A Quantitative Histomorphometric Classifier (QuHbIC) Identifies Aggressive Versus Indolent p16-Positive Oropharyngeal Squamous Cell Carcinoma

James S. Lewis Jr, MD[*,†], **Sahirzeeshan Ali, MS**[‡], **Jingqin Luo, PhD**[§], **Wade L. Thorstad, MD**[||], and **Anant Madabhushi, PhD**[‡]

[*] Department of Pathology and Immunology, Washington University, St Louis, MO

[†]Department of Otolaryngology Head and Neck Surgery, Washington University, St Louis, MO

[||]Department of Radiation Oncology, Washington University, St Louis, MO

[§]Division of Biostatistics, Case Western Reserve University, Cleveland, OH.

[‡]Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH.

Abstract

Human papillomavirus–related (p16-positive) oropharyngeal squamous cell carcinoma patients develop recurrent disease, mostly distant metastasis, in approximately 10% of cases, and the remaining patients, despite cure, can have major morbidity from treatment. Identifying patients with aggressive versus indolent tumors is critical. Hematoxylin and eosin-stained slides of a microarray cohort of p16-positive oropharyngeal squamous cell carcinoma cases were digitally scanned. A novel cluster cell graph was constructed using the nuclei as vertices to characterize and measure spatial distribution and cell clustering. A series of topological features defined on each node of the subgraph were analyzed, and a random forest decision tree classifier was developed. The classifier (QuHbIC) was validated over 25 runs of 3-fold cross-validation using case subsets for independent training and testing. Nineteen (11.9%) of the 160 patients on the array developed recurrence. QuHbIC correctly predicted outcomes in 140 patients (87.5% accuracy). There were 23 positive patients, of whom 11 developed recurrence (47.8% positive predictive value), and 137 negative patients, of whom only 8 developed recurrence (94.2% negative predictive value). The best other predictive features were stage T4 (18 patients; 83.1% accuracy) and N3 nodal disease (10 patients; 88.6% accuracy). QuHbIC-positive patients had poorer overall, disease-free, and disease-specific survival ($P < 0.001$ for each). In multivariate analysis, QuHbIC-positive patients still showed significantly poorer disease-free and disease-specific survival, independent of all other variables. In summary, using just tiny hematoxylin and eosin punches, a computer-aided histomorphometric classifier (QuHbIC) can strongly predict recurrence risk. With prospective validation, this testing may be useful to stratify patients into different treatment groups.

Keywords

human papillomavirus; oropharyngeal squamous cell carcinoma; p16; quantitative histomorphometrics; image analysis; prognosis

The incidence of oropharyngeal squamous cell carcinoma (OSCC) due to human papillomavirus (HPV) is increasing.[1,2] The virus-related tumors are clinically, molecularly, and biologically distinct. They tend to arise in white men in their 40s and 50s who are nonsmokers (or former light smokers) and who have higher rates of sexual exposure and sexual partners. They classically present due to cervical nodal metastases, and the primary tumors are usually small[3] with a distinct nonkeratinizing morphology.[4] The survival is quite favorable and clearly better than for HPV-negative tumors.[5] However, although most patients respond well to therapy, approximately 10% to 15% of patients develop progressive disease, predominantly in the form of distant metastases.[3,6] T-stage[5,7,8] and smoking status[1,9] are well established as strong prognostic factors for HPV-related OSCCs, but the current focus is to identify more robust prognostic biomarkers that can more tightly define the risk of progressive disease for individual patients. Clinical trials are underway to evaluate deintensified treatment regimens with the goal of reducing the substantial morbidity from surgery, radiation, and chemotherapy among these patients who typically do not die of their cancers and who survive long term after treatment. New biomarkers that separate the good biology cancers from the more aggressive ones would greatly assist in targeting some patients for deintensification and perhaps would allow for intensified treatment for the patients who are at higher risk.[10]

We previously identified morphologic features in HPV-related OSCC which correlate with disease recurrence.[11] Some tumors have profound nuclear anaplasia, and this is often associated with multinucleation of tumor cells. In a well-characterized cohort of surgically resected OSCCs, the presence of these changes correlated strongly and independently with disease recurrence.[11] In particular, patients whose tumors lacked these changes almost never suffered disease recurrence.

With the advent of digital pathology, imaging scientists have begun to develop computerized image analysis algorithms for making diagnostic (disease presence), prognostic (outcome prediction), and theragnostic (choice of therapy) predictions from high-resolution digitized histopathology images.[12] Computer-aided quantitative histomorphometric analysis is an emerging field which uses powerful computing to identify, characterize, and quantitate histologic features of tissues in a way that is beyond the visual capabilities of a human being.[13] Many quantitative features can be assessed, such as precise numeric measurements pertaining to the spatial arrangement and architecture of nuclei, shapes of nests and nuclei, and nuclear texture. This technology has already been shown to be useful for the detection of prostate adenocarcinomas in tissue sections[14] and also for predicting tumor biology and clinical behavior in breast[15] carcinomas. It has significant potential to transform the practice of pathology.

This study aimed to evaluate quantitative histomorphometric analysis for prognostication in HPV-related (p16-positive) OSCC utilizing a tissue microarray cohort of patients with robust clinical follow-up.

MATERIALS AND METHODS

Patient Selection

With approval from the Human Research Protection Office, new primary cases of OSCC from 1997 to 2008 were identified from Radiation Oncology and Otolaryngology Head and Neck Surgery databases. Clinical follow-up information, including survival data, smoking, Adult Comorbidity Evaluation (ACE-27) indices, and other clinical variables, was obtained from a clinician database (W.L.T.) and electronic medical records as well as hospital oncology data services. Whole-tissue sections of all cases were reviewed by the study pathologist (J.S.L.) and confirmed as SCC.[4,11,16] Other pathologic features were obtained by pathology report review. Tumors were histologically typed as keratinizing SCC, nonkeratinizing SCC, and nonkeratinizing SCC with maturation according to our established system.[4] All patients were treated either with primary surgery with or without postoperative radiation and chemotherapy or were treated definitively with radiation therapy with or without chemotherapy. Intensity-modulated radiation therapy was administered in all cases by a single radiation oncologist (W.L.T.). Patients were not clinically managed with regard to HPV status.

The paraffin blocks were obtained from department files, and a tissue microarray was constructed (Fig. 1). Depending on the amount of available biopsied or resected tumor tissue, 2 mm (or, if only limited amounts of tumor tissue were present, 0.6 mm) punches were taken in duplicates. As most of the cases (85% to 90%) were treated with primary surgery, the majority of cases on the array had the larger (2 mm) punches. The hematoxylin and eosin (H&E)-stained slides were digitally scanned at ×400 magnification using an Aperio Scanscope XT digital scanner. These images were then annotated to denote the cases that were p16 positive (see next paragraph) and further to designate those who developed recurrent disease of any kind in the clinical follow-up period versus those who did not. Cases for which there was <10% tumor tissue across the 2 punches or for which the tumor on the tissue cores were folded or markedly distorted were excluded (ie, not annotated for inclusion). The resulting anonymous and annotated digital images were then shared with coinvestigators by file sharing protocol.

Immunohistochemistry for p16

Immunohistochemical analysis was previously performed for p16 on 4-μm-thick whole-tumor formalin-fixed, paraffin-embedded sections of each case that was placed on the tissue microarray using a monoclonal antibody to p16 (Roche/MTM Laboratories; clone E6H4, monoclonal, 1:1 dilution) on a Ventana Benchmark automated immunostainer (Ventana Medical Systems Inc., Tucson, AZ) according to standard protocols with appropriate controls including normal tonsil (negative) and a known positive OSCC case. Antigen retrieval, standard on the machine, utilized the Ventana CC1, EDTA-Tris, pH 8.0 solution. Staining was read by the study pathologist (J.S.L.). Staining was considered positive only if there was nuclear and cytoplasmic staining (of any intensity) in >50% of tumor cells (Fig. 1) and negative if there was no (or only partial) staining of <50% of cells, as it is well established that only extensive expression correlates with the presence of transcriptionally active HPV.[17,18]

Quantitative Histomorphometric Analysis

Graph theory has emerged as a method to characterize the structure of large complex networks leading to a better understanding of dynamic interactions that exist between their components.[19] Topographical characteristics extracted from these graphs can define the network structure (topology) and relationships that exist within the node population.[20,21] Nodes with similar characteristics tend to cluster together, and the pattern of this clustering

provides information as to the shared properties, and therefore the function, of those individual nodes. In the context of image analysis and classification of digital pathology, some researchers have shown that spatial graphs and tessellations such as those obtained through the Voronoi, Delaunay, and minimum spanning tree, built using nuclei as vertices, may actually have biological context and potentially be predictive of disease severity.[20,21] These graphs have already been mined for quantitative features that have shown to be useful in the context of prostate and breast cancer grading.[21,22]

Despite their complex nature, cancer cells tend to self-organize in clusters and exhibit architectural organization, an attribute that forms the basis of many cancers.[23] This paper uses a computational model, Cell Cluster Graph (CCG), which relies solely on the organization of clusters of tumor cells.

The intuition behind CCGs is to capture clustering information of nuclei in large tissue specimens and to extract topological properties and attributes that can quantify tumor morphology efficiently. CCG is based on connecting the nuclei into nodes and analyzing the distribution and pattern of the nodes. A detailed mathematical description of the image analysis methods is available in Supplemental File 1 (Supplemental Digital Content 1, http://links.lww.com/PAS/A184).

Briefly, construction of CCGs can be achieved in 3 steps. The first step is to distinguish nuclei from the background. The second step is to identify nuclei clusters for node assignment. The last step is the link establishing where the pairwise spatial relation between the nodes is translated to the edges (links) of CCG with a certain probability. Figure 2 illustrates the workflow of CCG on oropharyngeal tissue specimen images. We extracted a set of features (graph metrics) from CCG to quantify tumor morphology. These features are listed in Table 1.

To show the utility of the CCG features in distinguishing between favorable and poor prognoses of p16-positive oropharyngeal carcinomas, we performed an unsupervised dimensionality reduction on the set of CCG features extracted from all 160 cases by means of principal component analysis. This technique reduced the original set of 7 CCG features to 3 principal components for each of the images, enabling us to more easily visualize whether the CCG features could separate between the progressors and nonprogressors in reduced dimension space.

Statistical Analysis

Associations between the binary classifier results and the other categorical clinical and pathologic variables were determined by Fisher exact tests. Overall, disease-free, and disease-specific survival were defined as the time interval between the start of treatment (either the date of surgical resection or start of radiation and/or chemotherapy) and the date of death due to any cause, the date of death or the date of first tumor recurrence, or the date of death in patients with known recurrent tumor present at the time of death, respectively. The Kaplan-Meier (KM) product limit method was used to estimate empirical survival probabilities as illustrated by KM curves. Log rank tests were applied to examine survival differences, indicating the significance of a categorical variable being prognostic for a survival endpoint. Multivariate Cox proportional hazard models were also used to investigate the independent prognostic ability of variables after accounting for the other classic clinical and pathologic variables. Hazard ratios, associated 95% confidence intervals, and Wald test P-values were reported. All tests were 2 sided, with the significance level set at 0.05. All analyses were performed in statistical software R 2.15.2 (http://cran.r-project.org).

RESULTS

Patient Demographics

The microarray consisted of 17 H&E slides. There were 160 p16-positive cases which had adequate tumor present for analysis. Demographic and clinical data for the overall cohort of patients are summarized in Table 2. The patients were primarily white men in their mid 50s, two thirds of whom were current or former smokers. More than 90% of patients had nodal metastases at presentation and approximately 85% to 90% of patients were treated with primary surgery with or without postoperative radiation. Approximately three quarters of patients' tumors were T1 or T2, and >90% had a nonkeratinizing morphology. Most of these clinical and pathologic features typify HPV-related (p16-positive) OSCC. Tumor recurred in only 19 of the 160 patients (11.9%), and 13 of these 19 (68.4%) were distant metastases (or 8.1% of the total number of patients).

Image Analysis

CCGs were generated after computer identification of nuclei through the watershed algorithm (Fig. 2). Results of simple cell cluster graphing into reduced dimension without supervised modeling and without regard to disease recurrence status (ie, unsupervised cluster analysis on CCG features extracted from all the images) are shown in Figure 3. This shows that the tumors from patients with recurrent disease cluster closely together with little overlap with patients who did not suffer disease recurrence. The quantitative histology-based image classifier (or QuHbIC) was then developed on the basis of Support Vector Machine (SVM) training on features (Table 1) extracted from the CCGs (mathematical details in Supplemental File 1, (Supplemental Digital Content 1, http://links.lww.com/PAS/A184). The SVM was trained with sets of progressor (+) and nonprogressor (−) instances to learn a decision boundary for separating the two classes. Hence, any new instance that was found to land on the (+) partition of the decision boundary was classified as a progressor (unfavorable) and, similarly, that found to land on the (−) partition was identified as a nonprogressor (favorable).

To ensure robustness of the classifier for training and testing data, a randomized 3-fold cross-validation procedure was implemented. Two subsets were considered as training data and the remaining as testing data, following which classification was performed. This was repeated until all 3 subsets were classified and the entire cross-validation procedure was iterated 10 times.

QuHbIC was positive in 23 of the 160 patients (14.4%). As shown in Table 3, 11 of the 23 (47.8%) positive patients developed recurrent disease compared with only 8 of the 137 (5.8%) negative patients ($P = 0.0001$). This translates into an 8-fold higher rate of recurrent disease in the classifier-positive patients versus the classifier-negative ones. For prediction of recurrent disease, QuHbIC had an accuracy of 87.5% and positive and negative predictive values of 47.8% and 94.2%, respectively. This was better than any other single clinical or pathologic feature. The best predictive ability among traditional clinical and pathologic variables (Table 3) was for patients with T3 or T4 versus T1 or T2 tumors (accuracy of 74.0% and positive and negative predictive values of 12.5% and 91.4%, respectively; $P = 0.076$), T4 versus T1, T2, or T3 tumors (accuracy of 83.1% and positive and negative predictive values of 27.8% and 90.4%, respectively; $P = 0.04$), and N3 versus N0, N1, or N2 tumors (accuracy of 88.6% and positive and negative predictive values of 50.0% and 91.2%, respectively; $P = 0.0019$). QuHbIC-positive patients were statistically significantly more likely to be male and to have received definitive radiation therapy for their treatment. There were no significant correlations between QuHbIC results and the other clinical and pathologic features.

Survival Analysis

QuHbIC-positive patients had statistically significantly worse overall, disease-free, and disease-specific survival in univariate analysis (Table 4). KM survival curves for QuHbIC are presented in Figure 4. Among the other variables, high T-stage (as T3 or T4 vs. others), high N-stage (as N2C or N3 vs. others), smoking (lifetime vs. never), and treatment with primary radiation (+/− chemotherapy) rather than primary surgery also correlated with worse overall, disease-free, and disease-specific survival in univariate analysis. All of these are what would be expected in a contemporary cohort of p16-positive OSCC.

Multivariate survival analysis controlling for the effect of the major prognostic variables is presented in Table 5. QuHbIC results still correlated with worse overall, disease-free, and disease-specific survival independent of the other variables, although this was only statistically significant for disease-free and disease-specific survival. High T-stage also still retained statistically significant correlation with poorer overall, disease-free, and disease-specific survival.

DISCUSSION

As the rates and overall prevalence of HPV-related OSCC increase in what has been termed a cancer "epidemic,"[10,24,25] it is incumbent that we learn to treat patients with these cancers in ways that are appropriate to their unique biology. The typical surgical, radiation, and chemotherapeutic regimens that have been developed for traditional head and neck SCC certainly result in cure for most patients with HPV-related OSCC,[1] but morbidity can be substantial.[26,27] The head and neck medical community has begun in recent years to much more clearly consider morbidity for the survivors of OSCC. Indeed, several clinical trials utilizing deintensified or targeted therapies for such patients are now underway. The concern is for the 10% to 15% of patients with HPV-related OSCC in whom the tumors are not necessarily indolent or treatment responsive.[1,6,7] Although most patients do well with today's treatments, we have all observed those patients with aggressive, refractory disease who either have distant metastases at presentation or who go on to develop them.[28] A major focus is to stratify patients with HPV-related OSCC into low-risk and high-risk groups.[1,9] Other than high T-stage[1,5,7] and, on a less consistent basis, smoking status[9] and high N-stage,[1,9] few clinical or pathologic features correlate significantly with disease recurrence after treatment. No robust molecular markers have emerged either.[29]

We recently published on the histologic features of tumor cell anaplasia and multinucleation in surgically resected OSCC as strongly correlating with disease recurrence independent of all other variables.[11] This work showed that morphologic features of these cancers can potentially be useful to predict their behavior. This study entailed review of all slides from fully resected primary tumors as well as nodal metastases, and the definition for positivity was a single high-power field with 3 or more large (anaplastic) nuclei and/or 3 or more overtly multinucleated cells. Given that the changes of anaplasia and multinucleation were usually very focal and perhaps their determination by pathologists somewhat subjective, we speculated that this finding by pathologists, alone, might not be practical in the routine practice setting. Further, one could never "rule out" such changes in patients who would be treated with definitive chemoradiation, in whom you only receive a small biopsy specimen for diagnosis before their treatment. We speculate, however, that the emerging technique of computer-aided quantitative histomorphometry might be better able to detect and quantify these morphologic changes, particularly the more subtle ones or ones in which quantitation is needed.

Quantitative histomorphometry has already been successful in predicting recurrence in breast cancer[15,30] and biochemical recurrence in prostate cancer. In men with prostate

cancer, quantitative histomorphometrics was able to predict biochemical recurrence in patients with long-term follow-up after prostatectomy.[31] In breast cancer, quantitative analysis of tissue specimen revealed significant contributions of stromal regions of tumors to the risk for disease progression.[30]

The first finding in our study that signaled that the morphology does correlate with behavior was that by simple, unsupervised clustering of the CCG data, there was a clear segregation of those patients who developed recurrent disease versus those who did not (Fig. 3). We then modeled the histomorphometric features to develop and refine a simple, binary test, QuHbIC, which, when positive, strongly enriched for patients who developed recurrent disease. Only 23 of 160 (14.4%) patients had a positive QuHbIC result, but these patients were 8 times more likely to develop recurrent disease (48% vs. 6%). Of the other major individual prognostic variables, having a T4 tumor made a patient 3 times more likely to develop recurrent disease, and having N3 nodal disease made a patient 6 times more likely to develop recurrent disease. Thus, QuHbIC was the single most predictive feature for recurrent disease in this patient cohort. Focusing specifically on distant metastasis (the primary type of disease recurrence and cause of patient death), of the 23 patients with positive QuHbIC, 9 (39%) developed distant metastasis compared with only 4 of 133 (3%) with negative results, a rate over 10 times higher. Also importantly, a negative QuHbIC result signaled a 50% lower likelihood of disease recurrence. Only 6% of patients with negative QuHbIC developed recurrent disease versus 12% for the entire cohort. Only 3% of patients with negative QuHbIC developed distant metastases versus 8% of the entire cohort.

Several other prognostic biomarkers have been investigated among HPV-related OSCC, and some have shown promise. High expression of cyclin D1[32–34] and EGFR[35,36] has been shown to correlate with worse survival in some studies. However, the studies have usually included a mixture of HPV-positive and HPV-negative tumors, and the results have only been inconsistently shown to be independent of HPV status. Other markers that have been studied include p21,[37] bcl-2, mdm-2, survivin,[38] and p14[37] but with less promising results. Additional limitations to these immunohistochemical biomarkers are that the expression levels are continuous so that the cutoff values for positive results (ie, what constitutes high expression) have been inconsistently applied, the methods for assessment have varied. There is also likely to be significant interobserver variability in the assessment of staining.

The quantitative histomorphometric approach utilized here has many potential benefits, such as the capability to simply utilize H&E-stained sections, to assess tumors on only very small, and randomly selected, tumor samples (in this case the punches on the microarrays were randomly taken from the central tumor and were either 2 or 0.6 mm only), and to provide a simple, binary result that, in theory, should be free of substantial variability. One can thus imagine a future clinical management workflow in which a quantitative histomorphometric classifier is utilized in the routine pathologic evaluation along with other clinical and pathologic variables to place patients in low-risk and high-risk groups that have their own specific treatment types and regimens. The low-risk patients would be candidates for deintensified or targeted therapies, and the high-risk patients would get standard, full-intensity (or even higher) treatment or some other form of appropriate multimodality treatment.

The molecular mechanisms underlying the morphologic changes we observe microscopically as pathologists and which are being detected by QuHbIC image analysis are not known. From the larger field of oncology, it is clear that tumor cell pleomorphism, multinucleation, and nuclear anaplasia are signals of genetic complexity. So the high-risk tumors in these studies almost certainly are ones that have undergone some form of genetic progression. A major focus of our future research will be to delineate the specific genes and

pathways involved in this progression with the goal of facilitating better classification of patients' tumors and hopefully for the development of targeted treatments with less morbidity than current radiation and chemotherapy.

Clearly, QuHbIC was modeled on the very cohort of patients for which we have reported outcome data here. However, the fact that histomorphometric data can be extracted and correlated with clinical outcomes is extremely promising in and of itself, and with sophisticated cross-validation techniques, one can strongly assert that QuHbIC results are predictive independent of the data set on which they were modeled.

QuHbIC has the potential of being another piece of information for surgeons and oncologists in treating their patients with OSCC. One can imagine its use in routine practice as follows: patients are evaluated and undergo biopsy or surgical resection of their cancers, the tissue is diagnosed as SCC and tested for p16, and, if positive, one slide is digitally scanned. Computerized image analysis is performed, and the QuHbIC classifier result is either "favorable" or "unfavorable." Those patients with "favorable" results are at low risk for recurrent disease and are therefore candidates for lesser doses of radiation or chemotherapy or for targeted, lower morbidity therapies. Those patients with "unfavorable" results are at much higher risk for recurrent disease and are therefore given full, multimodality therapy or even potentially more aggressive therapy.

The next step is testing whether QuHbIC still retains the strong correlation with disease recurrence, independent of other features, in a separate, large, validation cohort of p16-positive OSCC patients. If this is performed and is successful, then a prospective, randomized trial stratifying treatment by QuHbIC results would address whether it would be useful in actual clinical management of patients.

In summary, using a large and well-characterized tissue microarray cohort of p16-positive OSCC patients, QuHbIC, an image-based computer-aided quantitative histomorphometric classifier, was developed and tested and shows a very strong and independent correlation with disease recurrence. This testing holds promise as a risk classifier with the potential to guide clinical management of patients with OSCC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

1. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. 2010; 363:24–35. [PubMed: 20530316]

2. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papilloma-virus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol. 2011; 29:4294–4301. [PubMed: 21969503]

3. Dahlstrom KR, Calzada G, Hanby JD, et al. An evolution in demographics, treatment, and outcomes of oropharyngeal cancer at a major cancer center: a staging system in need of repair. Cancer. 2012; 119:81–89. [PubMed: 22736261]

4. Chernock RD, El-Mofty SK, Thorstad WL, et al. HPV-related nonkeratinizing squamous cell carcinoma of the oropharynx: utility of microscopic features in predicting patient outcome. Head Neck Pathol. 2009; 3:186–194. [PubMed: 20596971]

5. Ukpo OC, Flanagan JJ, Ma XJ, et al. High risk human papillomavirus E6/E7 mRNA detection by a novel in situ hybridization assay strongly correlates with p16 expression and patient outcomes in oropharyngeal squamous cell carcinoma. Am J Surg Pathol. 2011; 35:1343–1350. [PubMed: 21836494]

6. Huang SH, Perez-Ordonez B, Weinreb I, et al. Natural course of distant metastases following radiotherapy or chemoradiotherapy in HPV-related oropharyngeal cancer. Oral Oncol. 2013; 49:79–85. [PubMed: 22917550]

7. Haughey BH, Hinni ML, Salassa JR, et al. Transoral laser micro-surgery as primary treatment for advanced-stage oropharyngeal cancer: a United States multicenter study. Head Neck. 2011; 33:1683–1694. [PubMed: 21284056]

8. Rich JT, Milov S, Lewis JS Jr, et al. Transoral laser microsurgery (TLM) +/− adjuvant therapy for advanced stage oropharyngeal cancer: outcomes and prognostic factors. Laryngoscope. 2009; 119:1709–1719. [PubMed: 19572271]

9. Gillison ML, Zhang Q, Jordan R, et al. Tobacco smoking and increased risk of death and progression for patients with p16-positive and p16-negative oropharyngeal cancer. J Clin Oncol. 2012; 30:2102–2111. [PubMed: 22565003]

10. Sturgis EM, Ang KK. The epidemic of HPV-associated oropharyngeal cancer is here: is it time to change our treatment paradigms? J Natl Compr Canc Netw. 2011; 9:665–673. [PubMed: 21636538]

11. Lewis JS Jr, Scantlebury JB, Luo J, et al. Tumor cell anaplasia and multinucleation are predictors of disease recurrence in oropharyngeal squamous cell carcinoma, including among just the human papillomavirus-related cancers. Am J Surg Pathol. 2012; 36:1036–1046. [PubMed: 22743286]

12. Madabhushi A, Doyle S, Lee G, et al. Integrated diagnostics: a conceptual framework with examples. Clin Chem Lab Med. 2010; 48:989–998. [PubMed: 20491597]

13. Gurcan MN, Boucheron LE, Can A, et al. Histopathological image analysis: a review. IEEE Rev Biomed Eng. 2009; 2:147–171. [PubMed: 20671804]

14. Monaco JP, Tomaszewski JE, Feldman MD, et al. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. Med Image Anal. 2010; 14:617–629. [PubMed: 20493759]

15. Basavanhally A, Feldman M, Shih N, et al. Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: comparison to Oncotype DX. J Pathol Inform. 2011; 2:S1. [PubMed: 22811953]

16. Lewis JS Jr. p16 immunohistochemistry as a standalone test for risk stratification in oropharyngeal squamous cell carcinoma. Head Neck Pathol. 2012; 6(suppl 1):75–82. [PubMed: 22430769]

17. Lewis JS Jr, Chernock RD, Ma XJ, et al. Partial p16 staining in oropharyngeal squamous cell carcinoma: extent and pattern correlate with human papillomavirus RNA status. Mod Pathol. 2012; 25:1212–1220. [PubMed: 22596101]

18. Chen ZW, Weinreb I, Kamel-Reid S, et al. Equivocal p16 immunostaining in squamous cell carcinoma of the head and neck: staining patterns are suggestive of HPV status. Head Neck Pathol. 2012; 6:422–429. [PubMed: 22801997]

19. Demir C, Gultekin SH, Yener B. Learning the topological properties of brain tumors. IEEE/ACM Trans Comput Biol Bioinform. 2005; 2:262–270. [PubMed: 17044189]

20. Madabhushi A. Digital pathology image analysis: Opportunities and challenges. Imaging Med. 2009; 1:7–10.

21. Doyle, S.; Hwang, M.; Kinsuk, S., et al. International Symposium on Biomedical Imaging: From Nano to Micro. Institute of Electrical and Electronics Engineers; Washington, DC: 2007. Automated grading of prostate cancer using architectural and textural imaging features..

22. Tabesh A, Teverovskiy M, Pang HY, et al. Multifeature prostate cancer diagnosis and Gleason grading of histological images. IEEE Trans Med Imaging. 2007; 26:1366–1378. [PubMed: 17948727]

23. Epstein JI, Allsbrook WC Jr, Amin MB, et al. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am J Surg Pathol. 2005; 29:1228–1242. [PubMed: 16096414]

24. Sturgis EM, Cinciripini PM. Trends in head and neck cancer incidence in relation to smoking prevalence: an emerging epidemic of human papillomavirus-associated cancers? Cancer. 2007; 110:1429–1435. [PubMed: 17724670]

25. Ramqvist T, Dalianis T. An epidemic of oropharyngeal squamous cell carcinoma (OSCC) due to human papillomavirus (HPV) infection and aspects of treatment and prevention. Anticancer Res. 2011; 31:1515–1519. [PubMed: 21617204]

26. Mortensen GL, Paaske PB. Patients perceive tonsil cancer as a strike at psycho-socially "vital organs". Dan Med J. 2012; 59:A4504. [PubMed: 22951199]

27. Rogers SN, Ahad SA, Murphy AP. A structured review and theme analysis of papers published on "quality of life" in head and neck cancer: 2000-2005. Oral Oncol. 2007; 43:843–868. [PubMed: 17600755]

28. Muller S, Khuri FR, Kono SA, et al. HPV positive squamous cell carcinoma of the oropharynx. Are we observing an unusual pattern of metastases? Head Neck Pathol. 2012; 6:336–344. [PubMed: 22528827]

29. Grimminger CM, Danenberg PV. Update of prognostic and predictive biomarkers in oropharyngeal squamous cell carcinoma: a review. Eur Arch Otorhinolaryngol. 2011; 268:5–16. [PubMed: 20827554]

30. Beck AH, Sangoi AR, Leung S, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci Transl Med. 2012; 3:108ra113.

31. Ali S, Veltri R, Epstein JA, et al. Cell Cluster Graph for Prediction of Biochemical Recurrence in Prostate Cancer Patients From Tissue Microarrays in Proceedings of the International Society for Optics and Photonics. 2013:8676.

32. Hong AM, Dobbins TA, Lee CS, et al. Use of cyclin D1 in conjunction with human papillomavirus status to predict outcome in oropharyngeal cancer. Int J Cancer. 2010; 128:1532–1545. [PubMed: 20503270]

33. Scantlebury JB, Luo J, Thorstad WL, et al. Cyclin D1-a prognostic marker in oropharyngeal squamous cell carcinoma that is tightly associated with high-risk human papillomavirus status. Hum Pathol. 2013; 44:1672–1680. [PubMed: 23566410]

34. Yu Z, Weinberger PM, Haffty BG, et al. Cyclin d1 is a valuable prognostic marker in oropharyngeal squamous cell carcinoma. Clin Cancer Res. 2005; 11:1160–1166. [PubMed: 15709184]

35. Kumar B, Cordell KG, Lee JS, et al. EGFR, p16, HPV Titer, Bcl-xL and p53, sex, and smoking as indicators of response to therapy and survival in oropharyngeal cancer. J Clin Oncol. 2008; 26:3128–3137. [PubMed: 18474878]

36. Hong A, Dobbins T, Lee CS, et al. Relationships between epidermal growth factor receptor expression and human papillomavirus status as markers of prognosis in oropharyngeal cancer. Eur J Cancer. 2010; 46:2088–2096. [PubMed: 20537890]

37. Hafkamp HC, Mooren JJ, Claessen SM, et al. P21 Cip1/ WAF1 expression is strongly associated with HPV-positive tonsillar carcinoma and a favorable prognosis. Mod Pathol. 2009; 22:686–698. [PubMed: 19305381]

38. Preuss SF, Weinell A, Molitor M, et al. Nuclear survivin expression is associated with HPV-independent carcinogenesis and is an indicator of poor prognosis in oropharyngeal cancer. Br J Cancer. 2008; 98:627–632. [PubMed: 18212752]
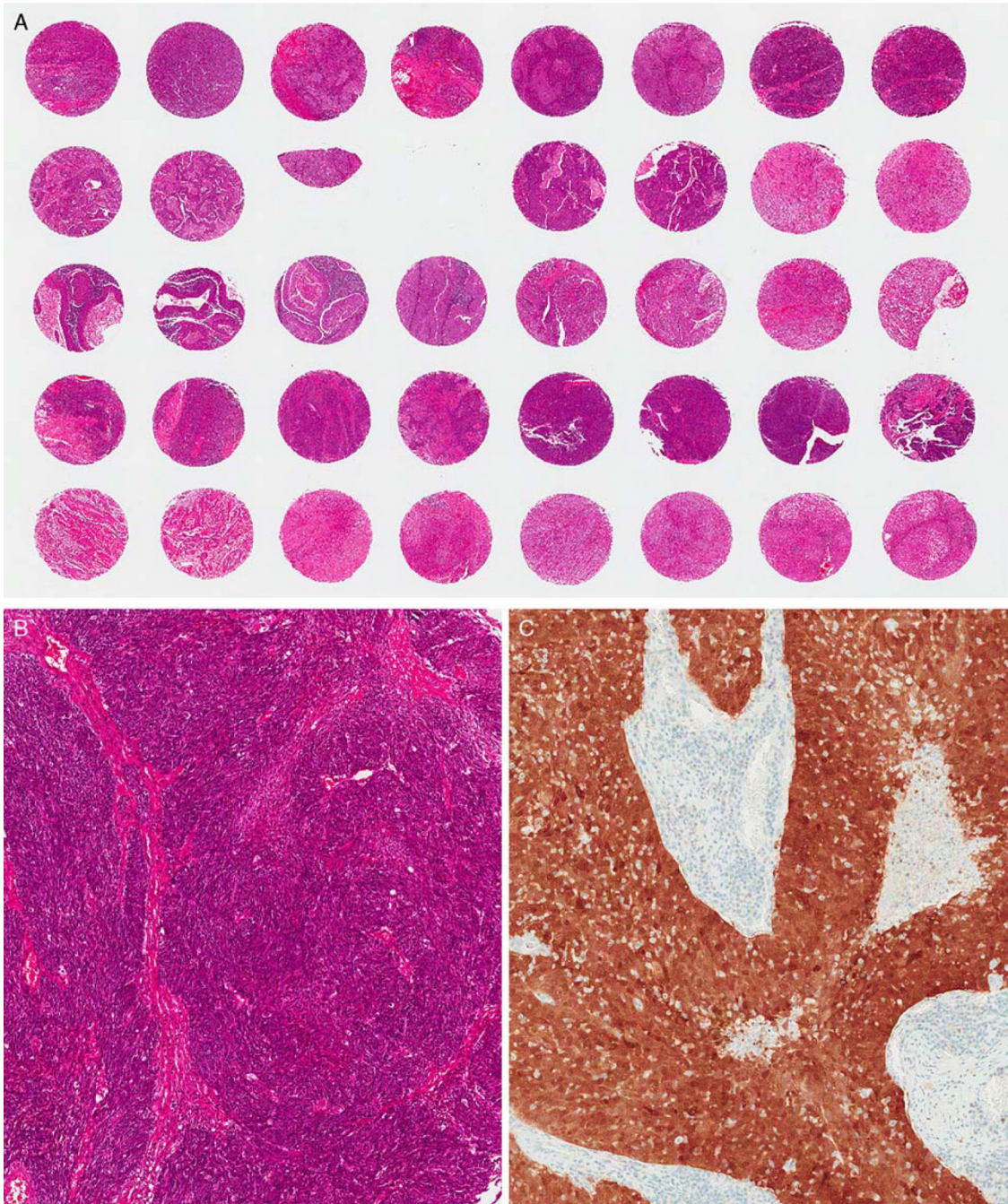
FIGURE 1.
A, Tissue microarray slide showing an array of 2 mm punches with H&E staining. B, A case on the array with tumor showing typical nonkeratinizing SCC morphology. C, p16 immunohistochemistry showing a tumor with strong and diffuse nuclear and cytoplasmic expression.
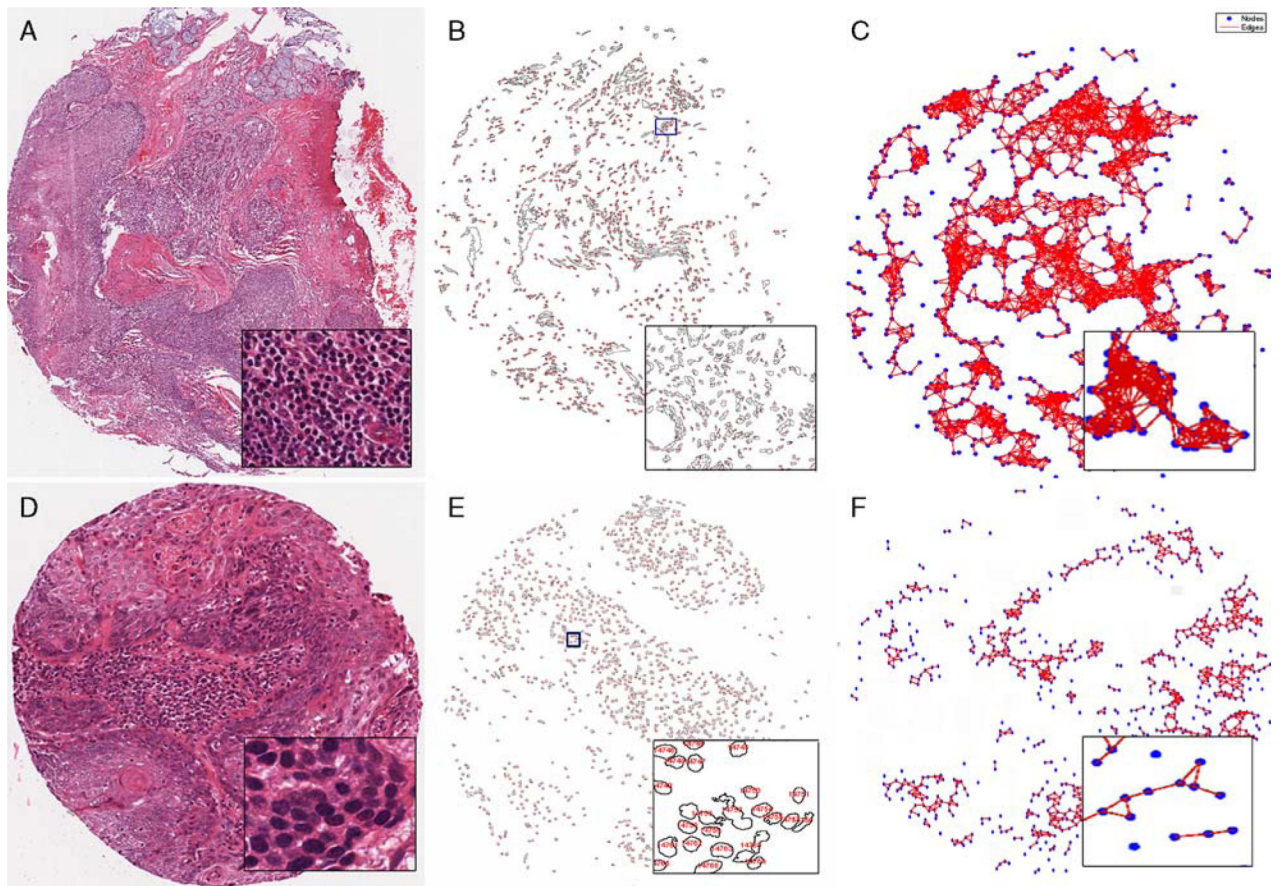
FIGURE 2.
Image analysis of the tissue microarrays. H&E images of tumors from representative patients who either did not develop recurrent disease (A) or who went on to develop recurrent disease (D). Nuclei were identified by computerized image analysis (B and E) and CCGs (C and F), shown by blue nodes for the nuclei with interconnecting red lines, generated. Insets of each panel show magnified areas of analysis.
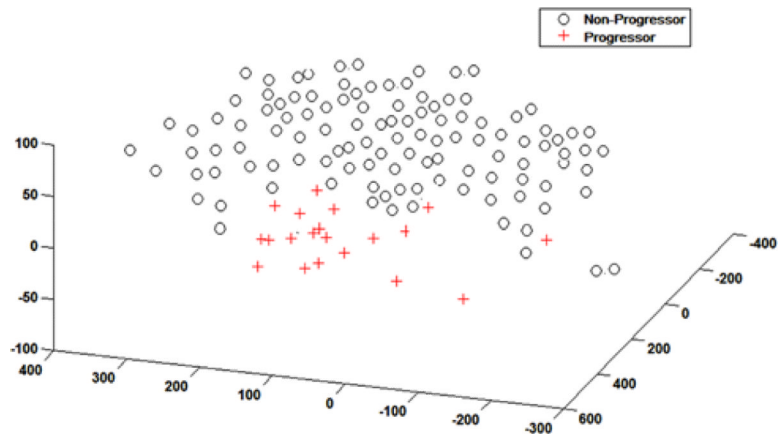
FIGURE 3.
Unsupervised clustering of histomorphometric data in reduced dimensionality space showing patients with recurrent disease (+) versus those with no recurrence (−). There is a clear separation between the two groups even without any modeling of the data.
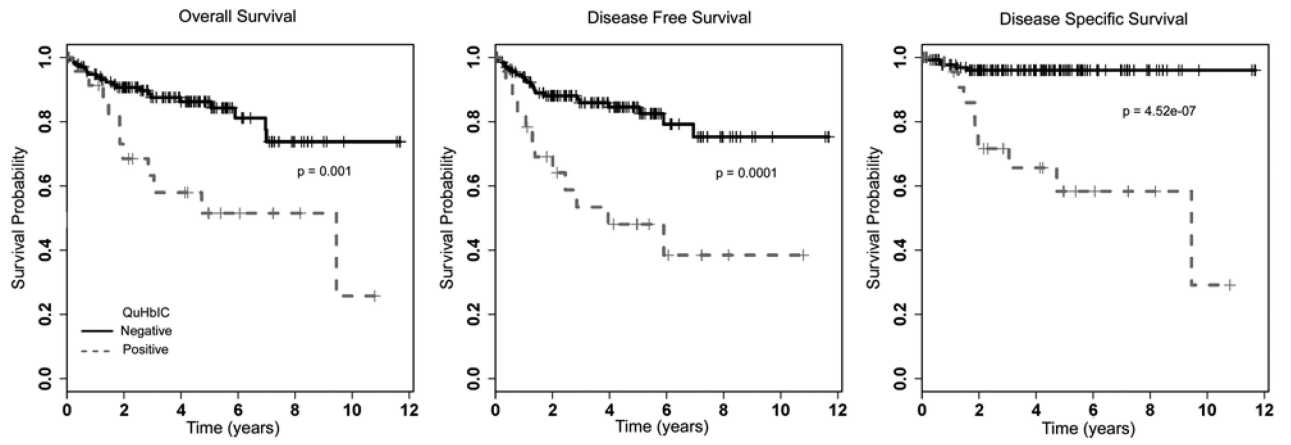
FIGURE 4.
K survival curves for QuHbIC results for the entire patient cohort.

TABLE 1

Features Characterized by the CCG Image Analysis

Features	Description
Clustering coefficient D	Ratio of total number of edges among the neighbors of the node and the node itself to the total number of edges that can exist among the neighbors of the node and the node itself per node
Giant connected component	Ratio between the number of nodes in the largest connected component in the graph and total the number of nodes
Average eccentricity	Average of node eccentricities, wherein the eccentricity of a node is the maximum shortest path length from the node to any other node in the graph
Clustering coefficient C	Ratio between the number of edges between the neighbors of node and the total possible number of edges between the neighbors of node

TABLE 2

Clinical and Pathologic Features for the Entire Cohort and by Respective QuHbIC Result

Characteristic	Entire Cohort = 160, N (%)	QuHbIC Positive = 23, N (%)	QuHbIC Negative = 137, N (%)	P^*
Age (y)	55.8	57.3 (50.9-64.4)	55.6 (49.0-61.4)	$0.50^\dagger$
Sex				
Female	16 (10.0)	0 (0.0)	16 (11.7)	0.13
Male	144 (90.0)	23 (100.0)	121 (88.3)	
Race				
White	151 (94.4)	21 (91.3)	130 (94.9)	0.62
Other	9 (5.6)	2 (8.7)	7 (5.1)	
Smoking				
Never	55 (32.3)	5 (21.7)	45 (34.1)	0.33
Ever (current or former)	105 (67.7)	18 (78.3)	87 (65.9)	
ACE Comorbidity Score				
0/1	122 (84.1)	17 (78.3)	105 (85.4)	0.35
2/3	23 (15.9)	5 (21.7)	18 (14.6)	
Tumor histology				
Keratinizing	7 (4.5)	1 (4.4)	6 (4.5)	0.78
Nonkeratinizing with maturation	45 (28.7)	5 (21.7)	40 (29.9)	
Nonkeratinizing	105 (66.9)	17 (73.9)	88 (64.7)	
T-stage				
T1/T2	116 (75.3)	16 (69.6)	100 (76.3)	0.60
T3/T4	38 (24.7)	7 (30.4)	31 (23.7)	
N-stage				
N0	12 (7.6)	1 (4.3)	11 (8.2)	0.88
N1/2a	49 (31.0)	8 (34.8)	41 (30.4)	
N2b/2c/3	97 (61.4)	14 (60.9)	83 (61.4)	
Overall stage				
I/II	10 (7.0)	1 (4.3)	9 (6.7)	1.0
III/IV	148 (93.0)	22 (95.7)	126 (93.3)	
Treatment				
Definitive radiation	18 (11.5)	9 (39.1)	9 (6.8)	**0.0001**
Surgery ± radiation	139 (88.5)	14 (60.9)	125 (93.2)	
Chemotherapy				
No	66 (49.6)	9 (40.9)	57 (51.4)	0.49
Yes	67 (50.3)	13 (59.1)	54 (48.6)	
Resection margin status				
Negative	115 (83.9)	13 (92.9)	102 (82.9)	0.47
Positive	22 (16.1)	1 (7.1)	21 (17.1)	
Disease recurrence				
No	19 (11.9)	12 (52.2)	129 (94.2)	**0.0001**
Yes	141 (88.1)	11 (47.8)	8 (5.8)	

Characteristic	Entire Cohort = 160, N (%)	QuHbIC Positive = 23, N (%)	QuHbIC Negative = 137, N (%)	P^*
Distant metastasis				
Negative	147 (91.9)	14 (60.9)	133 (97.1)	**0.0001**
Positive	13 (8.1)	9 (39.1)	4 (2.9)	
Death due to disease				
No	144 (91.7)	14 (60.9)	129 (95.5)	**0.0001**
Yes	13 (8.3)	9 (39.1)	6 (4.5)	

Values in bold are statistically significant, $P < 0.05$

[*] P-values were calculated from 2-sided Fisher exact test.

[†] P-value was calculated from Welch 2-sample t test.

TABLE 3

Correlation Between QuHbIC and Other Major Variables and Disease Recurrence

	QuHbIC Positive	QuHbIC Negative	Total
Recurrence, n (%)	11 (47.8)	8 (5.8)	19
No recurrence, n (%)	12 (52.2)	129 (94.2)	141
Total	23	137	160
Accuracy = 87.5%	PPV = 47.8%	NPV = 94.2%	***P* < 0.0001**

	T4	T1, T2, or T3	Total
Recurrence, n (%)	5 (27.8)	13 (9.6)	18
No recurrence, n (%)	13 (72.2)	123 (90.4)	136
Total	18	136	154
Accuracy = 83.1%	PPV = 27.8%	NPV = 90.4%	***P* = 0.040**

	N3	N0, N1, or N2	Total
Recurrence, n (%)	5 (50.0)	13 (8.8)	18
No recurrence, n (%)	5 (50.0)	135 (91.2)	140
Total	10	148	158
Accuracy = 88.6%	PPV = 50.0%	NPV = 91.2%	***P* = 0.0019**

P-values in bold are statistically significant, i.e., < 0.05.

TABLE 4

Univariate Log Rank Survival Analysis for the Major Clinical and Pathologic Variables and for the Image Analysis Classifier QuHbIC

Variable	Overall Survival, HR (95% CI)	P [†]	Disease-free Survival	P	Disease-specific Survival	P
Age	1.02 (0.98-1.06)	0.47	1.02 (0.99-1.06)	0.29	1 (0.95-1.06)	0.91
Sex						
Male vs. female	3.82 (0.52-28.02)	0.16	3.85 (0.53-28.22)	0.15	NA[*] (0-Inf)	0.18
Race						
Nonwhite vs. white	0.56 (0.08-4.14)	0.57	0.48 (0.07-3.55)	0.47	1.28 (0.17-9.88)	0.81
Smoking						
Ever vs. never	4.9 (1.48-16.18)	**0.0039**	3.98 (1.39-11.37)	**0.0053**	6.72 (0.88-51.43)	**0.033**
Treatment						
Surgery and/or postoperative radiation vs. definitive	0.21 (0.1-0.44)	**0.00001**	0.2 (0.1-0.42)	**< 0.00001**	0.09 (0.03-0.27)	**< 0.00001**
Overall stage						
III-IV vs. I-II	2.22 (0.3-16.32)	0.42	1 (0.24-4.17)	0.99	NA[*] (0-Inf)	0.33
Chemotherapy						
Yes vs. no	2.49 (1.09-5.69)	**0.026**	2.08 (0.95-4.57)	0.063	2.62 (0.74-9.27)	0.12
T-stage (%)						
T3-T4 vs. T1-T2	3.41 (1.64-7.08)	**0.00046**	2.74 (1.36-5.51)	**0.0033**	7.52 (2.26-25.02)	**0.00011**
N-stage (%)						
N1-3 vs. N0	3.06 (0.41-22.61)	0.25	1.41 (0.34-5.93)	0.63	NA[*] (0-Inf)	0.27
N-stage						
N2C or N3 vs. N0-N2B	2.62 (1.15-5.94)	**0.017**	2.27 (1.01-5.07)	**0.040**	5.55 (1.86-16.58)	**0.00055**
N-stage						
N2B-N3 vs. N0-N2A	2.28 (1-5.17)	**0.043**	1.59 (0.75-3.34)	0.22	2.07 (0.62-6.88)	0.23
Resection margin status						
Positive vs. negative	0.73 (0.17-3.16)	0.67	0.69 (0.16-2.97)	0.61	1.16 (0.14-9.98)	0.89
ACE Comorbidity Index						
2 or 3 vs. 0 or 1	1.45 (0.59-3.57)	0.41	1.37 (0.57-3.34)	0.48	1.65 (0.45-6.01)	0.44
QuHbIC						
Positive vs. negative	3.2 (1.53-6.7)	**0.0011**	3.64 (1.79-7.41)	**0.00013**	10.11 (3.36-30.38)	**< 0.00001**

CI indicates confidence interval; HR, hazard ratio.

Values in bold are statistically significant, $P < 0.05$

[*] Insufficient numbers in subgroups to generate a hazard ratio.

[†] P-values calculated from log rank tests.

TABLE 5

Multivariate Cox Model Analysis Controlling for Major Prognostic Variables

Variable	Overall Survival *P* HR (95% CI)	Disease-free Survival *P* HR (95% CI)	Disease-specific Survival *P* HR (95% CI)
Smoking	***P* = 0.038**	*P* = 0.063	*P* = 0.36
Ever vs. never	3.7 (1.07-12.82)	2.8 (0.94-8.32)	2.7 (0.32-23.03)
T-stage	***P* = 0.021**	*P* = 0.064	***P* = 0.029**
T3/T4 vs. T1/T2	2.74 (1.16-6.45)	2.2 (0.95-4.98)	4.4 (1.15-16.38)
Treatment	*P* = 0.11	*P* = 0.05	*P* = 0.05
Surgery/postoperative radiation vs. definitive	0.5 (0.2-1.17)	0.43 (0.19-0.99)	0.29 (0.09-1.01)
N-stage	*P* = 0.65	*P* = 0.62	*P* = 0.23
N2C/N3 vs. N0-N2B	1.24 (0.48-3.19)	1.3 (0.5-3.19)	2.2 (0.6-8.21)
QuHbIC	*P* = 0.081	***P* = 0.013**	***P* = 0.003**
Positive vs. negative	2.08 (0.91-4.75)	2.65 (1.22-5.75)	6.47 (1.87-22.35)

CI indicates confidence interval; HR, hazard ratio.

P-values in bold are statistically significant, i.e., < 0.05.