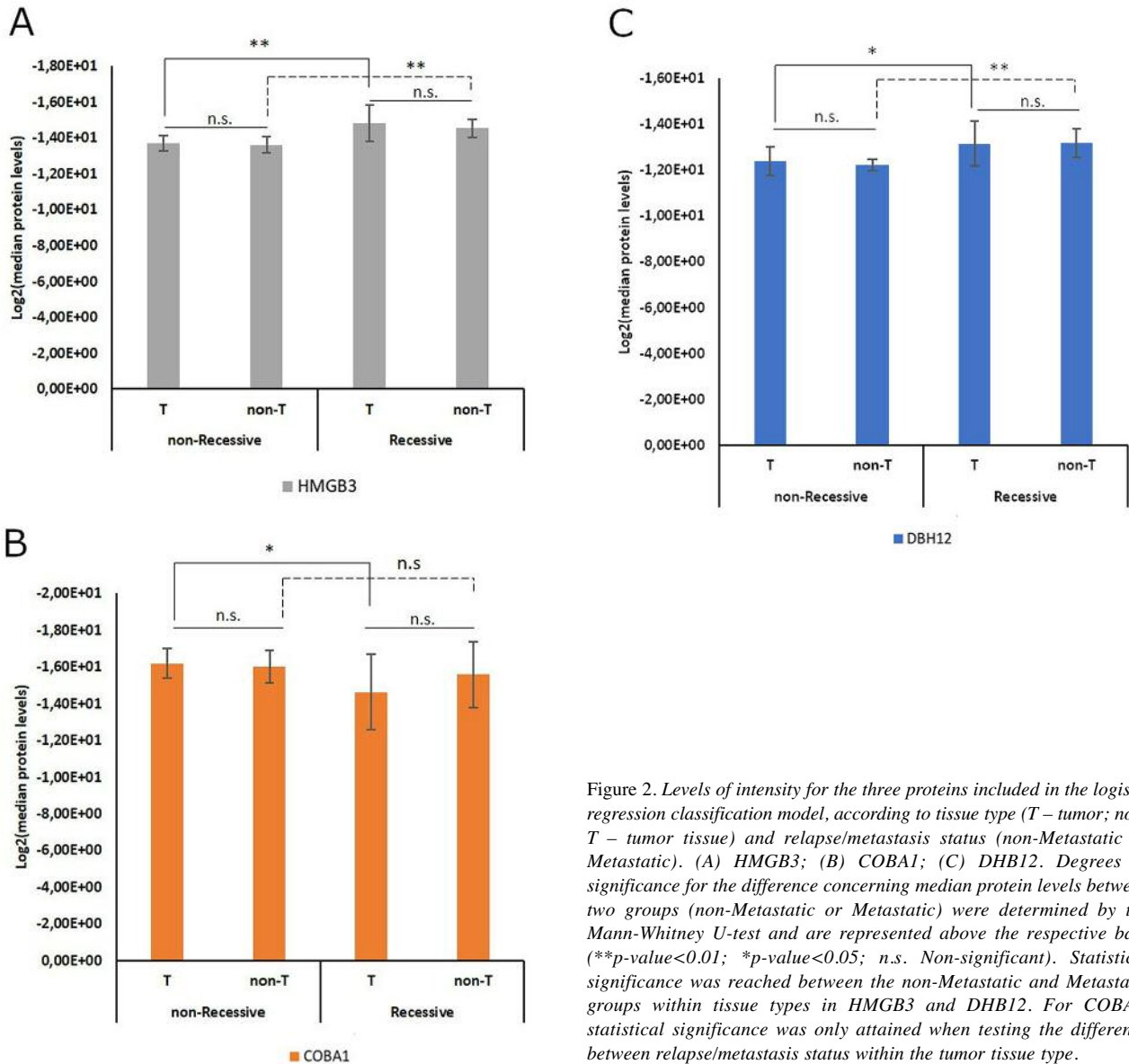
Figure 2. *Levels of intensity for the three proteins included in the logistic regression classification model, according to tissue type (T – tumor; non-T – tumor tissue) and relapse/metastasis status (non-Metastatic or Metastatic). (A) HMGB3; (B) COBA1; (C) DHB12. Degrees of significance for the difference concerning median protein levels between two groups (non-Metastatic or Metastatic) were determined by the Mann-Whitney U-test and are represented above the respective bars (**p-value<0.01; *p-value<0.05; n.s. Non-significant). Statistical significance was reached between the non-Metastatic and Metastatic groups within tissue types in HMGB3 and DHB12. For COBA1, statistical significance was only attained when testing the difference between relapse/metastasis status within the tumor tissue type.*

Discussion

Clinical cancer proteomics holds the potential to identify and quantify novel biomarkers for early detection, diagnosis, prediction of the clinical outcome and develop new therapeutic targets (23). Several genomic and proteomic biomarkers have been associated with HNSCC; however, none was successfully translated to routine clinical practice, yet (24). It is important to note that the capability to effectively treat HNSCC patients will depend upon the ability to identify if surgical margins are free of tumor, and predict tumor behavior, metastatic/recurrence potential, and response to therapy (25). In this study, we performed a global proteomic view of HNSCC samples, specifically in tumor and non-tumor tissue samples from tongue and floor of the mouth tumors, as an attempt to discover new biomarkers of early diagnosis of metastasis and relapses. We quantified more than one thousand proteins with different levels of intensities between the patients that develop metastasis or recurrence and those without metastasis or recurrence, which revealed the great biological complexity that underlies HNSCC. Moreover, these tumors display clinical heterogeneity, being their behavior unpredictable using only the available set of clinical markers; therefore, the development of a prognosis predictive model is a novel and promising strategy to increase the HNSCC survival rate and
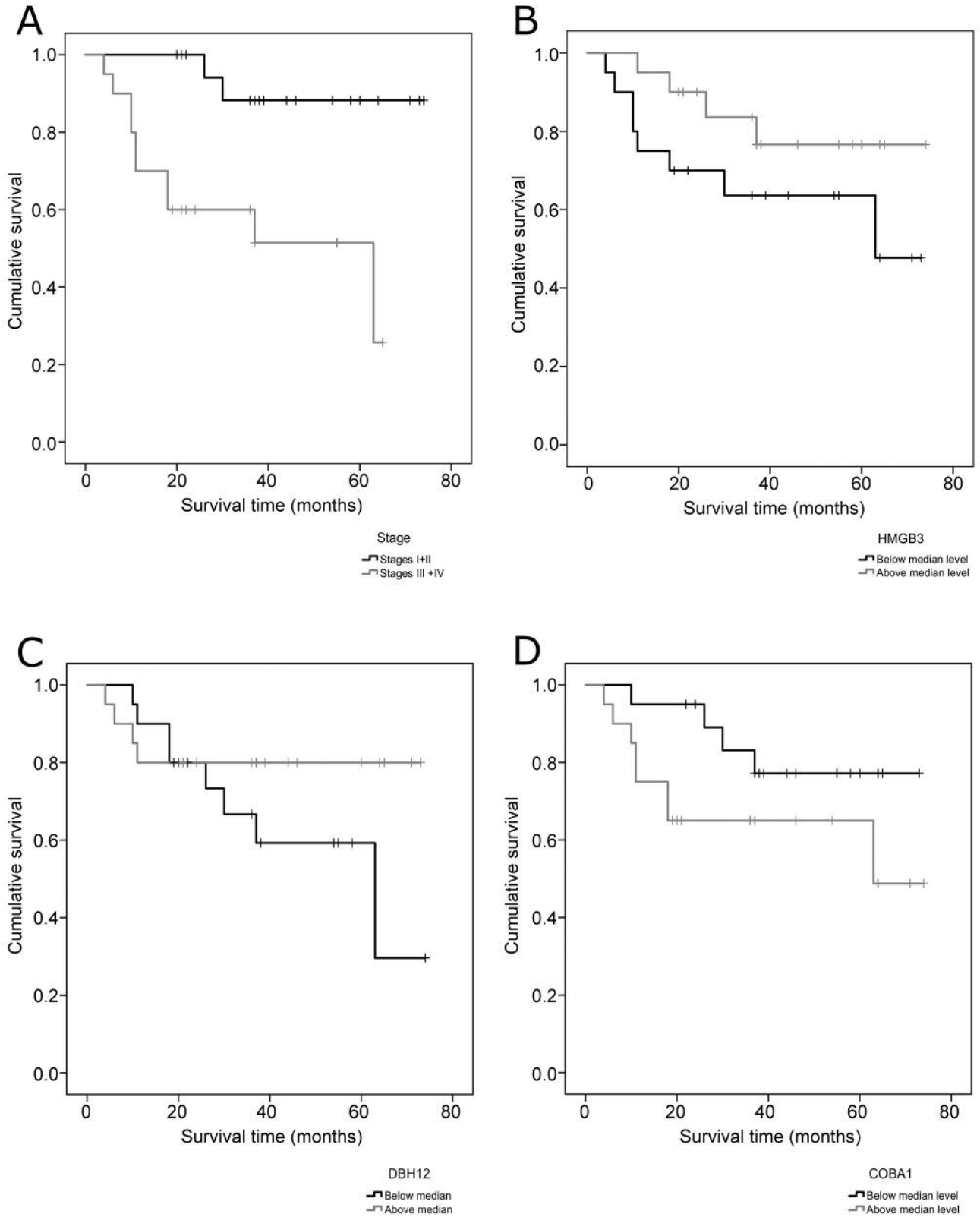
Figure 3. *Kaplan-Meier survival curves for the four variables included in the logistic regression model. (A) Clinical Stage; (B) HMGB3; (C) DBH12; (D) COBA1.*

Table V. *Mean estimated survival time in months and respective standard errors, for each of the four variables included in the logistic regression model, for the duration of the study.*

		Mean survival time estimate (months)	Standard error (months)
Stage	I+II	69	3.598
	III+IV	40	5.843
HMGB3	<Median	50	6.506
	>Median	62	5.160
DHB12	<Median	50	6.031
	>Median	60	5.843
COBA1	<Median	62	4.753
	>Median	50	6.628

improve the quality of life of the patients (4). The proteomic signature identified in the present study was used to build a predictive classification model of recurrence and metastasis development. This developed multivariate model was evaluated on the tumor and non-tumor samples presenting an average accuracy higher than 85 and 80%, respectively, and includes a categorical variable tumor staging along with DHB12, HMGB3, and COBA1proteins.

We verified that high levels of DBH12 and HMGB3 proteins could be, in our cohort, associated with a decreased risk of HNSCC patients to develop relapse or metastasis. On the other hand, high levels of COBA1 protein seem to be linked to an increased risk for these patients to develop relapse or metastasis. As expected, HNSCC patients with a higher tumor stage (III or IV) have an increased risk of relapse and metastasis development.

Considering the patients' survival, high level of DHB12 and HMGB3 and low level of COBA1 proteins were associated with higher survival time. As anticipated, we verified that patients on III and IV tumor stages survived less than those on I and II tumor stages. Altogether, these results suggest a specific set of three proteins that seem to have an important role in the development and prediction of HNSCC recurrence and metastasis.

To the best of our knowledge this is the first study to highlight these three proteins together with tumor stage as a predictor of recurrence or metastases development in tongue and floor of the mouth tumors, which could represent a significant step forward in the paradigm shift of clinical management of these patients as well as in the successful implementation of precision medicine.

The protein DHB12, encoded by the gene *HSD17B12* mapped at 11p11.2, is considered a multifunctional enzyme that performs the conversion of estrone to estradiol and elongate long-chain fatty acids (26). In HNSCC high expression of HSD17B12 mRNA is pointed out as a predictor

of metastasis in a four-gene model with the capability of successful prediction of 77% (27). Moreover, DHB12 seems also to play a role in breast cancer progression (28). In the presented study, our results showed an association between high DHB12 protein levels and decreased risk of metastasis and recurrence development as well as with better survival.

The protein HMGB3 is a member of a family of proteins that hold one or more high mobility group DNA-binding motifs and plays an important role in DNA replication, recombination, and repair (29). HMGB3 seems to play an important role in a wide variety of cancers, such as breast cancer, gastric cancer, bladder cancer, colorectal cancer, and esophageal squamous cell carcinoma. In breast cancer, the mRNA and protein levels of HMGB3 were higher in the tumor compared to adjacent benign specimens and exhibited an indirect correlation between the expression of HMGB3 mRNA and patient survival (30). Likewise, high-level expression of HMGB3 protein was detected in gastric adenocarcinoma cells and its overexpression was correlated with a poor prognosis, being postulated that HMGB3 may promote gastric cancer cell proliferation by regulating the cell cycle (31). Urinary bladder cancer patients with higher HMGB3 expression showed a poorer overall survival rate than those with low HMGB3 (p=0.0079, logrank test), indicating its potential role in tumor proliferation and migration (32). In colorectal cancer, HMGB3 was up-regulated at both mRNA and protein levels, exercising oncogene effect to promote cell growth and migration (33). Also, esophageal squamous cell carcinoma patients with higher HMGB3 expression have a shorter overall survival than those with lower HMGB3 expression (34). Despite, HMGB3 was reported to be overexpressed in a variety of human cancers, our results for tongue and floor of the mouth tumors found that high levels of HMGB3 protein was associated with low risk of relapse and metastasis and consequently with better survival, which seems to be opposite to the results presented for other cancer types. However, the role of HMGB3 in human cancers remains still relatively unclear. Recently, Mukherjee *et al.* demonstrated *in vitro* that HMGB3 plays a role in sensitizing cisplatin-resistant ovarian cancer cells to cisplatin treatment, probably via the transcriptional repression and deregulation of the ATR-CHK1 damage signaling pathway (35).

The collagen type XI α1 (*COL11A1*) gene encodes one of the two α chains of type XI collagen, a minor fibrillar collagen (36), that is located on chromosome 1 (1p21.1) and regulates the promoter of the major collagen type V (COL5A2) (37). Collagen is a major component of the extracellular matrix and is involved in the regulation of cell proliferation, differentiation, and migration (38). Increased expression of *COL11A1* gene was described to promote tumor progression in HNSCC (39), having this gene already shown differential expression between metastatic and nonmetastatic oral cavity/oropharynx squamous cell carcinoma (fold change=7.61; p=0.002) (40). Additionally,

Shen *et al*. found that expression of *COL11A1* was up-regulated in metastatic and recurrent non-small cell lung cancer (41). These results are in agreement with our proteomic findings for tongue and floor of the mouth tumors, where high levels of the protein collagen alpha-1(XI) chain (COBA1) were associated with an increased risk for HNSCC patients develop relapse or metastasis and consequently associated with worse survival, reinforcing the importance of this protein in the pathogenesis of HNSCC recurrences and metastases.

In our predictive model, high levels of DBH12 and HMGB3 that are proteins related to long-chain fatty acids elongation, conversion of estrone to estradiol, DNA replication and repair are correlated with a decreased risk of relapse or metastasis development as well as high levels of COBA1 protein, a component of the extracellular matrix, involved in the regulation of cell proliferation, differentiation and migration are correlated with an increased risk of relapse or metastasis development. Further studies must be conducted to clarify the role of the identified proteins in HNSCC pathogenesis as well as to confirm our observations in a larger clinical series. The clinical application of this classification model is promising since it is possible to identify newly diagnosed HNSCC patients with risk of recurrence/metastasis development and, consequently, monitor them closely, avoiding or performing early detection of the recurrences and use more aggressive and personalized treatments. Moreover, we highlighted in this study three proteins that could be candidates for targeted therapy studies.

In conclusion, considering that HNSCC patients exhibit poor overall prognosis and survival with a high risk to recur after the primary tumor diagnosis and treatment, the proteomic predictive model for recurrence and metastasis development that we propose may pave the way to more precise patient management and targeted drug design.

Conflicts of Interest

The Authors have no conflicts of interest to disclose.

Authors' Contributions

I.P.R. and S.I.A. performed the experiments. I.P.R. wrote the manuscript. L.E. performed the statistical analysis. F.M. and L.B. collected the samples and performed the critical interpretation of the data from the clinical point of view. B.M., I.M.C. and J.B.M. contributed to conception and design of the study and critically revised the manuscript. All Authors read and approved the final manuscript.

Acknowledgements

References

1 Descamps G, Wattiez R and Saussez S: Proteomic study of hpv-positive head and neck cancers: Preliminary results. Biomed Res Int 2014: 430906, 2014. PMID: 3955617. DOI: 10.1155/2014/430906

2 Cimpean AM, Balica RA, Doros IC, Balica NC, Gaje PN, Popovici RA and Raica M: Epidermal growth factor receptor (egfr) and keratin 5 (k5): Versatile keyplayers defining prognostic and therapeutic sub-classes of head and neck squamous cell carcinomas. Cancer Genomics Proteomics *13(1)*: 75-81, 2016. PMID: 26708602.

3 Sciubba JJ: Oral cancer. The importance of early diagnosis and treatment. Am J Clin Dermatol *2(4)*: 239-251, 2001. PMID: 11705251. DOI: 10.2165/00128071-200102040-00005

4 Ribeiro IP, Caramelo F, Esteves L, Menoita J, Marques F, Barroso L, Migueis J, Melo JB and Carreira IM: Genomic predictive model for recurrence and metastasis development in head and neck squamous cell carcinoma patients. Sci Rep *7(1)*: 13897, 2017. PMID: 5654944. DOI: 10.1038/s41598-017-14377-x

5 Ribeiro IP, Marques F, Caramelo F, Pereira J, Patricio M, Prazeres H, Ferrao J, Juliao MJ, Castelo-Branco M, de Melo JB, Baptista IP and Carreira IM: Genetic gains and losses in oral squamous cell carcinoma: Impact on clinical management. Cell Oncol (Dordr) *37(1)*: 29-39, 2014. PMID: 24353162. DOI: 10.1007/s13402-013-0161-5

6 Ribeiro IP, Caramelo F, Marques F, Domingues A, Mesquita M, Barroso L, Prazeres H, Juliao MJ, Baptista IP, Ferreira A, Melo JB and Carreira IM: Wt1, msh6, gata5 and pax5 as epigenetic oral squamous cell carcinoma biomarkers - a short report. Cell Oncol (Dordr) *39(6)*: 573-582, 2016. PMID: 27491556. DOI: 10.1007/s13402-016-0293-5

7 Argiris A, Karamouzis MV, Raben D and Ferris RL: Head and neck cancer. Lancet *371(9625)*: 1695-1709, 2008. PMID: 18486742. DOI: 10.1016/S0140-6736(08)60728-X

8 Takes RP, Baatenburg de Jong RJ, Schuuring E, Hermans J, Vis AA, Litvinov SV and van Krieken JH: Markers for assessment of nodal metastasis in laryngeal carcinoma. Arch Otolaryngol Head Neck Surg *123(4)*: 412-419, 1997. PMID: 9109790. DOI: 10.1001/archotol.1997.01900040048008

9 Marur S, D'Souza G, Westra WH and Forastiere AA: HPV-associated head and neck cancer: A virus-related cancer epidemic. Lancet Oncol *11(8)*: 781-789, 2010. PMID: 5242182. DOI: 10.1016/S1470-2045(10)70017-6

10 Yonezawa K, Nishiumi S, Kitamoto-Matsuda J, Fujita T, Morimoto K, Yamashita D, Saito M, Otsuki N, Irino Y, Shinohara M, Yoshida M and Nibu K: Serum and tissue metabolomics of head and neck cancer. Cancer Genomics Proteomics 10(5): 233-238, 2013. PMID: 24136976.

11 Harris TM, Du P, Kawachi N, Belbin TJ, Wang Y, Schlecht NF, Ow TJ, Keller CE, Childs GJ, Smith RV, Angeletti RH, Prystowsky MB and Lim J: Proteomic analysis of oral cavity squamous cell carcinoma specimens identifies patient outcome-associated proteins. Arch Pathol Lab Med *139(4)*: 494-507, 2015. PMID: 4380852. DOI: 10.5858/arpa.2014-0131-OA

12 Karpathiou G, Stachowitz ML, Dumollard JM, Gavid M, Froudarakis M, Prades JM and Peoc'h M: Gene expression comparison between the primary tumor and its lymph node metastasis in head and neck squamous cell carcinoma: A pilot study. Cancer Genomics Proteomics *16(3)*: 155-161, 2019. PMID: 6542642. DOI: 10.21873/cgp.20121

13 Matta A, DeSouza LV, Shukla NK, Gupta SD, Ralhan R and Siu KW: Prognostic significance of head-and-neck cancer biomarkers previously discovered and identified using itraq-labeling and multidimensional liquid chromatography-tandem mass spectrometry. J Proteome Res *7(5)*: 2078-2087, 2008. PMID: 18407684. DOI: 10.1021/pr7007797

14 Balog J, Sasi-Szabo L, Kinross J, Lewis MR, Muirhead LJ, Veselkov K, Mirnezami R, Dezso B, Damjanovich L, Darzi A, Nicholson JK and Takats Z: Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. Sci Transl Med *5(194)*: 194ra193, 2013. PMID: 23863833. DOI: 10.1126/scitranslmed.3005623

15 Bennike TB, Kastaniegaard K, Padurariu S, Gaihede M, Birkelund S, Andersen V and Stensballe A: Proteome stability analysis of snap frozen, rnalater preserved, and formalin-fixed paraffin-embedded human colon mucosal biopsies. Data Brief *6(*942-947, 2016. PMID: 4753390. DOI: 10.1016/j.dib.2016.01.061

16 Anjo SI, Santa C and Manadas B: Swath-ms as a tool for biomarker discovery: From basic research to clinical applications. Proteomics *17(3-4)*, 2017. PMID: 28127880. DOI: 10.1002/pmic.201600278

17 Anjo SI, Santa C, Saraiva SC, Freitas K, Barah F, Carreira B, Araújo I and Manadas B: Neuroproteomics using short gelc-swath: From the evaluation of proteome changes to the clarification of protein function. *In*: Current proteomic approaches applied to brain function. Humana Press: New York, NY, *pp*. 107-138, 2017. DOI: 10.1007/978-1-4939-7119-0_8

18 Anjo SI, Santa C and Manadas B: Short gelc-swath: A fast and reliable quantitative approach for proteomic screenings. Proteomics *15(4)*: 757-762, 2015. PMID: 25418953. DOI: 10.1002/pmic.201400221

19 Anjo SI, Simões I, Castanheira P, Grãos M and Manadas B: Use of recombinant proteins as a simple and robust normalization method for untargeted proteomics screening: Exhaustive performance assessment. Talanta *205:* 120163, 2019. PMID: 31450411. DOI: 10.1016/j.talanta.2019.120163

20 Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R and Aebersold R: Targeted data extraction of the ms/ms spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. Mol Cell Proteomics *11(6)*: O111.016717, 2012. PMID: 3433915. DOI: 10.1074/mcp.O111.016717

21 Tang WH, Shilov IV and Seymour SL: Nonlinear fitting method for determining local false discovery rates from decoy database searches. J Proteome Res *7(9)*: 3661-3667, 2008. PMID: 18700793. DOI: 10.1021/pr070492f

22 Sennels L, Bukowski-Wills JC and Rappsilber J: Improved results in proteomics by use of local and peptide-class specific false discovery rates. BMC Bioinformatics *10:* 179, 2009. PMID: 2709624. DOI: 10.1186/1471-2105-10-179

23 Matta A, Ralhan R, DeSouza LV and Siu KW: Mass spectrometry-based clinical proteomics: Head-and-neck cancer biomarkers and drug-targets discovery. Mass Spectrom Rev *29(6)*: 945-961, 2010. PMID: 20945361. DOI: 10.1002/mas. 20296

24 Sudha R, Kawachi N, Du P, Nieves E, Belbin TJ, Negassa A, Angeletti RH and Prystowsky MB: Global proteomic analysis distinguishes biologic differences in head and neck squamous carcinoma. Lab Invest *87(8)*: 755-766, 2007. PMID: 17558418. DOI: 10.1038/labinvest.3700598

25 Belbin TJ, Bergman A, Brandwein-Gensler M, Chen Q, Childs G, Garg M, Haigentz M, Hogue-Angeletti R, Moadel R, Negassa A, Owen R, Prystowsky MB, Schiff B, Schlecht NF, Shifteh K, Smith RV and Zheng X: Head and neck cancer: Reduce and integrate for optimal outcome. Cytogenet Genome Res *118(2-4)*: 92-109, 2007. PMID: 18000360. DOI: 10.1159/000108290

26 Moeller G and Adamski J: Integrated view on 17beta-hydroxysteroid dehydrogenases. Mol Cell Endocrinol *301(1-2)*: 7-19, 2009. PMID: 19027824. DOI: 10.1016/j.mce.2008.10.040

27 Rickman DS, Millon R, De Reynies A, Thomas E, Wasylyk C, Muller D, Abecassis J and Wasylyk B: Prediction of future metastasis and molecular characterization of head and neck squamous-cell carcinoma based on transcriptome and genome analysis by microarrays. Oncogene *27(51)*: 6607-6622, 2008. PMID: 19027824. DOI: 10.1038/onc.2008.251

28 Song D, Liu G, Luu-The V, Zhao D, Wang L, Zhang H, Xueling G, Li S, Desy L, Labrie F and Pelletier G: Expression of aromatase and 17beta-hydroxysteroid dehydrogenase types 1, 7 and 12 in breast cancer. An immunocytochemical study. J Steroid Biochem Mol Biol *101(2-3)*: 136-144, 2006. PMID: 16930994. DOI: 10.1016/j.jsbmb.2006.06.015

29 Stros M: Hmgb proteins: Interactions with DNA and chromatin. Biochim Biophys Acta *1799(1-2)*: 101-113, 2010. PMID: 20123072. DOI: 10.1016/j.bbagrm.2009.09.008

30 Elgamal OA, Park JK, Gusev Y, Azevedo-Pouly AC, Jiang J, Roopra A and Schmittgen TD: Tumor suppressive function of mir-205 in breast cancer is linked to hmgb3 regulation. PLoS One *8(10)*: e76402, 2013. PMID: 3788717. DOI: 10.1371/journal.pone.0076402

31 Tang HR, Luo XQ, Xu G, Wang Y, Feng ZJ, Xu H, Shi YW, Zhang Q, Wu LG, Xue CQ, Wang CW and Wu CY: High mobility group-box 3 overexpression is associated with poor prognosis of resected gastric adenocarcinoma. World J Gastroenterol *18(48)*: 7319-7326, 2012. PMID: 3544037. DOI: 10.3748/wjg.v18.i48.7319

32 Li M, Cai Y, Zhao H, Xu Z, Sun Q, Luo M, Gu L, Meng M, Han X and Sun H: Overexpression of hmgb3 protein promotes cell proliferation, migration and is associated with poor prognosis in urinary bladder cancer patients. Tumour Biol *36(6)*: 4785-4792, 2015. PMID: 25647262. DOI: 10.1007/s13277-015-3130-y

33 Zhang Z, Chang Y, Zhang J, Lu Y, Zheng L, Hu Y, Zhang F, Li X, Zhang W and Li X: Hmgb3 promotes growth and migration in colorectal cancer by regulating wnt/beta-catenin pathway. PLoS One *12(7)*: e0179741, 2017. PMID: 5497964. DOI: 10.1371/journal.pone.0179741

34 Gao J, Zou Z, Gao J, Zhang H, Lin Z, Zhang Y, Luo X, Liu C, Xie J and Cai C: Increased expression of hmgb3: A novel independent prognostic marker of worse outcome in patients with esophageal squamous cell carcinoma. Int J Clin Exp Pathol *8(1)*: 345-352, 2015. PMID: 25755721.

35 Mukherjee A, Huynh V, Gaines K, Reh WA and Vasquez KM: Targeting the high mobility group box 3 protein sensitizes chemoresistant ovarian cancer cells to cisplatin. Cancer Res, 2019. PMID: 31061066. DOI: 10.1158/0008-5472.CAN-19-0542

36 Mio F, Chiba K, Hirose Y, Kawaguchi Y, Mikami Y, Oya T, Mori M, Kamata M, Matsumoto M, Ozaki K, Tanaka T, Takahashi A, Kubo T, Kimura T, Toyama Y and Ikegawa S: A functional polymorphism in col11a1, which encodes the alpha 1 chain of type xi collagen, is associated with susceptibility to lumbar disc herniation. Am J Hum Genet *81(6)*: 1271-1277, 2007. PMID: 2276353. DOI: 10.1086/522377

37 Yoshioka H, Greenwel P, Inoguchi K, Truter S, Inagaki Y, Ninomiya Y and Ramirez F: Structural and functional analysis of the promoter of the human alpha 1(xi) collagen gene. J Biol Chem *270(1)*: 418-424, 1995. PMID: 7814404. DOI: 10.1074/jbc.270.1.418

38 Prockop DJ, Kivirikko KI, Tuderman L and Guzman NA: The biosynthesis of collagen and its disorders (first of two parts). N Engl J Med *301(1)*: 13-23, 1979. PMID: 449904. DOI: 10.1056/NEJM197907053010104

39 Sok JC, Lee JA, Dasari S, Joyce S, Contrucci SC, Egloff AM, Trevelline BK, Joshi R, Kumari N, Grandis JR and Thomas SM: Collagen type xi alpha1 facilitates head and neck squamous cell cancer growth and invasion. Br J Cancer *109(12)*: 3049-3056, 2013. PMID: 3859935. DOI: 10.1038/bjc. 2013.624

40 Schmalbach CE, Chepeha DB, Giordano TJ, Rubin MA, Teknos TN, Bradford CR, Wolf GT, Kuick R, Misek DE, Trask DK and Hanash S: Molecular profiling and the identification of genes associated with metastatic oral cavity/pharynx squamous cell carcinoma. Arch Otolaryngol Head Neck Surg *130(3)*: 295-302, 2004. PMID: 15023835. DOI: 10.1001/archotol.130.3.295

41 Shen L, Yang M, Lin Q, Zhang Z, Zhu B and Miao C: Col11a1 is overexpressed in recurrent non-small cell lung cancer and promotes cell proliferation, migration, invasion and drug resistance. Oncol Rep *36(2)*: 877-885, 2016. PMID: 27373316. DOI: 10.3892/or.2016.4869