Research Article

Identification of Malignancy-Associated Changes in Histologically Normal Tumor-Adjacent Epithelium of Patients with HPV-Positive Oropharyngeal Cancer

James Jabalee [ID],[1] **Anita Carraro** [ID],[1] **Tony Ng** [ID],[2] **Eitan Prisman**,[3,4] **Cathie Garnis**,[1] **and Martial Guillaud** [ID][1]

[1]*British Columbia Cancer Research Centre, Department of Integrative Oncology, Vancouver, BC, Canada*
[2]*Department of Pathology, Vancouver General Hospital and University of British Columbia, Vancouver, BC, Canada*
[3]*Vancouver General Hospital, Division of Otolaryngology-Head and Neck Surgery, Vancouver, BC, Canada*
[4]*University of British Columbia, Vancouver, BC, Canada*

Correspondence should be addressed to Martial Guillaud; mguillau@bccrc.ca

The incidence of HPV-positive oropharyngeal cancer (HPV+ OPC) is increasing, thus presenting new challenges for disease detection and management. Noninvasive methods involving brush biopsies of diseased tissues were recently reported as insufficient for tumor detection in HPV+ OPC patients, likely due to differences between the site of tumor initiation at the base of involuted crypts and the site of brush biopsy at the crypt surface. We hypothesized that histologically normal surface epithelial cells in the oropharynx contain changes in nuclear morphology that arise due to tumor proximity. We analyzed the nuclear phenotype of matched tumor, tumor-adjacent normal, and contralateral normal tissues from biopsies of nine HPV+ OPC patients. Measurements of 89 nuclear features were used to train a random forest-based classifier to discriminate between normal and tumor nuclei. We then extracted voting scores from the trained classifier, which classify nuclei on a continuous scale from zero ("normal-like") to one ("tumor-like"). In each case, the average score of the adjacent normal nuclei was intermediate between the tumor and contralateral normal nuclei. These results provide evidence for the existence of phenotypic changes in histologically normal, tumor-adjacent surface epithelial cells, which could be used as brush biopsy-based biomarkers for HPV+ OPC detection.

1. Introduction

Oropharyngeal cancers (OPCs), which include malignancies of the tonsils, posterior pharyngeal wall, soft palate, and tongue base, have undergone a dramatic epidemiological shift. While the rate of tobacco and alcohol-related head and neck cancers is declining in North America, the incidence of HPV-positive OPC (HPV+ OPC) has been steadily rising since the early 1980s and is especially prevalent among young individuals (<60 years of age) [1]. These observations highlight the need for improved means of detecting and managing the disease.

The current standard of care, chemoradiation therapy (CRT), is associated with severe negative side effects, both acute and chronic [2–4]. Given the relatively young age and high survival rate of this cohort, patients often must live with significant morbidities for decades [5]. To avoid the drawbacks of CRT, transoral robotic surgery (TORS), in which the tumor is resected through the mouth, is currently being investigated as an alternative treatment option [6]. However, to facilitate TORS and circumvent the need for concurrent CRT and its associated toxicities, tumors must be diagnosed at an early stage (T3 or less) [6]. Thus, a method of screening for the disease in high-risk populations would increase the number of TORS-eligible patients, resulting in reduced morbidity and improved patient quality of life.

Previous work attempted to create a noninvasive screening method for HPV+ OPC by sampling the oropharyngeal

epithelium via brush biopsy and testing for the presence of HPV16 DNA via PCR [7]. Unfortunately, viral DNA was present only in a subset of HPV+OPC patients in which the tumor tissue was directly accessible, and not in those where disease did not present at the oropharyngeal surface [7]. The most likely explanation is that oropharyngeal tumors often initiate at the base of deep invaginations, or crypts, whereas brushing samples only the normal surface epithelial cells, which do not exhibit HPV infection [8]. Identifying biomarkers present in the superficial epithelium adjacent to a tumor is an essential step in the development of noninvasive, brush biopsy-based screening methods for detecting HPV+OPC.

Malignancy-associated changes (MACs) are subtle morphological changes that occur in histologically normal cells due to their proximity to a tumor and may include changes in the size, shape, and chromatin structure of the nucleus [9]. MACs have been shown to be reproducibly measured via image cytometry for numerous cancer types, including lung [10], breast [11], colon [12], cervical [13, 14], and oral [15–17] cancer, thus making them potentially useful diagnostic biomarkers. Here we use an in-house imaging system to measure >100 nuclear features for each cell nucleus of epithelial tissue samples derived from HPV+OPC patients to determine their utility as diagnostic biomarkers.

2. Materials and Methods

2.1. Sample Source. Nine tumors from OPC patients were collected via head and neck surgeries performed at Vancouver General Hospital as part of a randomized control trial comparing TORS to radiation to treat early-stage OPC. The use of this tissue was approved by the University of British Columbia Ethics Committee (ID#: H15-01121). Informed consent was obtained from each patient prior to surgery. Eligible patients had tumors < 4 cm in diameter (stage T1/T2) and lymph node metastases of <3 cm on either side of the neck (stage N0/N1/N2). A biopsy of the tumor, including adjacent histologically normal surface epithelium, and a biopsy of contralateral normal tissue were collected from each patient. Patient demographics are included in Table S1.

2.2. Sample Preparation. Nondiagnostic tissues were fixed in molecular fixative by the study pathologist, Dr. T. Ng, and paraffin embedded. Four-micrometer-thick sections were cut, and a subset was stained with hematoxylin and eosin (H&E). These slides were used by the pathologist to identify three areas of interest: (1) tumor, (2) adjacent normal epithelium (an), and (3) contralateral normal epithelium (cln) (Figure 1). Sections immediately neighboring the H&E-stained slides were stained with Feulgen-thionin [18], which is stoichiometric for DNA, and areas of interest were imaged as described below.

2.3. Image Cytometry. A semiautomated quantitative imaging system was used [19, 20]. Regions of interest identified in H&E-stained slides were manually delineated in adjacent Feulgen-thionin-stained sections by an experienced technician. The imaging system then automatically located and

focused individual objects, imaged them, and classified them as intact, in-focus, or "junk" nuclei using a random forest-based algorithm as previously described [17]. Nuclei spanning the full width of the epithelium were included in the analysis. The result of this process was then confirmed by an experienced technician. An illumination wavelength of 600 ± 5 nm was used, corresponding to the absorption peak of the Feulgen-thionin stain. An effective pixel sampling space in the plane of the sample of $0.34\,\mu m^2$ and an effective pixel sampling area of $0.116\,\mu m^2$ were used. Imaging system characteristics were in conformity with the European Society of Analytical Cellular Pathology [21]. For each nucleus, the software calculated >100 features related to size, shape, DNA quantity, and chromatin distribution as previously described [22]. Features found to be highly sensitive to the variability in staining intensity as assessed by our own experience were removed prior to the final analysis.

2.4. Nuclear Classification. A random forest learning algorithm was used to classify in-focus and intact nuclei as belonging to either contralateral normal or tumor tissue. This tree-based classification algorithm has the advantage of being robust against overfitting [23] and has been previously shown to perform well on a similar data set [17]. Nuclei from tumor and contralateral normal, but not adjacent normal, epithelial cells were used to train the classifier. Nuclei were randomly split into training (80%) and test (20%) sets. The number of features used by the classifier was tuned using 10-fold cross-validation (Table 1). Five hundred random trees were generated, and subsampling was done without replacement. Once the relevant parameters had been tuned, model performance was assessed on the test set using a receiver operating characteristic (ROC) curve [24].

Although all nuclei derived from normal tissue can be considered normal, not all nuclei derived from tumor tissue can be considered abnormal. Indeed, tumors are composed of a collection of cell types, including stromal cells, lymphocytes, fibroblasts, and others in addition to cancer cells, and the proportions of these cell types vary by tumor. Thus, any nucleus from tumor tissue that is classified as abnormal is considered more likely to have been derived from abnormal tissue than normal tissue. We selected only those epithelial nuclei derived from abnormal regions of tumor tissue for further analysis.

2.5. Large-Scale DNA Organization (LDO) Score. We extracted the voting score generated by the random forest classifier, which we term the large-scale DNA organization (LDO) score. The LDO score refers to the proportion of trees that classify a nucleus as "normal-like" and ranges from 0 (all trees vote "normal-like") to 1 (no trees vote "normal-like"). Intermediate scores represent a phenotype intermediate between "normal-like" and "tumor-like."

2.6. Statistical Analysis. Statistical analyses were performed using R statistical software (version 3.2.5). The random forest algorithm was implemented, parameters were tuned, and performance was evaluated using the CARET package [25], and ROC curves were generated using the pROC package
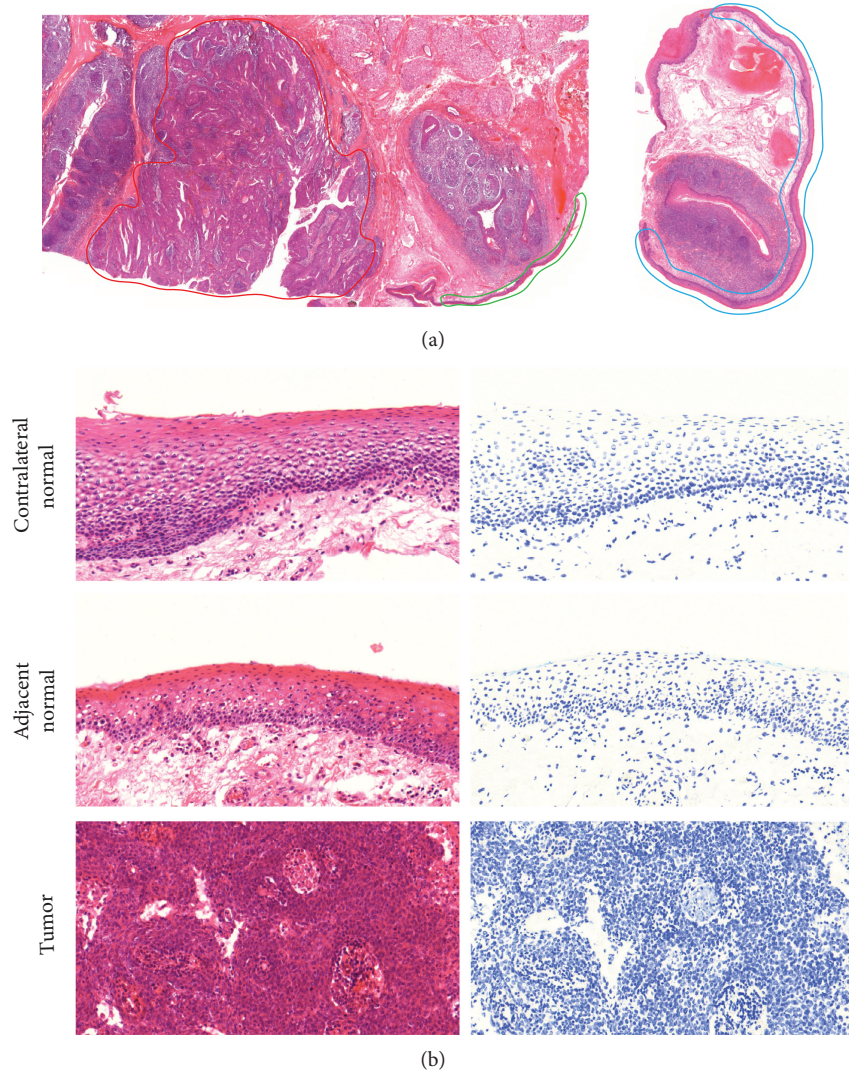
(a)

(b)

FIGURE 1: Regions of interest. (a) Tumor (left) and contralateral normal (right) biopsies were collected at the time of surgery. Three regions of interest corresponding to the tumor (red), adjacent normal epithelium (green), and contralateral normal epithelium (blue) were outlined by the study pathologist and used for analysis. (b) Close-up of each region of interest stained with H&E (left) or Feulgen-thionin (right).

TABLE 1: Parameters of the nuclear classification model.

Initial group	Classification groups	# of trees	# of features per tree	Sample size	Sampling
Intact, in-focus nuclei identified by technician	Contralateral normal, tumor	500	35	Contralateral normal: 5572 Tumor: 6653	Without replacement

[26]. For box plots, the center line of the box represents the median, and the box limits represent the 25th and 75th percentiles. The upper and lower whiskers extend to the 5th and 95th percentiles.

3. Results

A total of 4830 tumor and 3352 contralateral normal nuclei were used to train the random forest classifier. When applied to the test set, the model showed an area under the ROC curve of 0.90. A total of 84.7% of the 1207 tumor and 80.8% of the 837 contralateral normal test set nuclei were correctly classified (Table 2).

Examples of contralateral normal and tumor nuclei and values of select features are given in Figures 2(a) and 2(b). Of the top ten most important features of our model (based on the mean decrease in accuracy), eight represent chromatin texture features, while the remainder represent morphological and photometric features (the meaning of each feature is given in Table S2; for additional details, see

TABLE 2: Performance of the nuclear classification model.

| | | Predicted | | Correct classification rate (%) |
		Contralateral normal	Tumor	
A: Training set				
Actual	Contralateral normal	**2519**	883	75.5
	Tumor	661	**4169**	86.3
B: Test set				
Actual	Contralateral normal	**647**	154	80.8
	Tumor	190	**1053**	84.7

Correctly classified nuclei represented in bold.

Doudkine et al. [22]). All ten features are statistically significantly different between contralateral normal and tumor nuclei (two-tailed t-test of log-10-transformed values with Benjamini-Hochberg correction, $p < 0.001$). If cells of the adjacent normal epithelium do not contain MACs, we would expect that there would be no significant difference between contralateral normal and adjacent normal nuclei for any feature. However, of the top 10 most important model features, we find that four (fractal1_area, long_runs2, run135_percent, and long_runs1) show an intermediate phenotype between that of the contralateral normal and tumor (i.e., a statistically significant difference between adjacent normal and contralateral normal, and between adjacent normal and tumor), while three (max_radius, long135_runs, and fractal_dimen) are significantly different from the contralateral normal but not the tumor (ANOVA of log-10-transformed values with Tukey's post hoc honestly significant difference test, $p < 0.001$; Figure 2(c)).

Large-scale DNA organization (LDO) scores are a measure of how "normal-like" or "tumor-like" a specific nucleus is within the context of our model. The distribution of LDO scores of tumor and contralateral normal samples of the test set is shown in Figure 3. Contralateral normal samples are skewed to the right while tumor samples are skewed to the left, indicating that nuclei from tumor-derived epithelial cells are more likely to be abnormal. To further demonstrate that adjacent normal epithelial cells contain MACs, we used our model to calculate the LDO scores for the full data set consisting of 4189 contralateral normal, 5192 adjacent normal, and 6037 tumor nuclei from nine patients. If adjacent normal nuclei contain MACs, then the distribution of LDO scores should be intermediate between that of the contralateral normal and tumor. For each patient, the mean of the distribution of LDO scores from contralateral normal epithelium is close to zero (0.16 ± 0.15) and that of tumor is close to one (0.88 ± 0.13), as expected, whereas adjacent normal nuclei display an intermediate phenotype (0.49 ± 0.25; Figure 4). Furthermore, the distribution of LDO scores of adjacent normal nuclei is statistically significantly different from both the contralateral normal and tumor for each of the nine patients (ANOVA with Tukey's post hoc honestly significant difference test, $p < 0.001$). These results strongly suggest the presence of MACs in histologically normal tumor-adjacent epithelial cell nuclei.

4. Discussion

The rising incidence of HPV+ OPC presents new challenges for the detection and management of oropharyngeal malignancies. Given the relatively young age and good health of this cohort, much attention is being given to methods that deescalate the intensity of therapy without sacrificing its effectiveness, thus reducing morbidity and improving long-term patient quality of life [27]. TORS is a promising means for achieving this goal as it provides an alternative to standard CRT for individuals with early-stage tumors. TORS has similar disease outcomes when compared to CRT [28–30] but with improved functional outcomes [28–35]. Importantly, TORS has been found to result in significantly lower rates of dysphagia requiring either a gastrostomy tube or tracheotomy [28, 34], which are among the most common and challenging complications associated with CRT [2, 3, 34]. Strikingly, Moore et al. report preservation of swallowing function in >90% of TORS patients [28], whereas the rate of dysphagia requiring a gastrostomy tube following CRT approaches 45% [3, 36]. Patients treated with TORS also tend to score higher on factors related to quality of life, including eating, speech, and social, than did those treated with adjuvant radiation therapy or chemoradiation therapy [31, 32, 35]. Unfortunately, the use of TORS is limited to patients with early-stage tumors exhibiting limited nodal involvement [6]. A method of screening high-risk individuals would increase the number of TORS-eligible patients, thus avoiding the need for CRT and improving patient quality of life.

We sought to determine if MACs exist in cells of the histologically normal oropharyngeal surface epithelium that exists adjacent to tumor tissues. Previous reports describe MACs as being robustly identified in patients with early-stage tumors at other organ sites, making them ideal candidates for early tumor detection [11, 15, 37–41]. Furthermore, surface epithelial cells are easily accessible via brush biopsy, which is an attractive sampling approach for the oropharynx as it is cheap, rapid, safe, painless, and noninvasive. Although previous studies have reported that brush biopsies are problematic for general OPC screening because they cannot directly sample tumors arising in tonsillar crypts [7], brush biopsies of clearly visible and accessible OPC lesions have been reported to be effective for HPV detection [7, 42]. Further, brush biopsies in the adjacent oral cavity, where
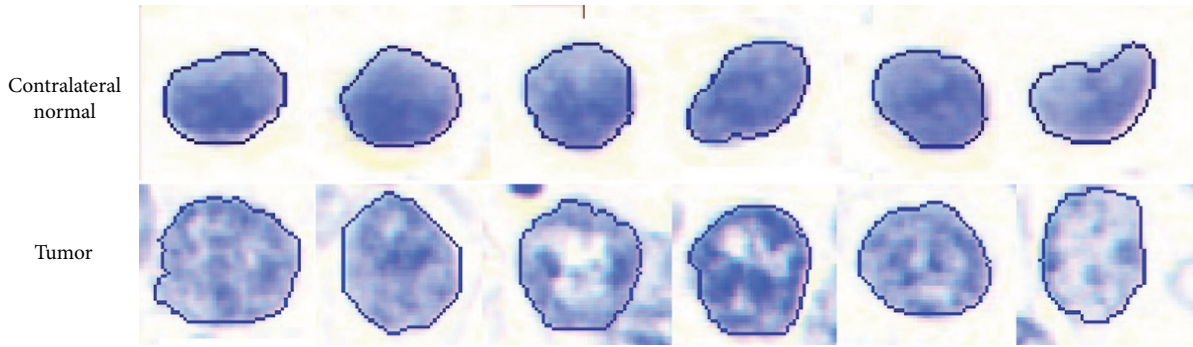
(a)

	Fractal1_area	Long_runs1	Long_runs2	Max_radius
Contralateral normal	6352.1	12.0	2.6	18.8
Tumor	13,399.5	8.3	1.5	21.9

(b)
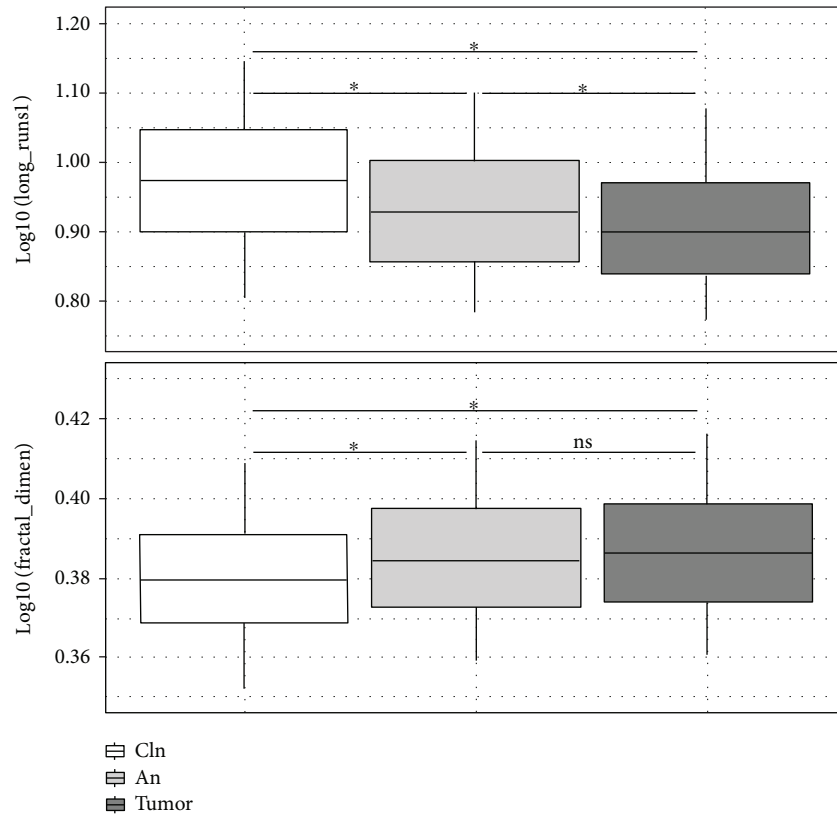
(c)

FIGURE 2: Examples of nuclear morphology of contralateral normal and tumor cells and the distribution of select features. (a) Images of Feulgen-thionin-stained nuclei. (b) Mean of select features for the nuclei pictured in (a). (c) Distribution of select features in contralateral normal (cln), adjacent normal (an), and tumor nuclei. Significance tested using ANOVA of log-10-transformed values with Tukey's post hoc honestly significant difference test. $^*p < 0.001$. ns: not significant.

tissues are less convoluted, have been reported to have high sensitivity and specificity for the detection of DNA aneuploidy, a nuclear morphological marker of neoplastic cell transformation [43]. Thus, the identification of MACs in surface epithelium represents the initial step towards the development of a brush biopsy-based screening method for early-stage oropharyngeal tumors in high-risk populations. Additional research is required to determine if the differences in nuclear phenotype observed in this study can be similarly identified in epithelial cells collected via brush biopsy.
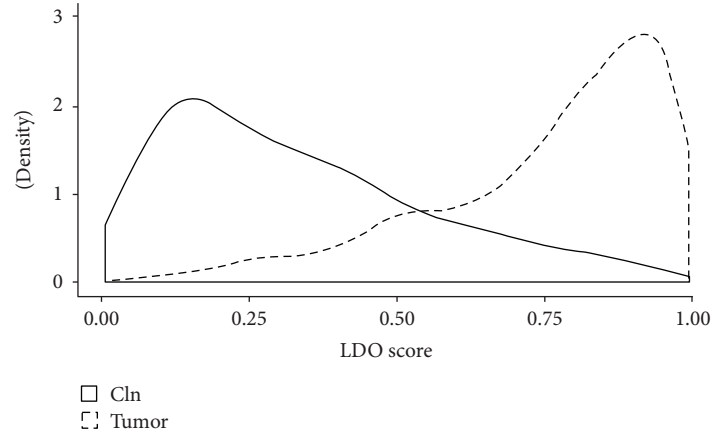
FIGURE 3: Distribution of LDO scores of nuclei derived from contralateral normal (cln) and tumor tissue of the test set.
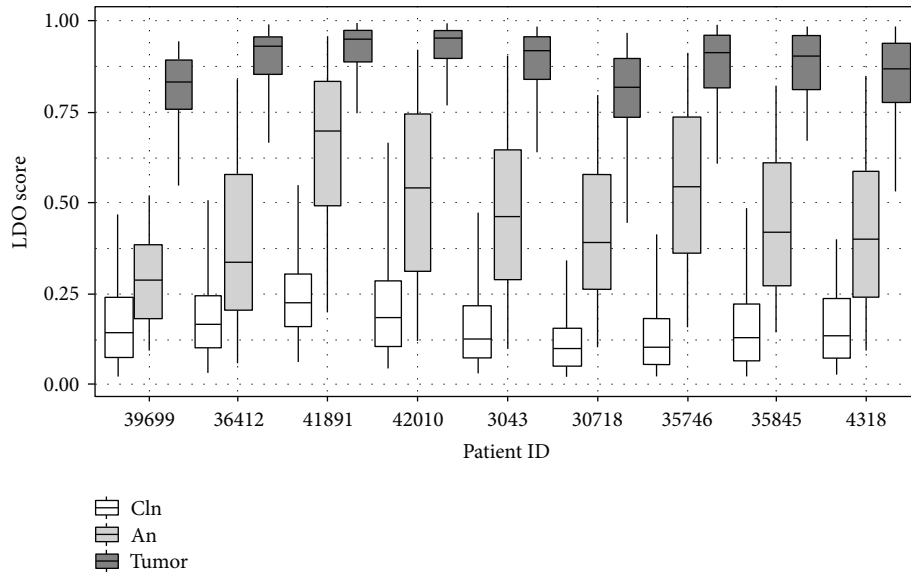

FIGURE 4: Distribution of the LDO scores of nuclei derived from contralateral normal (cln), adjacent normal (an), and tumor tissue grouped by patient. In each case, the mean LDO of the adjacent normal epithelium is intermediate between that of the contralateral normal and the tumor.

Image cytometry shows great clinical promise and has been adopted for the screening of cervical cancer with great success [44]. We used image cytometry to measure nuclear features of contralateral normal, adjacent normal, and tumor epithelial cell nuclei to identify phenotypic differences among them. Of the ten most important nuclear features identified by our model (based on the mean decrease in accuracy), eight were associated with chromatin texture (Table S1). Changes in chromatin are known to accompany malignant transformation and have recently been shown to influence gene expression [45] and genome stability [46], suggesting an active role for chromatin reorganization in tumorigenesis. Adjacent normal epithelial cells show values intermediate between the contralateral normal and tumor for a number of these features, including fractal1_area (which measures the intensity contrast between condensed and uncondensed chromatin, thus revealing the organization

of heterochromatin and euchromatin) and long_runs1 (which measures the random distribution of chromatin within the nucleus). Notably, since the contralateral normal epithelium used in our study was obtained from patients with OPC tumors, it is possible that these cells also contain MACs, but to a lesser degree than adjacent normal cells. MACs have been described in cells contralateral to the tumor of lung cancer patients [10], demonstrating that long-range effects are possible. Thus, analysis of cells from epithelia of individuals without cancer is required to overcome this issue.

Additional evidence for the existence of MACs in adjacent normal epithelial cells comes from our observation that the average LDO score of the adjacent normal epithelium is intermediate between that of the contralateral normal and tumor for each patient examined. Similarly, hyperplastic and mild/moderate dysplastic cells of the oral cavity also display an average LDO score intermediate between that of

normal and tumor tissues [17]. Increased LDO is associated with molecular alterations such as loss of heterozygosity [47] and correlates with an increased risk of cancer progression [17, 47]. Interestingly, the standard deviation of the distribution of LDO scores for adjacent normal nuclei is consistently higher than that of either contralateral normal or tumor nuclei, suggesting greater heterogeneity in nuclear phenotype among adjacent normal epithelial cells.

Even with a small patient cohort, we provide strong evidence that the overall chromatin organization of adjacent normal cells differs from that of the contralateral normal and tumor. Since adjacent normal and contralateral normal epithelia are histologically identical and derive from the same anatomical structure (i.e., surface oropharyngeal epithelium), it is reasonable to conclude that the observed differences are due to the differing proximity of each epithelium to the tumor, a finding that conforms to the definition of MACs.

5. Conclusions

Malignancy-associated changes are reliably detected in histologically normal (normal-appearing) oropharyngeal epithelial cells located adjacent to a tumor and could be used as a noninvasive means of detecting early-stage oropharyngeal tumors.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the University of British Columbia and with the 1964 Declaration of Helsinki and its later amendments.

Consent

Informed consent was obtained from all individual participants included in the study.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

Supplementary Materials

Supplementary Table 1: it provides an overview of patient demographics for all individuals included in this study. Supplementary Table 2: it provides a description of the most important nuclear features as defined by our model. *(Supplementary Materials)*

References

[1] A. K. Chaturvedi, E. A. Engels, R. M. Pfeiffer et al., "Human papillomavirus and rising oropharyngeal cancer incidence in the United States," *Journal of Clinical Oncology*, vol. 29, no. 32, pp. 4294–4301, 2011.

[2] A. Trotti, L. A. Bellm, J. B. Epstein et al., "Mucositis incidence, severity and associated outcomes in patients with head and neck cancer receiving radiotherapy with or without chemotherapy: a systematic literature review," *Radiotherapy and Oncology*, vol. 66, no. 3, pp. 253–262, 2003.

[3] M. Machtay, J. Moughan, A. Trotti et al., "Factors associated with severe late toxicity after concurrent chemoradiation for locally advanced head and neck cancer: an RTOG analysis," *Journal of Clinical Oncology*, vol. 26, no. 21, pp. 3582–3589, 2008.

[4] D. J. Givens, L. H. Karnell, A. K. Gupta et al., "Adverse events associated with concurrent chemoradiation therapy in patients with head and neck cancer," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 135, no. 12, pp. 1209–1217, 2009.

[5] W. K. Mydlarz, J. Y. K. Chan, and J. D. Richmon, "The role of surgery for HPV-associated head and neck cancer," *Oral Oncology*, vol. 51, no. 4, pp. 305–313, 2015.

[6] A. C. Nichols, K. Fung, C. Chapeskie et al., "Development of a transoral robotic surgery program in Canada," *Journal of Otolaryngology–Head and Neck Surgery*, vol. 42, no. 1, pp. 8–11, 2013.

[7] C. Fakhry, B. T. Rosenthal, D. P. Clark, and M. L. Gillison, "Associations between oral HPV16 infection and cytopathology: evaluation of an oropharyngeal "pap-test equivalent" in high-risk populations," *Cancer Prevention Research*, vol. 4, no. 9, pp. 1378–1384, 2011.

[8] T. A. Berman and J. T. Schiller, "Human papillomavirus in cervical cancer and oropharyngeal cancer: one cause, two diseases," *Cancer*, vol. 123, no. 12, pp. 2219–2229, 2017.

[9] B. Palcic, "Nuclear texture: can it be used as a surrogate endpoint biomarker?," *Journal of Cellular Biochemistry*, vol. 19, pp. 40–46, 1994.

[10] C. MacAulay, S. Lam, P. W. Payne, J. C. LeRiche, and B. Palcic, "Malignancy-associated changes in bronchial epithelial cells in biopsy specimens," *Analytical and Quantitative Cytology and Histology*, vol. 17, no. 1, pp. 55–61, 1995.

[11] B. Susnik, A. Worth, and B. Palcic, "Malignancy-associated changes in the breast. Changes in chromatin distribution in epithelial cells in normal-appearing tissue adjacent to carcinoma," *Analytical and Quantitative Cytology and Histology*, vol. 17, no. 1, pp. 62–68, 1995.

[12] A. G. Montag, P. H. Bartels, H. E. Dytch, E. Lerma-Puertas, F. Michaelassi, and M. Bibbo, "Karyometric features in nuclei near colonic adenocarcinoma. Statistical analysis," *Analytical and Quantitative Cytology and Histology*, vol. 13, no. 3, pp. 159–167, 1991.

[13] M. Bibbo, A. G. Montag, E. Lerma-Puertas, H. E. Dytch, S. Leelakusolvong, and P. H. Bartels, "Diagnostic and prognostic relevance of malignancy-associated changes in cervical smears," *Analytical and Quantitative Cytology and Histology*, vol. 11, no. 5, pp. 281–285, 1989.

[14] H. U. Kasper, G. Haroske, U. Geissler, W. Meyer, and K. D. Kunze, "Diagnostic and prognostic relevance of malignancy-associated changes in cervical smears," *Analytical and Quantitative Cytology and Histology*, vol. 19, no. 6, pp. 482–488, 1997.

[15] M. Us-Krasovec, J. Erzen, M. Zganec et al., "Malignancy associated changes in epithelial cells of buccal mucosa: a potential cancer detection test," *Analytical and Quantitative Cytology and Histology*, vol. 27, no. 5, pp. 254–262, 2005.

[16] S. P. Singh, A. Sahu, A. Deshmukh, P. Chaturvedi, and C. Murali Krishna, "*In vivo* Raman spectroscopy of oral buccal mucosa: a study on malignancy associated changes (MAC)/ cancer field effects (CFE)," *Analyst*, vol. 138, no. 14, pp. 4175–4182, 2013.

[17] J. Baik, Q. Ye, L. Zhang et al., "Automated classification of oral premalignant lesions using image cytometry and random forests-based algorithms," *Cellular Oncology*, vol. 37, no. 3, pp. 193–202, 2014.

[18] B. P. Whitaker, W. W. Stigliano, F. L. Carson, and J. A. Lynn, "Thionin Feulgen stain for DNA (ploidy) quantitation by image analysis," *Journal of Histotechnology*, vol. 16, no. 2, pp. 113–116, 1993.

[19] D. Garner, G. Fergusson, and B. Palcic, "Automated cervical cancer screening," in *The Cyto-Savant System*, H. K. Grohs and O. A. N. Husain, Eds., pp. 305–317, Igaku-shoin Medical Publishers, Inc., Hong Kong, 1994.

[20] R. Kamalov, M. Guillaud, D. Haskins et al., "A java application for tissue section image analysis," *Computer Methods and Programs in Biomedicine*, vol. 77, no. 2, pp. 99–113, 2005.

[21] G. Haroske, J. P. Baak, H. Danielsen et al., "Fourth updated ESACP consensus report on diagnostic DNA image cytometry," *Analytical Cellular Pathology*, vol. 23, no. 2, pp. 89–95, 2001.

[22] A. Doudkine, C. MacAulay, N. Poulin, and B. Palcic, "Nuclear texture measurements in image cytometry," *Pathologica*, vol. 87, no. 3, pp. 286–299, 1995.

[23] C. Chen, A. Liaw, and L. Breiman, *Using Random Forest to Learn Imbalanced Data*, vol. 110, University of California, Berkely, 2004.

[24] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[25] M. Kuhn, J. Wing, S. Weston, A. Williams, C. Keefer, and A. Engelhardt, "Caret: classification and regression training," *R package version 6.0-76 The Comprehensive R Archive Network*, 2016, https://CRAN.R-project.org/package=caret.

[26] X. Robin, N. Turck, A. Hainard et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.

[27] L. Masterson, D. Moualed, Z. W. Liu et al., "De-escalation treatment protocols for human papillomavirus-associated oropharyngeal squamous cell carcinoma: a systematic review and meta-analysis of current clinical trials," *European Journal of Cancer*, vol. 50, no. 15, pp. 2636–2648, 2014.

[28] E. J. Moore, S. M. Olsen, R. R. Laborde et al., "Long-term functional and oncologic results of transoral robotic surgery for oropharyngeal squamous cell carcinoma," *Mayo Clinic Proceedings*, vol. 87, no. 3, pp. 219–225, 2012.

[29] R. V. Smith, B. A. Schiff, M. Garg, and M. Haigentz, "The impact of transoral robotic surgery on the overall treatment of oropharyngeal cancer patients," *The Laryngoscope*, vol. 125, pp. S1–S15, 2015.

[30] D. C. Ling, B. V. Chapman, J. Kim et al., "Oncologic outcomes and patient-reported quality of life in patients with oropharyngeal squamous cell carcinoma treated with definitive transoral robotic surgery versus definitive chemoradiation," *Oral Oncology*, vol. 61, pp. 41–46, 2016.

[31] A. M. Hurtuk, A. Marcinow, A. Agrawal, M. Old, T. N. Teknos, and E. Ozer, "Quality-of-life outcomes in transoral robotic surgery," *Otolaryngology–Head and Neck Surgery*, vol. 146, no. 1, pp. 68–73, 2012.

[32] F. D. Leonhardt, H. Quon, M. Abrahão, B. W. O'Malley Jr, and G. S. Weinstein, "Transoral robotic surgery for oropharyngeal carcinoma and its impact on patient-reported quality of life and function," *Head & Neck*, vol. 34, no. 2, pp. 146–154, 2012.

[33] P. T. Dziegielewski, T. N. Teknos, K. Durmis et al., "Transoral robotic surgery for oropharyngeal cancer: long-term quality of life and functional outcomes," *JAMA Otolaryngology–Head & Neck Surgery*, vol. 139, no. 11, pp. 1099–1108, 2013.

[34] K. Motz, H. Y. Chang, H. Quon, J. Richmon, D. W. Eisele, and C. G. Gourin, "Association of transoral robotic surgery with short-term and long-term outcomes and costs of care in oropharyngeal cancer surgery," *JAMA Otolaryngology—Head Neck Surgery*, vol. 143, no. 6, pp. 580–588, 2017.

[35] R. Sethia, A. C. Yumusakhuylu, I. Ozbay et al., "Quality of life outcomes of transoral robotic surgery with or without adjuvant therapy for oropharyngeal cancer," *The Laryngoscope*, vol. 128, pp. 403–411, 2017.

[36] B. H. Haughey, M. L. Hinni, J. R. Salassa et al., "Transoral laser microsurgery as primary treatment for advanced-stage oropharyngeal cancer: a United States multicenter study," *Head & Neck*, vol. 33, no. 12, pp. 1683–1694, 2011.

[37] H. E. Nieburgs, A. F. Goldberg, B. Bertini, J. Silaqi, B. Pacheco, and H. Reisman, "Malignancy associated changes (MAC) in blood and bone marrow cells of patients with malignant tumors," *Acta Cytologica*, vol. 11, no. 5, pp. 415–423, 1967.

[38] M. Guillaud, A. Doudkine, D. Garner, C. MacAulay, and B. Palcic, "Malignancy associated changes in cervical smears: systematic changes in cytometric features with the grade of dysplasia," *Analytical Cellular Pathology*, vol. 9, no. 3, pp. 191–204, 1995.

[39] N. Ikeda, C. MacAulay, S. Lam et al., "Malignancy associated changes in bronchial epithelial cells and clinical application as a biomarker," *Lung Cancer*, vol. 19, no. 3, pp. 161–166, 1998.

[40] E. C. Mommers, N. Poulin, C. J. Meijer, J. P. Baak, and P. J. Diest, "Malignancy-associated changes in breast tissue detected by image cytometry," *Analytical Cellular Pathology*, vol. 20, no. 4, pp. 187–195, 2000.

[41] Y. Man and H. E. Nieburgs, "A subset of cell clusters with malignant features in morphologically normal-appearing and hyperplastic tissues," *Cancer Detection and Prevention*, vol. 30, no. 3, pp. 239–247, 2006.

[42] M. A. Broglie, W. Jochum, D. Förbs, R. Schönegg, and S. J. Stoeckli, "Brush cytology for the detection of high-risk HPV infection in oropharyngeal squamous cell carcinoma," *Cancer Cytopathology*, vol. 123, no. 12, pp. 732–738, 2015.

[43] J. M. Ma, T. J. Zhou, R. Wang et al., "Brush biopsy with DNA-image cytometry: a useful and noninvasive method for monitoring malignant transformation of potentially malignant oral disorders," *European Archives of Oto-Rhino-Laryngology*, vol. 271, no. 12, pp. 3291–3295, 2014.

[44] D. Garner, "Clinical application of DNA ploidy to cervical cancer screening: a review," *World Journal of Clinical Oncology*, vol. 5, no. 5, pp. 931–965, 2014.

[45] L. M. Almassalha, A. Tiwari, P. T. Ruhoff et al., "The global relationship between chromatin physical topology, fractal structure, and gene expression," *Scientific Reports*, vol. 7, article 41061, 2017.

[46] A. Canela, Y. Maman, S. Jung et al., "Genome organization drives chromosome fragility," *Cell*, vol. 170, no. 3, pp. 507–521.e18, 2017.

[47] M. Guillaud, L. Zhang, C. Poh, M. P. Rosin, and C. MacAulay, "Potential use of quantitative tissue phenotype to predict malignant risk for oral premalignant lesions," *Cancer Research*, vol. 68, no. 9, pp. 3099–3107, 2008.