A gene expression-based immune content predictor for survival and postoperative radiotherapy response in head and neck cancer

Yingqin Li,1,2 Xiaohong Hong,1,2 Yuan Zhang,1,2 Yingqing Li,1,2 Yuan Lei,1 Qingmei He,1 Xiaojing Yang,1 Yelin Liang,1 Jun Ma,1 and Na Liu1

1Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou, Guangdong 510060, China

The immune infiltration in the tumor microenvironment has been demonstrated to be relevant to radiotherapy response. Here, we sought to understand the immune infiltration in head and neck cancer (HNC) and evaluate its significance in predicting prognosis and radiotherapy response. Using RNA sequencing data of 522 retrospective head and neck squamous cell carcinomas (HNSCCs), we constructed an immune content score based on genes related to 6 prognostic infiltrating cell types. Unsupervised hallmark pathway clustering demonstrated an immune-related tumor cluster containing the immune content score. Patients with high immune content scores exhibited favorable overall survival and disease-free survival (DFS). Moreover, the immune content score was an independent prognostic factor for DFS in HNSCC. Interestingly, the immune content score was strongly associated with radiation response pathways. These results were also extended to nasopharyngeal carcinoma. Furthermore, patients in the low immune content score group significantly gained overall survival benefits from postoperative radiotherapy (PORT), whereas patients in the high immune content score group did not. Therefore, this study identifies the immune content score as a prognostic tool, which might have a potential association with PORT response, thereby facilitating outcome prediction and treatment decision in HNC.

INTRODUCTION

Head and neck cancer (HNC) is the sixth most common cancer with global incidence.1 Because of the aggressiveness of local invasion and metastasis, >50% of all patients with HNC are initially diagnosed at a locally advanced stage.2 Radiotherapy is one of the major treatment modalities for patients with HNC, and postoperative radiotherapy (PORT) is routinely used for patients with resectable, locally advanced head and neck squamous cell carcinomas (HNSCCs).3–6 Although radiotherapy effectively improves patient survival in HNC, the TNM stage-dependent outcomes remain heterogeneous, and some may suffer from serious radiation toxicity without benefit.7–9 Thus, there is an urgent need to identify novel biomarkers to predict benefits from radiotherapy to improve survival and avoid overtreatment toxicity for patients with HNC.

Tumor initiation, progression, and metastasis have been demonstrated to depend on the bidirectional interactions between tumor cells and the surrounding environment, which form the tumor microenvironment (TME).10,11 The immune content, determined by the types of tumor-infiltrating immune cells, can yield information that is relevant to prognosis and treatment response.12,13 Parallel studies have emphasized the prognostic role of the immune content and specific immune cell types in HNC.2,8,14 For example, total tumor-infiltrated lymphocytes (TILs) and CD8+ T cell (TC) infiltration have been reported to be associated with survival and response to chemoradiotherapy in HNC.15–17 However, the immune content landscapes and their impacts on radiotherapy remain unclear in HNC.

Within the past decade, rapid technical and bioinformatics advances have enabled the comprehensive molecular characterization of tumors using gene expression profiling (GEP).18,19 Based on immune-specific genes or expression signatures, tumor-infiltrating immune cells can be characterized and quantified from bulk tumor GEP data with bioinformatics methods.20–22 CIBERSORT, a computational approach to deconvolute GEP data, allows high-sensitivity discrimination of immune infiltrates within tumors.23,24 This method estimates the relative proportion of various immune cell types, offering an opportunity to reanalyze available genomic data to...
explore the correlation between tumor-infiltrating cells and clinical outcomes.25–29

Here, we applied the computational algorithm CIBERSORT to draw a map of infiltrated immune cell types in HNC. The immune content score was constructed based on genes related to 6 prognostic infiltrating cell types and positively correlated with favorable survival in both the HNSCC and nasopharyngeal carcinoma (NPC) cohorts. Moreover, PORT was preferentially beneficial in patients with low immune content score. Therefore, this study identifies the immune content score as a prognostic tool, which might be able to predict response to radiotherapy, thereby facilitating precise treatment decisions in HNC.

RESULTS
CIBERSORT-based computation of immune content associated with survival
CIBERSORT analysis revealed the immune infiltrated cell composition in the HNSCC cohort (Table S1). When related to clinical outcomes, we identified 6 infiltrated immune cell types with different proportions between patients with good and poor survival. Cases with good prognosis exhibited higher fractions of CD8+ TCs, activated memory CD4+ TCs, follicular helper TCs, resting natural killer (NK) cells, and plasma cells (PCs), whereas cases with poor prognosis showed higher fractions of M0-type macrophages (p < 0.05, Table S2).

Development and validation of a prognostic immune content score model
Using single-sample gene set enrichment analysis (ssGSEA) analysis, we constructed an immune content model by calculating the scores according to 160 genes representing the 6 prognosis-associated infiltrating cell types. Notably, the calculated immune content scores could be clustered with all the immune-related hallmark pathways (inflammatory response, interferon alpha, interferon gamma, interleukin [IL]-2, IL-6, and the complement pathway) (Figure 1A). Although NPC was a distinct type of HNC characterized with different epidemiology, etiology, and treatment strategies, a similar pattern of clustering between immune content scores and immune infiltrated cell composition was observed in the NPC cohort (Figure 1B).

We then evaluated the clinical implications of the immune content score by correlating with clinicopathological characteristics and outcomes. As shown in Table 1, the immune content score was significantly correlated with N stage (p = 0.031), human papillomavirus (HPV) infection (assessed by immunohistochemistry p16 status, p = 0.043), and primary tumor site (p = 0.005) in the HNSCC cohort. Survival analysis showed that patients with a high immune content score had significantly favorable overall survival (OS; 5-year OS, 52.4% versus 41.6%, p = 0.046) and disease-free survival (DFS; 3-year DFS, 66.2% versus 56.3%, p = 0.017) compared with patients with a low immune content score in HNSCC (Figures 1C and 1D). In multivariate analysis, the immune content score was retained as an independent prognostic factor for DFS in HNSCC (hazard ratio [HR] 0.433, 95% confidence interval [CI] 0.193–0.975, p = 0.043, Table S3).

In the NPC cohort, there was no significant association between the immune content score and clinical stage or mutation burden (p > 0.05, Table S4). Consistently, an increased immune content score was markedly associated with better DFS (3-year DFS, 88.7% versus 65.4, p = 0.028, Figure 1E), and the immune content score remained an independent prognostic factor for DFS in the NPC cohort (HR 0.268, 95% CI 0.074–0.966, p = 0.044, Table S5). Taken together, these results suggest that the prognostic value of the immune content score could be extended from HNSCC to NPC.

Immune infiltrated cell composition in patients with different immune content scores
In both the HNSCC and NPC cohorts, the proportions of resting memory CD4+ TCs and M0-type macrophages were >10% of immune infiltrated cells (Figure 2A; Table S1). In addition, the proportion of activated memory CD4+ TCs was most positively correlated with the proportion of CD8+ TCs (Pearson correlation = 0.53) in HNSCC patients, whereas the resting memory CD4+ TC fraction was most negatively correlated with the CD8+ TC fraction (Pearson correlation = −0.56) in NPC patients (Figure 2B).

Furthermore, we explored the immune cell profiles according to the immune content score. In the HNSCC cohort, the percentages of 14 infiltrated cell types were different between patients with low and high immune content scores (p < 0.05, Figure 2C). Cases with high immune content scores exhibited significantly increased proportions of naive and memory B cells (BCs), PCs, CD8+ TCs, activated memory CD4+ TCs, follicular helper TCs, M0-type macrophages, and decreased proportions of follicular helper TCs, M0-type macrophages, activated dendritic cells (DCs), activated mast cells, and eosinophils. Similarly, the fractions of naive and memory BCs were significantly upregulated and the fractions of follicular helper TCs and activated DCs were downregulated in the high immune content score group in the NPC cohorts (p < 0.05, Figure 2C).

Immune content score predicts response to postoperative radiotherapy
We performed GSEA to evaluate which biological pathways are correlated with immune content score across the HNSCC and NPC cohorts. Although immune-related pathways were only a small minority of the 5,815 pathways examined, 19 of the top 25 gene sets were enriched in immune-related pathways in the HNSCC cohort and 22 of the top 25 gene sets were enriched in the NPC cohort (Figures 3A and 3B). Our immune content model was constructed with genes representing TCs, NK cells, PCs, and macrophages, but other immune pathways were some of the most enriched in HNSCC and NPC cases. Thus, the immune content score may reflect individual immune status.
GSEA analysis also showed that the immune content score was significantly correlated with the radiation response pathways in both the HNSCC (GO_Response_To_Gamma_Radiation and GO_Response_To_Radiation) and NPC (TSAI_Response_To_Ionizing_Radiation and RASHI_Response_To_Ionizing_Radiation) cohorts, with false discovery rates (FDRs) < 0.10 and p values < 0.05 (Figures 3C and 3D). In clinical practice, PORT is one of the primary treatment strategies in HNSCC. Then, we evaluated the clinical value of the immune content score in the prediction of response to PORT.

To balance treatment selection bias, two paired HNSCC cohorts with or without PORT were generated by propensity score matching (PSM). As shown in Table S6, no significant difference was observed in clinicopathological characteristics between patients with and without PORT treatment. In the low immune content score group, patients who received PORT had significantly longer survival than patients who did not receive PORT (5-year OS, 64.7% versus 33.6%, p = 0.030, Figure 3E). In contrast, patients with a high immune content score did not benefit from PORT (p > 0.05, Figure 3E). Furthermore, multivariate analysis showed that PORT remained independently prognostic for DFS in the low immune content score group (HR 0.226, 95% CI 0.06–0.851, p = 0.028, Table S7) but not in the high immune content score group (p > 0.05). Taken together, these findings suggest that our prognostic model using the immune content score may stratify patients who are likely to respond to PORT.

**DISCUSSION**

In this retrospective cohort-based study, we constructed an immune content score model with prognostic and predictive value in forecasting HNC patient response to PORT. Our results showed that patients with a low immune content score had significantly poorer survival than patients with a high immune content score. Furthermore, our model showed that low-score patients received benefit from PORT, suggesting that the model is a promising strategy for identifying eligible candidates for PORT in HNC.
To date, the tumor immune infiltrates have been mainly investigated by immunohistochemistry, immunofluorescence, or flow cytometry, with a limited repertoire of phenotypic markers. The rapid advances in high-throughput technology have motivated its application in routine oncology settings and generated an unprecedented amount of GEP data. Based on immune-specific genes or expression signatures, tumor-infiltrating immune cells can be characterized and quantified from bulk tumor GEP data with bioinformatics methods, such as the CIBERSORT deconvolution algorithm. Here, we revealed the composition of immune infiltrated cells from two public HNC RNA sequencing (RNA-seq) datasets using CIBERSORT and constructed an immune content model by calculating the gene scores of survival-related infiltrated cell types. Moreover, the immune content score was significantly associated with survival, which is consistent with the findings in prostate, bladder, and breast cancer.

Table 1. Correlation between immune content score and clinicopathological characteristics in HNSCC cohort

| Variable            | Low immune content score group (n = 261) | High immune content score group (n = 261) | p value |
|---------------------|------------------------------------------|-------------------------------------------|---------|
| Median age (IQR)    | 61 (52–68)                               | 61 (54–69)                                | 0.895   |
| Gender              |                                          |                                           |         |
| female              | 58 (22.6)                                | 78 (29.9)                                | 0.059   |
| male                | 202 (77.4)                               | 183 (70.1)                               |         |
| Smoking             |                                          |                                           |         |
| non-smoker          | 51 (20.0)                                | 66 (25.9)                                | 0.114   |
| former and current smoker | 204 (80.0)                        | 189 (74.1)                               |         |
| Alcohol history     |                                          |                                           |         |
| no                  | 78 (31.0)                                | 85 (32.8)                                | 0.651   |
| yes                 | 174 (69.0)                               | 174 (67.2)                               |         |
| T stage             |                                          |                                           |         |
| T1–T2               | 96 (36.9)                                | 103 (39.9)                               | 0.483   |
| T3–T4               | 164 (63.1)                               | 155 (60.1)                               |         |
| N stage             |                                          |                                           |         |
| N0–N1               | 176 (67.7)                               | 151 (58.5)                               | 0.031*  |
| N2–N3               | 84 (32.3)                                | 107 (41.5)                               |         |
| TNM stage           |                                          |                                           |         |
| I–II                | 61 (24.1)                                | 57 (22.4)                                | 0.539   |
| III–IV              | 192 (75.9)                               | 198 (77.6)                               |         |
| HPV status p16 IHC  |                                          |                                           |         |
| negative            | 35 (76.1)                                | 38 (57.6)                                | 0.043*  |
| positive            | 11 (23.9)                                | 28 (42.4)                                |         |
| HPV status ISH      |                                          |                                           |         |
| negative            | 27 (87.1)                                | 38 (69.1)                                | 0.062   |
| positive            | 4 (12.9)                                 | 17 (30.9)                                |         |
| Primary tumor site  |                                          |                                           |         |
| oral cavity         | 163 (62.5)                               | 153 (58.6)                               | 0.005*  |
| oropharynx          | 27 (10.3)                                | 53 (20.3)                                |         |
| larynx and hypopharynx | 71 (27.2)                           | 55 (21.1)                                |         |

Values are presented as n (%), except for median age (IQR). HPV, human papillomavirus; IHC, immunohistochemistry; IQR, interquartile range; ISH, in situ hybridization. *p < 0.05; p value was determined by χ² or Fisher’s exact tests. +Variables included in this table had <20% of available values except for HPV p16 IHC and HPV status ISH.

In the past decade, the TME has been recognized to play a central role in the initiation and progression of tumors. Tumor-infiltrating immune cells in the TME have been demonstrated as potential biomarkers for clinical outcome and therapeutic responses. Therefore, quantification of tumor-infiltrating immune cell types holds promise for unveiling the multifaceted role of the immune system in HNC. To date, the tumor immune infiltrates have been mainly investigated by immunohistochemistry, immunofluorescence, or flow cytometry, with a limited repertoire of phenotypic markers. The rapid

The GSEA revealed that the immune content score was significantly related to radiation response pathways in both the HNSCC and NPC cohorts. Clinically, radiotherapy after surgery is routinely recommended for locoregionally advanced HNSCC patients who are at high risk of tumor recurrence. Although radiotherapy effectively improves patient survival in HNSCC, the stage-dependent outcomes are heterogeneous with PORT. Novel strategies are thus urgently needed to change from uniform stage-dependent treatment to biomarker-guided treatment selection for PORT in HNSCC. Thus, we evaluated whether the immune content score could predict benefit from PORT. The results showed that patients with low immune content scores could gain a survival benefit from PORT, whereas patients with high immune content scores did not. Therefore, these findings suggest that patients with high immune content scores are at low risk and may have a good prognosis when treated with surgery alone, and PORT brings little benefit but does add unnecessary radiation-related toxicities. Alternatively, other therapeutic strategies, such as targeted therapy or immunotherapy, may be another good option for the patients in the no-benefit group. Meanwhile, patients with low scores have poor prognoses and might be recommended for PORT, since they would gain survival and disease control benefits from PORT.

Our study had several limitations. The primary limitation of this work is that the estimation of infiltrated immune cell types and immune...
Figure 2. The landscape of immune infiltrated cells in HNC

(A) Immune infiltrated cells in the whole cohort (upper left) and individual patients in the HNSCC (bottom) and NPC (upper right) cohorts. (B) Correlation matrix of all immune cell proportions in the HNSCC (left) and NPC (right) cohorts. Variables have been ordered by average linkage clustering. (C) The difference in immune infiltrated cells between patients with low and high immune content scores in the HNSCC (left) and NPC (right) cohorts. The group with low immune content score is marked in blue, and the group with high immune content score is marked in red.
content is computationally predicted. To further validate the clinical value of the immune content score, multicenter and prospective studies are required. Moreover, our immune content score was constructed based on genes related to 6 infiltrating cell types, including CD8+ TCs, activated memory CD4+ TCs, follicular helper TCs, resting NK cells, PCs, and M0-type macrophages. The fundamental mechanisms underlying the infiltrated immune cell types and radiation response remain unknown in HNC, which deserves further experimental exploration.

In conclusion, this work illustrates the comprehensive immune landscape in HNC. Clinically, we developed an immune content score model that could effectively predict prognosis and PORT benefit in HNC. This model might aid clinicians in identifying patients who will benefit from PORT, facilitating the development of individual therapeutic strategies.

**MATERIALS AND METHODS**

**Gene expression datasets**
We retrospectively included 635 HNC human samples with publicly available RNA-seq data and clinical characteristics (age, sex, stage, survival, treatment, and so on) in our study. A HNSCC cohort of 522 samples was obtained from The Cancer Genome Atlas (TCGA) data portal (level 3 data, https://www.cancer.gov/tcga/, accessed...
March 27, 2019). For the NPC cohort, a dataset of 113 samples was collected from the Gene Expression Omnibus (GEO, GSE102349). Genes expressed in at least 50% of the samples were retained for analyses. Additional details about each cohort can be found in the original publications.

CIBERSORT analysis and the immune content score
A CIBERSORT-based deconvolution of data from 522 HNSCC samples was carried out according to the CIBERSORT instructions (https://cibersort.stanford.edu) using a 547-gene signature matrix customized by normalizing referenced microarray data (Affymetrix) to distinguish 22 immune infiltrated cell types. CIBERSORT was used to estimate the relative proportion of different immune cell types. According to overall survival, HNSCC cases were stratified into “good” (alive) and “poor” (death) prognosis subgroups, and the differences in the proportions of infiltrating immune cell types between groups were analyzed. Then, 160 genes from the differentiated infiltrated cytotypes were employed to calculate the immune content score using ssGSEA (GenePattern module, https://cloud.genepattern.org). Then, we assigned patients with high and low immune content scores according to median value. Hallmark pathway lists were downloaded from the Molecular Signatures Database (MSigDB, http://software.broadinstitute.org/gsea/msigdb/), and pathway scores were generated by computing gene expression levels within each pathway via ssGSEA. Scores were hierarchically clustered via heatmap analysis.

Gene set enrichment analysis
The gene expression profiles of HNSCC and NPC datasets were used to conduct GSEA to identify gene signatures between groups with different immune content scores. We employed the MSigDB collection of the Hallmark, C2 Canonical pathways (C2: CP), and C5 GO Biological Process (C5: GO-BP) gene sets (n = 5,815) and the GSEA algorithm to rank correlation coefficients for each gene according to the immune content score. GSEA results were visualized by normalized enrichment scores (p < 0.05 and FDR < 0.25) and an enrichment heatmap with the top 100 ranked genes (50 with positive score and 50 with a negative score) between groups.

Propensity score matching analysis
A PSM algorithm was applied to balance bias and resample two paired HNSCC cohorts with or without PORT. The propensity score was calculated for each patient weighted by clinicopathological characteristics, including age, sex, T stage, N stage, smoking, and primary tumor site. Since only 112 of the 522 patients in the HNSCC cohort have HPV infection status information (accessed by immunohistochemistry p16 status), the factor of HPV infection status was excluded during patient matching. Groups were fully matched at a ratio of 1:1 with the nearest neighbor algorithm and a 0.1 caliper.

Statistical analysis
All statistical analyses were performed with SPSS 22.0 software (IBM, Armonk, NY, USA). All data are presented as the mean ± SD from at least three independent experiments. Comparisons between the immune content score and categorical variables (age, sex, smoking status, alcohol history, T stage, N stage, clinical stage, HPV status, primary site, and mutation burden) were performed with a chi-square test or Fisher’s exact test. The Kaplan-Meier method was used to construct survival curves, and log-rank tests were applied to compare the differences between groups for OS and DFS in the HNSCC cohort, whereas only DFS analysis was available in the NPC cohort. Multivariate analysis was carried out to determine independent prognostic factors with a Cox proportional hazards regression model. Two-tailed p values < 0.05 were considered significant.

Availability of data and materials
Data are available in the public, open access repositories from The Cancer Genome Atlas Data Portal (https://www.cancer.gov/tcga/) and the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/, GSE102349).

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.omto.2021.06.013.

ACKNOWLEDGMENTS
This study was supported by grants from the National Natural Science Foundation of China, Grants (81803049), Natural Science Foundation of Guangdong Province, China (2019A1515011651 and 2017A030312003), Fundamental Research Funds for the Central Universities, China (19ykpy175), and Planned Science and Technology Project of Guangdong Province, China (2019B020230002).

AUTHOR CONTRIBUTIONS
Yingqin Li, J.M., and N.L. designed the study; Yingqin Li, X.H., Y.Z., Yingqing Li, Y. Lei, Q.H., X.Y., and Y. Liang acquired and analyzed the data; Yingqin Li, N.L., and J.M. wrote the manuscript.

DECLARATION OF INTERESTS
The authors declare no competing interests.

REFERENCES
1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68, 394–424.
2. Budach, V., and Tinhofer, I. (2019). Novel prognostic clinical factors and biomarkers for outcome prediction in head and neck cancer: a systematic review. Lancet Oncol. 20, e313–e326.
3. Karam, S.D., and Raben, D. (2019). Radiomunotherapy for the treatment of head and neck cancer. Lancet Oncol. 20, e404–e416.
4. Cooper, J.S., Pajak, T.F., Forastiere, A.A., Jacobs, I., Campbell, B.H., Saxman, S.B., Kish, J.A., Kim, H.E., Cmelak, A.J., Rotman, M., et al.; Radiation Therapy Oncology Group 9501/Intergroup (2004). Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. N. Engl. J. Med. 350, 1937–1944.
5. Bernier, J., Domenge, C., Orsahin, M., Matuszewksa, K., Lefèbvre, J.L., Greiner, R.H., Giralt, J., Maingon, P., Rolland, F., Bolla, M., et al.; European Organization for Research and Treatment of Cancer Trial 22931 (2004). Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. N. Engl. J. Med. 350, 1945–1952.
