Data Mining Identifies Six Proteins that Can Act as Prognostic Markers for Head and Neck Squamous Cell Carcinoma

Zeng-hong Wu¹,², Yun-Tang³, and Qing Cheng¹

Abstract
Head and neck squamous cell carcinoma (HNSCC) is a malignant tumor of the upper aerodigestive tract affecting the oral cavity, lips, paranasal sinuses, larynx, and nasopharynx. Proteogenomics combines proteomics and genomics and employs mass spectrometry and high-throughput sequencing technologies to identify novel peptides. The aim of this study was to identify potential protein biomarkers for clinical treatment of HNSCC. To achieve this, we utilized two sets of data, one on proteins from The Cancer Proteome Atlas (TCPA) and the other on gene expression from The Cancer Genome Atlas (TCGA) database, to evaluate dysfunctional proteogenomics microenvironment. Univariate Cox regression analysis was performed to examine the relationship between protein signatures and prognosis. A total of 19 proteins were significantly associated with overall survival (OS) of patients, of which E2F transcription factor 1 (E2F1; HR = 4.557, 95% CI = 1.810 to 11.469) and enhancer of zeste homolog 2 (EZH2; HR = 0.430, 95% CI = 0.187 to 0.984) were the most differentially expressed between patients with longer and shorter OS, respectively. Furthermore, multivariate Cox regression analysis on six proteins (ERALPHA, HER3, BRAF, P27, RAPTOR, and E2F1) was performed to build the prognostic model. The receiver operating characteristic curves were used to determine whether the expression pattern of survival-related proteins could provide an early prediction of the occurrence of HNSCC. Herein, we found an AUC of 0.720. Based on an online database, we identified novel protein markers for the prognosis of HNSCC. The findings of the present study may provide new insights into the development of new and reliable tools for precise cancer intervention.

Keywords
head and neck squamous cell carcinoma, proteomics, TCPA, TCGA, prognosis

Introduction
Head and neck squamous cell carcinoma (HNSCC) is a malignant tumor of the upper aerodigestive tract affecting the oral cavity, lips, paranasal sinuses, larynx, and nasopharynx. Currently, it is the sixth most prevalent cancer globally and accounts for more than 650,000 new cases and 90,000 deaths annually². Known risk factors for the disease include smoking cigarettes, infection with human papillomavirus (HPV), and alcohol consumption³. One of the main challenges in HNSCC treatment is late diagnosis because symptoms tend to occur in the advanced stages of the disease. Consequently, the 5-yr survival rate of individuals with HNSCC is less than 50% and can drop to less than 35% in patients that develop local recurrence or metastasis⁴. Once in the advanced stages, treatment interventions can affect the structure and function of organs involved in speech, taste, smell, swallowing, as well as other vital organs of the body. These eventualities reduce the overall quality of life of the affected individuals⁵,⁶.

Proteogenomics is a field of research that combines proteomics and genomics and utilizes mass spectrometry and high-throughput sequencing technologies to identify novel peptides. Mass spectrometry–based matching or

¹ Department of Otorhinolaryngology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China
² Department of Infectious Diseases, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
³ Department of Critical Care Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Submitted: February 28, 2020. Revised: April 27, 2020. Accepted: April 30, 2020.

Corresponding Author:
Qing Cheng, Department of Otorhinolaryngology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China.
Email: 15971507750@163.com
identification of homologous peptides provides evidence of gene expression at protein level. As such, it allows in-depth analysis of interactions between host and respective pathogens. Combining advanced genomics with proteomics has the potential of revealing useful biomarkers that can drive the development of new diagnostic techniques and therapies. Many studies have explored the role of proteogenomics in assessing the prognosis of cancers, but even so, the mechanism with which tumors progress remains unclear. Ma et al. found that C-C motif chemokine ligand protein mediates progression and angiogenesis of Epstein Barr virus (EBV)–mediated nasopharyngeal carcinoma. Moreover, patients with oral squamous cell carcinoma exhibit overexpressed CD44 molecule S100A7, and S100P proteins. Consistently, these proteins are considered as potential biomarkers for early detection of malignant tumors. A few integrative tools for complete proteogenomics analyses have been developed; however, there are currently no high-throughput technologies for evaluating protein biomarkers that can predict the overall survival (OS) or response to HNSCC immunotherapy. Because high-throughput expression data are available, it is feasible to use global information on gene expression to analyze the association between genomics and prognoses of tumors. In this study, we utilized two sets of data, one on proteins, from The Cancer Proteome Atlas (TCPA), and the other on gene expression, from The Cancer Genome Atlas (TCGA) database, to evaluate dysfunctional proteogenomics microenvironment with an aim of identifying useful protein biomarkers for clinical treatment of HNSCC.

Materials and Methods

Data Collection

Data was collected from TCPA, which contained a protein profile of 8,167 tumor samples. These proteins were primarily reflective of the gene expression of genetic maps in TCGA database. They provide a unique opportunity to validate the reliability of data in TCGA and identify model cell lines for functional research. TCGA has generated many platforms for data on cancer genomics as well as proteomics using the Reverse Phase Protein Array, which measures the level of about 150 proteins and 50 phosphoproteins in tumors. Data on proteomics was therefore downloaded from TCPA (level 4), together with gene expression profiles from TCGA.

Establishing Prognostic Related Gene Signatures

Univariate Cox regression analysis was used to identify the prognostic genes and establish genetic characteristics. The gene signature was calculated as risk score = (CoefficientmRNA1 × expression of mRNA1) + (CoefficientmRNA2 × expression of mRNA2) + ... + (CoefficientmRNA n × expression mRNA n). Based on the median, each parameter was categorized into a low-risk and a high-risk group. Kaplan–Meier
survival analysis was performed to compare the survival rate between the study group and the control group.

**Establishing a Predictive Nomogram**

Nomograms are often used to determine the prognosis of cancer because they can simplify statistical prediction models and provide individualized numerical assessment of the likelihood of an event (such as relapse or death). In the present study, a receiver operating characteristic (ROC) curve was plotted to evaluate the accuracy of the proteins as prognostic markers for HNSCC. Univariate and multifactorial Cox regression analyses were performed to compare the association between the expression of specific proteins and clinicopathological variables. Data with missing values were deleted from the survey. These data were analyzed using R packages in R (version 3.5.3 http://www.r-project.org/). Statistical significance regarding Perl language for data matrix and all data processing was set at $P < 0.05$.

**Results**

**Prognostic Protein Signatures**

Data on proteomics were first retrieved from TCPA. The corresponding genomic data on clinical survival was downloaded from TCGA database. Univariate Cox regression analysis identified 19 proteins that were significantly associated with OS of patients, of which $E2F1$ protein was identified ($E2F$ transcription factor 1; hazard ratio (HR) = 4.657, 95% confidence interval (CI) = 1.810 to 11.469) and $EZH2$ protein was identified (enhancer of zeste homolog 2; HR = 0.430, 95% CI = 0.187 to 0.984) were the most differentially expressed between patients with longer and shorter OS, respectively (Table S1). The volcano plot was drawn to visualize the expression profile of survival-related proteins. The red and green dots indicate high-risk
and low-risk proteins, respectively (Fig. 1A). Multivariate Cox regression analysis on the six proteins Estrogen receptor alpha (ERALPHA), Erb-b2 receptor tyrosine kinase 3 (HER3), B-Raf proto-oncogene serine/threoninekinase (BRAF), Cyclin-dependent kinase inhibitor p27 (P27), mTOR regulation-related protein (RAPTOR), and E2F transcription factor 1 (E2F1) was performed to build the prognostic model. The formula for our model was: Risk Score = 0.84 × ERALPHA + 0.89 × HER3 + (−0.64) × BRAF + (−0.86) × P27 + 1.79 × RAPTOR + 1.33 × E2F1. Significant differences in expression levels of ERALPHA (P = 2.934e−12), HER3 (P = 3.779e−07), BRAF (P = 7.187e−08), P27 (P = 1.059e−06), RAPTOR (P = 2.694e−11), and E2F1 (P = 6.311e−12) were found between the high-risk and the low-risk groups.

**Survival Results and Multivariate Examination.** OS analysis showed that the expression of high-risk proteins was associated with better prognosis of HNSCC compared with low-risk proteins (P < 0.01; Fig. 1B). Also, the comprehensive prognostic signature model of the HNSCC revealed that HNSCC patients expressing high-risk proteins exhibited lower mortality rates and high OS (Fig. 2). Cox regression analysis showed that our prognostic model was an independent factor for OS (Fig. 3). The ROC curves were used to determine whether the expression pattern of survival-related proteins could provide an early prediction for the occurrence of HNSCC. We found an area under roc curve (AUC) of 0.720, which indicated that the sensitivity and specificity of this prognostic model were moderate. Clinicopathological factors associated with HNSCC were also explored. The

**Fig. 3.** The results of Cox regression analysis indicate that our prognostic model is an independent prognostic factor for overall survival.
AUC of age, gender, grade, stage, tumor (T), metastasis (M), and node (N) were 0.546, 0.476, 0.526, 0.639, 0.578, 0.444, and 0.698, respectively (Fig. 4).

Creating Predictive Nomogram. A nomogram was created by combining clinical factors and the prognostic model. The final model comprised seven factors captured in TCGA, which included age, gender, grade, stage, T, M, and N (Fig. 5). Combining the prognostic model with clinical factors improved the accuracy of predicting OS for 1, 2, and 3 yr. Analysis of co-expression of the six proteins identified 24 negatively correlated and 68 positive correlated proteins (Table S2).

Discussion

Analysis of cancer proteomics provides new insights into changes that occur in the early stages of tumorigenesis, and therefore assist in the identification of novel biomarkers for early disease diagnosis and prognosis. Tumor cells present unique proteomic characteristics regarding their growth and survival compared to normal cells. These unique proteomic signaling pathways can be used as prognostic markers for the occurrence and development of cancer. Also, they can act as therapeutic targets in cancer treatment. Herein, we identified new and reliable prognostic proteomics signatures based on online datasets. These proteomic profiles are likely to manifest in patients with HNSCC and thus can be utilized as biomarkers and targets for therapeutic interventions.

In the present study, proteomic data was downloaded from TCPA (level 4), together with clinical survival data from TCGA. In total, 19 survival-related proteins were identified. The expression of E2F1 protein correlated with high OS. E2F1 (also known as E2F-1, RBAP1, RBBP3, RBP3) is a member of the E2F transcription factor family. The E2F family plays a vital role in controlling the cell cycle and inhibiting tumor-enhancing proteins. A recent study found that E2F1 activates IncRNA small nucleolar RNA host gene 3 (SNHG3), thereby promoting the migration and proliferation of cells in individuals with non-small-cell lung cancer by transforming the growth factor–β pathway13. Similarly, therapeutic targets that suppress the myeloid zinc finger 1 (MZF1) and MZF1 antisense RNA 1 (MZF1-AS1)/poly(ADP-ribose) polymerase 1 (PARP1)/E2F1 axis can inhibit the synthesis of proline and progression of neuroblastoma14. A study identified miRNA-205 as a putative HNSCC oncogene and regulator of E2F1 protein expression15. Notably, HNSCC caused by HPV infection expresses E6 and E7 viral...
oncogenes. These proteins suppress tumor protein p53 (TP53) and retinoblastoma 1 (RB1) and activate E2F1, a cell cycle regulator. At present, the role of E2F1 in HNSCC has not been fully elucidated. Given this, we aimed to explore the relationship between the expression of E2F1 and the occurrence and progression of HNSCC. EZH2 and the catalytic component of Polycomb inhibition complex 2 are associated with the expression of homologous genes (Hox) and early stages of inactivating X chromosome. Overexpression of EZH2 is thus associated with high incidence and poor prognosis of tumors. Maintaining the stability of mitochondrial membrane potential in HNSCC is dependent on EZH2. Studies have shown that signal transducer and activator of transcription 3 (STAT3) signaling through overexpression/activation of phosphotidylinositol-3-kinase (PI3K) regulates HOX transcript antisense RNA (HOTAIR) and EZH2 in HNSCC, thus promoting the progression of the disease.

The potential of utilizing the pattern of survival-related proteins in early prediction of future occurrence of HNSCC was examined using the ROC curve. The AUC was 0.720. We examined six proteins (ERalpha, HER3, BRAF, P27, RAPTOR, and E2F1) in the multivariate Cox regression analysis to build the prognostic model. Estrogen receptor genes (ERalpha and ERbeta) play an important role in the progression of tumors, particularly breast cancer. Notably, the ERbeta gene has been shown to play a more significant role than ERalpha in the etiology of colorectal cancer. In breast cancer, the expression of ERbeta alone is an independent risk factor for poor prognosis of patients undergoing tamoxifen adjuvant therapy. The expression of ERalpha and ERbeta in HNSCC upregulates the estrogen receptor ligands, thus activating transcription and transduction of cytoplasmic signals. Also, ERalpha was shown to be associated with epidermal growth factor receptor (EGFR) in HNSCC cells treated with dasatinib. HER3 (Erb-b2 receptor tyrosine kinase 3) is a member of the human EGFR family and consists of four type-1 transmembrane tyrosine kinase receptors: HER1 to HER4. The upregulation of HER3 is associated with poor survival rates in various malignancies, including HNSCC and breast, colorectal, gastric, and prostate cancer. Studies have also found that upregulating HER3 using cetuximab hinders EGFR-mediated inhibition of HNSCC. The small nuclear ribonucleoprotein polypeptide E (B-Raf) proto-oncogene serine/threonine kinase (BRAF) encodes a protein belonging to the serine/threonine-protein kinase family. Mutation in BRAF is a marker for response to respective targeted therapy in patients with HNSCC. Therefore, factors that promote BRAF mutation drive oncogenesis. Accordingly, BRAF mutations are frequent in thyroid cancer. P27 is a cyclin-dependent kinase inhibitor and plays an essential role in the negative regulation of the cell cycle. Overexpression of p27 inhibits the proliferation of cancer cells, increases apoptosis, and arrests cell cycle in HNSCC. The expression of p27 can be used to predict how HNSCC responds to chemotherapy.
Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD
Zeng-hong Wu https://orcid.org/0000-0003-4977-6478

Supplemental Material
Supplemental material for this article is available online.

References
1. Xiao N. Comprehensive analysis of competitive endogenous RNAs network associated with head and neck squamous cell carcinoma. Sci Rep. 2018;8(1):10544.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jamal A. Global cancer statistics 2012. CA Cancer J Clin. 2012;65(2):87–108.
3. Magnes T, Egle A, Greil R, Melchardt T. Update on squamous cell carcinoma of the head and neck: ASCO annual meeting. Memo. 2017;10(4):220–223.
4. Chin D, Boyle GM, Williams RM, Ferguson K, Pandey N, Pedley J, Campbell CM, Theile DR, Parsons PG, Coman WB. Novel markers for poor prognosis in head and neck cancer. Int J Cancer. 2005;113(5):789–797.
5. Bressan V, Stevanin S, Bianchi M, Aleo G, Bagnasco A, Sasso L. The effects of swallowing disorders, dysgeusia, oral mucositis and xerostomia on nutritional status, oral intake and weight loss in head and neck cancer patients: a systematic review. Cancer Treat Rev. 2016;45(5):105–119.
6. Coskun HH, Medina JE, Robbins KT, Silver CE, Stevanin S, Bianchi M, Aleo G, Bagnasco A, Sasso L. Metabolomic approach in head and neck cancer: a systematic review. Cancer Treat Rev. 2016;45(5):105–119.
7. Menschaut G, Vandekerckhove TTM, Baggerman G, Schoofs L, Luyten W, Van Criekinge W. Peptidomics Coming of age: a review of contributions from a bioinformatics angle. J Proteome Res. 2010;9(5):2051–2061.
8. Ma W, Feng L, Zhang S, Zhang H, Zhang X, Qi X, Zhang Y, Feng Q, Xiang T, Zeng Y-X. Induction of chemokine (C-C motif) ligand 5 by Epstein-Barr virus infection enhances tumor angiogenesis in nasopharyngeal carcinoma. Cancer Sci. 2018;109(5):1710–1722.
9. Sivadasan P, Gupta MK, Sathe G, Sudheendra HV, Sunny SP, Renu D, Hari PS, Gowda H, Suresh A, Kuriakose MA, Siddeeshwara R. Salivary proteins from dysplastic leukoplakia and oral squamous cell carcinoma and their potential for early detection. J Proteomics. 2020;10(212):103574.
10. Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, Yang JY, Broom BM, Verhaak RGW, Kane DW, Wakefield C, et al. TCPA: a resource for cancer functional proteomics data. Nat Met. 2013;10(11):1046–1047.
11. Akbani R, Ng PKS, Werner HMJ, Shahnoradgoli M, Zhang F, Ju Z, Liu W, Yang J-Y, Yoshihara K, Li J, Ling S, et al. A pancancer proteomic perspective on the cancer genome atlas. Nat Commun. 2014;5(6):3887.
12. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364–1370.
13. Shi J, Li J, Yang S, Hu X, Chen J, Feng J, Shi T, He Y, Mei Z, He W, Xie J, et al. LncRNA SNHG3 is activated by E2F1 and promotes proliferation and migration of no small cell lung cancer cells through activating TGF-β pathway and IL/JAK2/STAT3 pathway. J Cell Physiol. 2019;235(3):2891–2900.
14. Fang E, Wang X, Yang F, Hu A, Wang J, Li D, Song H, Hong M, Guo Y, Liu Y, Li H, et al. Therapeutic targeting of MZF1 || S1 /PARP1/E2F1 Axis inhibits proline synthesis and neuroblastoma progression. Advanced Sci. 2019;6(19):1900581.
15. Alla V, Kowtharapu BS, Engelmann D, Emmrich S, Schmitz U, Steder M, Pützer BM. E2F1 confers antitumor drug resistance by targeting ABC transporter family members and Bcl-2 via the p73/DNp73-miR-205 circuitry. Cell Cycle. 2012;11(16):3067–3078.
16. Beck TN, Golemis EA. Genomic insights into head and neck cancer. Can Head Neck. 2016;1(2):1.
17. Schuttengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. Cell. 2007;128(4):735–745.
18. Zhou X, Ren Y, Kong L, Cai G, Sun S, Song W, Wang Y, Jin R, Qi L, Mei M, Wang X, et al. Targeting EZH2 regulates tumor growth and apoptosis through modulating mitochondria dependent cell-death pathway in HNSCC. Oncotarget. 2015;6(32):33720–33732.
19. Sun S, Wu Y, Guo W, Yu F, Kong L, Ren Y, Wang Y, Yao X, Jing C, Zhang C, Liu M, et al. STAT3/HOTAIR signaling axis regulates HNSCC growth in an EZH2-dependent manner. Clin Cancer Res. 2018;24(11):2665–2677.
20. Slattery ML, Sweeney C, Murtaugh M, Ma KN, Wolff RK, Potter JD, Caan BJ, Samowitz W. Associations between ERalpha, ERbeta, and AR genotypes and colon and rectal cancer. Cancer Epidemiol Biomarkers Prev. 2005;14(12):2936–2942.
21. Gruvberger-Saal SK, Benda hl P-O, Saal LH, Laakso M, Gruvberger B, Apholz M, Brattstrom B, Borg A, Fernö M. Estrogen receptor beta expression is associated with tamoxifen response in ERα-negative breast carcinomas. Breast Cancer Res Treat. 2015;16(3):43.
22. Egloff AM, Rothstein ME, Seethala R, Siegfried JM, Grandis JR, Stabile LP. Cross-talk between estrogen receptor and epithelial growth factor receptor in head and neck squamous cell carcinoma. Clin Cancer Res Official J Am Assoc Cancer Res. 2007;13(7):1987.
23. Egloff AM, Rothstein ME, Seethala R, Siegfried JM, Grandis JR, Stabile LP. Cross-talk between estrogen receptor and epithelial growth factor receptor in head and neck squamous cell carcinoma. Clin Cancer Res. 2005;11(21):6529–6540.
24. Lin Y-C, Wu M-H, Wei T-T, Chuang S-H, Chen K-F, Cheng A-L, Chen C-C. Degradation of epidermal growth factor receptor induces dasatinib-induced apoptosis in head and neck squamous cell carcinoma cells. Neoplasia. 2012;14(6):463–475.
25. Zhang J, Saba NF, Chen GZ, Shin DM. Targeting HER (ERBB) signaling in head and neck cancer: an essential update. Mol Aspects Med. 2015;45(10):74–86.
26. Slamon DJ, Clark GM. Amplification of c-erbB-2 and aggressive human breast tumors? Sci. 1988;240(4860):1795–1798.
27. Maurer CA, Friess H, Kretschmann B, Zimmermann A, Stauffer A, Baer HU, Korc M, Buchler MW. Increased
expression of erbB3 in colorectal cancer is associated with concomitant increase in the level of erbB2. Hum Pathol. 1998;29(8):771–777.

27. Wang D, Qian G, Zhang H, Magliocca KR, Nannapaneni S, Amin ARMR, Rossi M, Patel M, El-Deiry M, Wadsworth JT, Chen Z, et al. HER3 targeting sensitizes HNSCC to cetuximab by reducing HER3 activity and HER2/HER3 dimerization - evidence from cell line and patient derived xenograft models. Clin Can Res. 2017;23(3):677–686.

28. Weber A, Langhanki L, Sommerer F, Markwarth A, Wittekind C, Tannapfel A. Mutations of the BRAF gene in squamous cell carcinoma of the head and neck. Oncogene. 2003;22(30):4757–4759.

29. Robb RN, Yang L, Williams T. Abstract 5186: Vemurafenib selectively radiosensitizes BRAF V600E mutant papillary and anaplastic thyroid carcinoma cells. Cancer Research. 2017;77(13 Supplement):5186–5186.

30. Park SY, Lee DW. Effects of p27 overexpression on head and neck squamous cell carcinoma cell lines. Head Neck. 2005;37(8):1187–1192.

31. Moreno GC, Hermson M, Garcia P, Juana M. p27 and BCL2 expression predicts response to chemotherapy in head and neck squamous cell carcinomas. Oral Oncology. 2014;50(2):128–134.

32. Yonezawa K, Tokunaga C, Oshiro N. Raptor, a binding partner of target of rapamycin. Biochemi Biophys Res Commun. 2004;313(2):0–441.

33. Marsh CA, Patel V, Dinali ME. Abstract LB-7: HPV-associated HNSCC: widespread mTOR pathway activity and preclinical evaluation of mTOR inhibitors rapamycin and RAD-001. Cancer Res. 2011;71(8 Supplement):LB–7-LB–7.

34. Coppock JD, Vermeer PD, Vermeer DW. mTOR inhibition as an adjuvant therapy in a metastatic model of HPV+ HNSCC. Oncotarget. 2016;7(17):24228–24241.