

## An Imaging Biomarker of Tumor-Infiltrating Lymphocytes to Risk-Stratify Patients With HPV-Associated Oropharyngeal Cancer

Germán Corredor, PhD <sup>1,2</sup> Paula Toro, MD,<sup>1</sup> Can Koyuncu, PhD <sup>1</sup> Cheng Lu, PhD,<sup>1</sup> Christina Buzzy, PhD <sup>1</sup> Kaustav Bera, MD,<sup>1</sup> Pingfu Fu, PhD <sup>3</sup> Mitra Mehrad, MD,<sup>4</sup> Kim A. Ely, MD,<sup>4</sup> Mojgan Mokhtari, MD,<sup>1</sup> Kailin Yang, MD, PhD <sup>5</sup> Deborah Chute, MD,<sup>6</sup> David J. Adelstein, MD,<sup>7</sup> Lester D. R. Thompson, MD <sup>8</sup> Justin A. Bishop, MD,<sup>9</sup> Farhoud Faraji, MD, PhD <sup>10</sup> Wade Thorstad, MD <sup>11</sup> Patricia Castro, PhD <sup>12</sup> Vlad Sandulache, MD, PhD,<sup>13,14,15</sup> Shlomo A. Koyfman, MD,<sup>5</sup> James S. Lewis Jr, MD <sup>4</sup> Anant Madabhushi, PhD <sup>1,2,\*</sup>

<sup>1</sup>Department of Biomedical Engineering, Center of Computational Imaging and Personalized Diagnostics, Case Western Reserve University, Cleveland, OH, USA; <sup>2</sup>Louis Stokes Cleveland VA Medical Center, Cleveland, OH, USA; <sup>3</sup>Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; <sup>4</sup>Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA; <sup>5</sup>Department of Radiation Oncology, Cleveland Clinic, Cleveland, OH, USA; <sup>6</sup>Department of Anatomic Pathology, Cleveland Clinic, Cleveland, OH, USA; <sup>7</sup>Department of Medicine, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; <sup>8</sup>Department of Pathology, Southern California Permanente Medical Group, Woodland Hills, CA, USA; <sup>9</sup>Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA; <sup>10</sup>Division of Otolaryngology-Head and Neck Surgery, Department of Surgery, UC San Diego Health, La Jolla, CA, USA; <sup>11</sup>Department of Radiation Oncology, Washington University in St. Louis, St. Louis, MS, USA; <sup>12</sup>Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA; <sup>13</sup>Department of Otolaryngology-Head and Neck Surgery, Baylor College of Medicine, Houston, TX, USA; <sup>14</sup>ENT Section, Operative Care Line, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA; and <sup>15</sup>Center for Translational Research on Inflammatory Disease (CTRID), Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA

\*Correspondence to: Anant Madabhushi, PhD, Center of Computational Imaging and Personalized Diagnostics, Case Western Reserve University, 2071 Martin Luther King Drive, Cleveland, OH 44106-7207, USA (e-mail: anant.madabhushi@case.edu).

### Abstract

**Background:** Human papillomavirus (HPV)-associated oropharyngeal squamous cell carcinoma (OPSCC) has excellent control rates compared to nonvirally associated OPSCC. Multiple trials are actively testing whether de-escalation of treatment intensity for these patients can maintain oncologic equipoise while reducing treatment-related toxicity. We have developed OP-TIL, a biomarker that characterizes the spatial interplay between tumor-infiltrating lymphocytes (TILs) and surrounding cells in histology images. Herein, we sought to test whether OP-TIL can segregate stage I HPV-associated OPSCC patients into low-risk and high-risk groups and aid in patient selection for de-escalation clinical trials. **Methods:** Association between OP-TIL and patient outcome was explored on whole slide hematoxylin and eosin images from 439 stage I HPV-associated OPSCC patients across 6 institutional cohorts. One institutional cohort (n = 94) was used to identify the most prognostic features and train a Cox regression model to predict risk of recurrence and death. Survival analysis was used to validate the algorithm as a biomarker of recurrence or death in the remaining 5 cohorts (n = 345). All statistical tests were 2-sided. **Results:** OP-TIL separated stage I HPV-associated OPSCC patients with 30 or less pack-year smoking history into low-risk (2-year disease-free survival [DFS] = 94.2%; 5-year DFS = 88.4%) and high-risk (2-year DFS = 82.5%; 5-year DFS = 74.2%) groups (hazard ratio = 2.56, 95% confidence interval = 1.52 to 4.32; P < .001), even after adjusting for age, smoking status, T and N classification, and treatment modality on multivariate analysis for DFS (hazard ratio = 2.27, 95% confidence interval = 1.32 to 3.94; P = .003). **Conclusions:** OP-TIL can identify stage I HPV-associated OPSCC patients likely to be poor candidates for treatment de-escalation. Following validation on previously completed multi-institutional clinical trials, OP-TIL has the potential to be a biomarker, beyond clinical stage and HPV status, that can be used clinically to optimize patient selection for de-escalation.

Oropharyngeal squamous cell carcinoma (OPSCC) has shown an important increase in incidence over the past several decades

(1). Transcriptionally active human papillomavirus (HPV) has become the most common cause of OPSCC in the United States

Received: May 18, 2021; Revised: August 3, 2021; Accepted: November 19, 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

(1), with an estimated prevalence of 16 000 patients annually (2). Patients with HPV-associated OPSCC demonstrate improved treatment response to chemoradiotherapy as well as better long-term survival, compared with patients with non-HPV associated OPSCC (3). This has led to the consideration of treatment de-intensification to reduce therapy-related morbidity in low-risk HPV-associated OPSCC patients who are never or light smokers and have low-stage disease (4-6).

A recent phase II randomized controlled trial (NRG-HN002) (6) aimed to identify a suitable de-escalation strategy for low-risk HPV-associated OPSCC patients, defined as those with T1-T2/N1-N2b/M0 or T3/N0-N2b/M0 (American Joint Committee on Cancer [AJCC] 7th ed.) and a 10 or less pack-year smoking history. The patients were randomly assigned to either of 2 arms: 1) dose-reduced radiotherapy (RT) (60 Gy) with weekly cisplatin and 2) accelerated RT (60 Gy) alone. Overall, the dose-reduced chemoradiation arm met the prespecified threshold of a 2-year disease-free survival (DFS) of 85% and is being tested against the standard of care in the NRG-HN005 randomized phase III trial (7). However, a considerable percentage of patients in both arms experienced local-regional failure, and the RT arm alone did not meet the prespecified DFS threshold (6). These data are promising with respect to de-escalation as a general strategy for HPV-associated OPSCC but suggest that additional stratification may be necessary to avoid undertreatment of aggressive tumor biology even in early stage disease.

Tumor-infiltrating lymphocytes (TILs) have been reported in multiple studies as associated with disease outcome in diverse cancer types [eg, breast and lung (8)]. In the context of HPV-associated OPSCC, TILs appear to have a protective effect through an adaptive host immune response directed against viral antigens, and specific lymphocytes against HPV-associated cancer (HPV-16 E7 T cells) have been identified (3). Previous work (9) showed that an increased density of TILs is associated with low risk of recurrence in low-stage HPV-associated OPSCC. Conversely, decreased TIL infiltration, associated with tobacco exposure, has been correlated with both advanced stage at presentation and locoregional recurrence (10). Unfortunately, quantification of TILs is difficult to implement in routine clinical practice because it is hard to standardize across pathologists and may necessitate the use of ancillary methods such as CD4 or CD8 immunohistochemistry (10).

In this study, we employed image processing and machine learning to develop OP-TIL, an imaging biomarker that quantitatively characterizes the spatial patterns of TILs and surrounding nucleated cells in digitized hematoxylin and eosin (H&E) slides of HPV-associated OPSCC patients. OP-TIL was trained to distinguish between the majority of low-risk patients (overall stage I [T1-T2/N0-N1/M0, AJCC 8th ed.] and never or light smokers) who will have a favorable DFS and be appropriate candidates for therapy de-escalation vs those patients who will have a poor DFS and for whom de-intensification would be inappropriate. We performed the analysis on patients with less than 10 pack-year, matching the inclusion criteria of smoking history in NRG-HN002 (6) and also analyzed patients with less than 30 pack-year, a risk cutoff suggested by previous works (10-12).

## Methods

### Dataset

A total of 1485 OPSCC patients with pretreatment primary tumor specimens (biopsies and resections) was retrospectively

collected from 6 institutions: the Michael E. DeBakey Veterans Affairs Medical Center (D1), Johns Hopkins University (D2), Washington University in St. Louis (D3), Southern California Permanente Medical Group (D4), Cleveland Clinic (D5), and Vanderbilt University Medical Center (D6). Fine-needle aspiration specimens were not included in this study because they do not provide tissue with preserved architecture. Corresponding clinicopathologic and outcome information from patients was obtained from the institutions at which the datasets were collected after obtaining the respective institutional review board approvals.

Because the AJCC 8th ed. HPV positivity is determined by p16 testing of tumor tissue (13), immunohistochemistry was performed at the respective institutions in routine clinical practice. Only cases classified as p16 positive by accepted standards [strong and diffuse, block-like nuclear, and cytoplasmic staining present in  $\geq 70\%$  of the tumor specimen (1)] were included in this study. H&E glass slides from each patient were re-reviewed by our collaborating pathologists at each institution for selecting the most representative single tumor slide (ie, the slide that contains the tumor bulk and its leading tumor edge). Patients for whom cautery (thermal) artifact was extensive or for whom minimal tumor was present were excluded. Cases were digitized as whole slide images at 40x resolution (0.25  $\mu\text{m}/\text{pixel}$  resolution) using a Ventana iScan HT scanner.

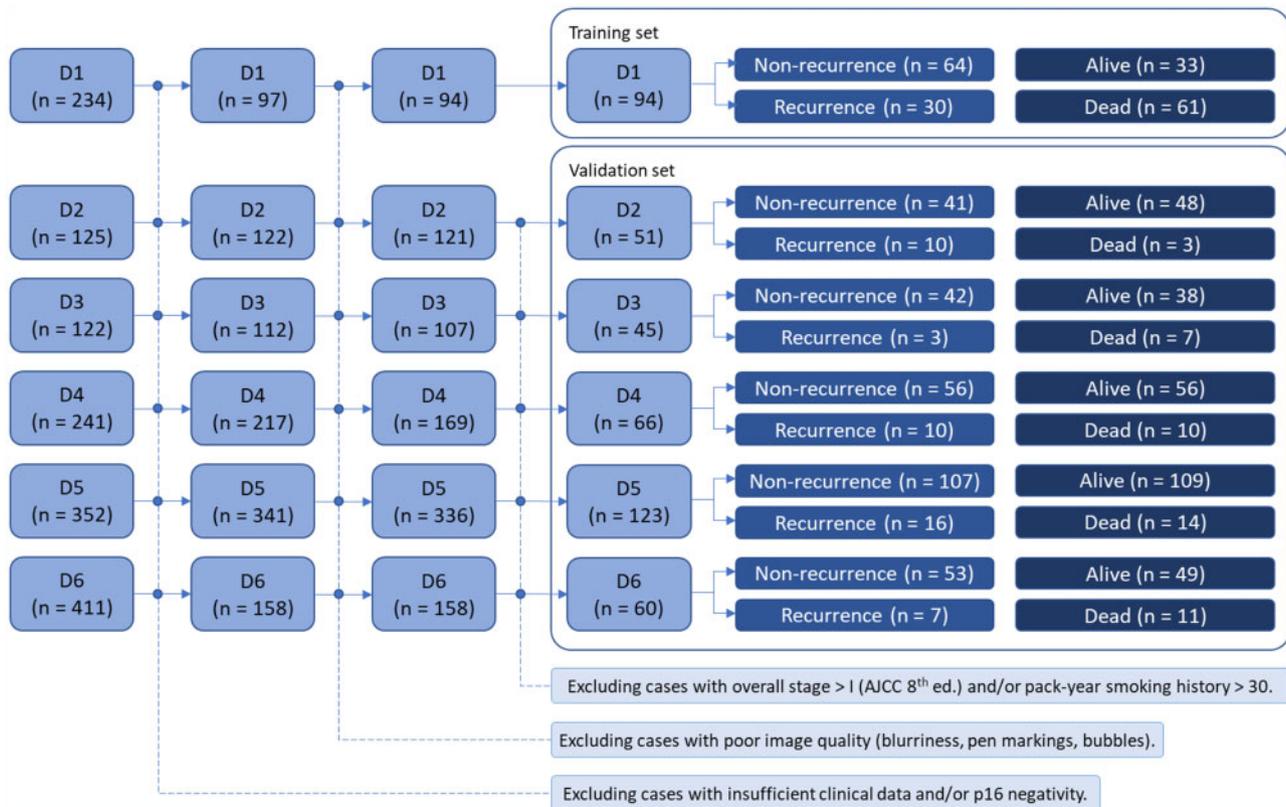
After reviewing the clinical data and p16 status, 438 of the patients were excluded because of negative or equivocal p16 results, insufficient clinical data (ie, no recurrence or death data), or insufficient tumor in the specimens. Image quality was checked using HistoQC (14), an open-source quality control tool for digital pathology slides; 62 cases were identified as inappropriate for the study because the presence of large blurry areas, obstructive dotting pen markings, or subcoverslip bubbles.

Cohort D1 ( $n = 94$ ) was employed for feature discovery and model training because it was previously employed in a study (10) that demonstrated association between TILs and patient survival. Cohorts D2 ( $n = 51$ ), D3 ( $n = 45$ ), D4 ( $n = 66$ ), D5 ( $n = 123$ ), and D6 ( $n = 60$ ) were used for independently evaluating the prognostic ability of OP-TIL in low-risk patients, so patients with overall stages higher than I and with more than 30 pack-year smoking history were excluded from the analysis, as well. Figure 1 illustrates the inclusion and exclusion criteria for patient selection.

### Identification of Cell Types

Figure 2 illustrates the building blocks of the introduced approach. First, for computational feasibility, each whole slide image was split into nonoverlapping 2048  $\times$  2048-pixel image tiles. Tiles containing at least 65% of tissue were retained, and the remaining were discarded. Image color normalization (15) was applied to each tile to compensate staining variations of slides acquired from different institutions.

Then, 2 cell types were automatically identified in each image tile: TILs and non-TILs (Figure 2, see Nuclei segmentation and TIL detection). In this study, following the definition employed by the International TILs Working Group (16,17), we consider TIL as any lymphocyte located within tumoral tissue, which includes both intratumoral and stromal lymphocytes. Although non-TILs include cancer cells (majority), macrophages, fibroblasts, and neutrophils, among other nucleated cells, in this study, they were not differentiated but treated as a single type. First, every individual cell nucleus in each tile was



**Figure 1.** Patient selection workflow for the datasets included in this study. D1 was employed for feature discovery and model training, and datasets D2-D6 were used for independent validation of the prognostic ability of the OP-TIL classifier. AJCC = American Joint Committee on Cancer.

segmented automatically using the deep-learning (18,19) model developed by Mahmood et al. (20). Subsequently, another validated machine learning model (21) was used to classify each segmented nucleus as either TIL or non-TIL based off visual features (texture, shape, and color). More details about the segmentation and TIL detection models are provided in the [Supplementary Methods](#) (available online).

### Feature Extraction

Once TILs were detected, the spatial interplay between TILs and non-TILs was characterized. The procedure involved first constructing clusters of TILs. A TIL was linked to another TIL if the distance between them was below a predefined threshold. A very large threshold value generates just a single graph, limiting the analysis of spatial interactions to within the microenvironment, whereas a small value (close to zero) produces multiple sparse subgraphs, leading to an analysis of individual nuclei. The value of the threshold was empirically determined to be 95 pixels (approximately 24  $\mu\text{m}$ ) because it showed a reasonable trade-off in the number of generated subgraphs. This action was repeated for each TIL until all of them were interrogated, resulting in a set of disconnected subgraphs of TILs (Figure 2, see Construction of cell subgraphs). Then, for each TIL subgraph, a polygon was constructed as the smallest convex set that contains all the subgraph cells (the convex hull) (Figure 2, see Construction of cell clusters). TILs not linked to other cells cannot form clusters, so they were excluded from the subsequent analysis. This process was also repeated for non-TILs.

From the constructed clusters of TILs and non-TILs within each tile, features related to density, architecture, and

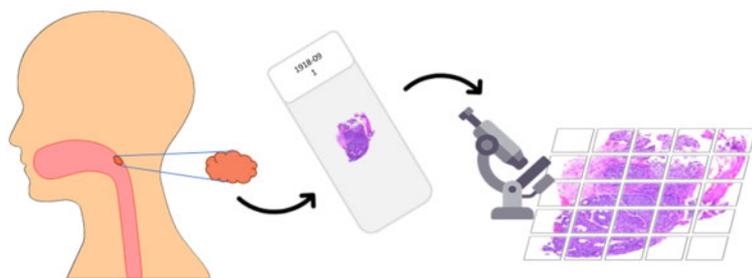
colocalization were extracted (Table 1; Supplementary Figure 1, available online). The final feature vector for each patient was obtained by computing 8 metrics (total, mean, standard deviation, median, minimum, maximum, skewness, and kurtosis) for each feature across all its constituent tiles (2952 features per patient). Some of these metrics aim to model the inpatient heterogeneity (23).

### Statistical Analysis

DFS is the time from the date of diagnosis to the date of first occurrence of local, regional, or distant recurrence, or death from any cause and was censored at the date of last follow-up for those alive without recurrence. Overall survival (OS) is the time interval between the date of diagnosis and the date of death and was censored at the date of last follow-up for those alive.

A Cox proportional hazards regression model (24) in conjunction with the least absolute shrinkage and selection operator (25) was used to identify the top OP-TIL features (ie, those that are most associated with patient outcome) in the training set (D1) along with their coefficients (which indicate the importance of each feature) for both DFS and OS. The assumption of proportionality was verified using a time-varying coefficient model. Then, DFS and OS risk scores were computed for each patient in D1 as a linear combination of the top feature values and their respective coefficients. Finally, the median value of all risk scores of patients in D1 was computed and used as a cutoff to discriminate among patients at low or high risk. A patient whose risk score value is higher than the median is considered high risk, and a patient with a risk score value lower than the median is considered low risk.

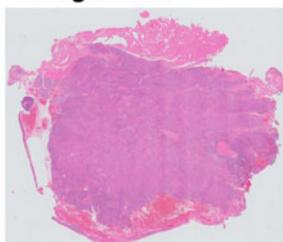
## 1 Sample acquisition and digitization



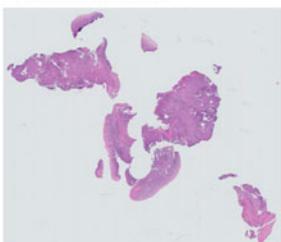
Samples from 94 patients were used to develop **OP-TIL**, an imaging biomarker for identifying low-risk HPV-associated OPSCC patients who could not benefit from therapy de-escalation. Its prognostic ability was tested on samples from 345 patients (overall stage I and light/never smokers) treated at 5 independent institutions.

## 2 Whole slide images

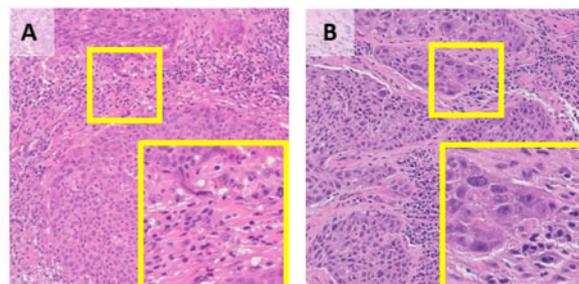
A. Higher risk



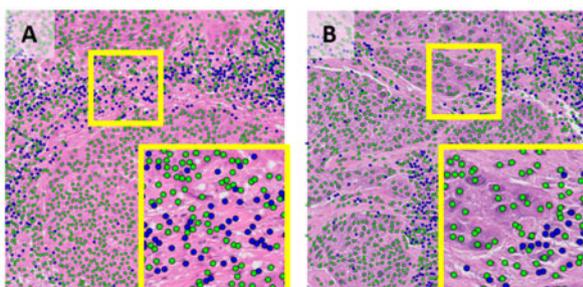
B. Lower risk



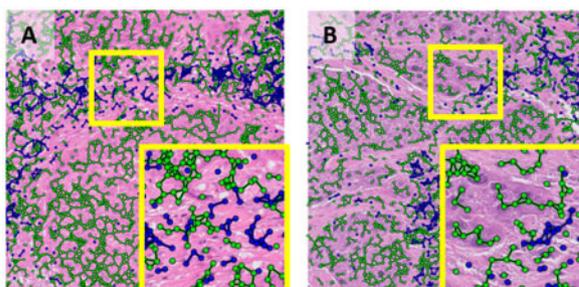
## 3 Enlarged fields of view



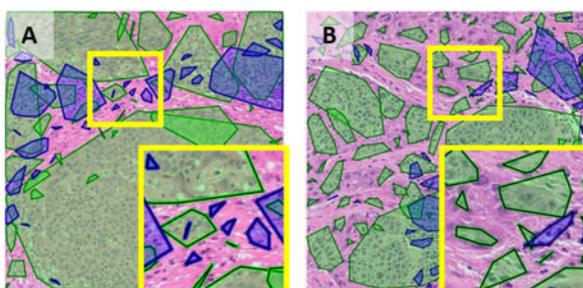
## 4 Nuclei segmentation & TIL detection



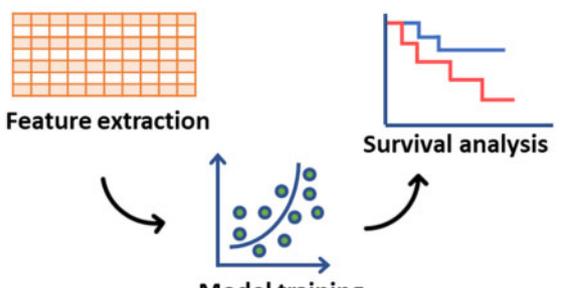
## 5 Construction of cell subgraphs



## 6 Construction of cell clusters



## 7 Model development & validation



**Figure 2.** Illustration of OP-TIL building blocks. A) Corresponds to a patient with higher risk of disease recurrence and B) corresponds to a patient who has a lower risk. TILs are represented with blue and non-TILs with green (non-TILs include different cells in the tumor microenvironment, such as cancer cells, macrophages, and fibroblasts, among others). HPV = human papillomavirus; OPSCC = oropharyngeal squamous cell carcinoma; TIL = tumor-infiltrating lymphocyte.

**Table 1.** Set of OP-TIL features, related to density, architecture, and colocalization extracted from the different cell clusters of TILs and non-TILs<sup>a</sup>

Set	No. of features	Extracted from	Description
1	34	Nuclei clusters	Number, size, and density of clusters of each type.
2	32	Nuclei clusters	Area intersected between clusters of the same and different type.
3	160	Nuclei clusters	Number of clusters surrounding a specific cluster type.
4	6	Centroids of the nuclei clusters	A new cluster was built for each type by drawing a convex hull containing all its centroids. From the resulting new clusters, the intersected area was computed.
5	102	Centroids of the nuclei clusters	Measures from global graphs of each type (Voronoi diagram, Delaunay triangulation, and minimum spanning tree).
6	16	Centroids of the nuclei clusters	Closeness of each nuclei type. This value was computed using a published metric (22) that assigns to each node (the cluster centroid) a value depending on how it is arranged with respect to other nodes. A high value indicates that that specific node is proximal to multiple nodes, whereas a low value implies that node is isolated.
7	19	Individual nuclei	Features related to quantity and compactness of TILs with respect to non-TILs.

<sup>a</sup>TIL = tumor-infiltrating lymphocyte.

For validation purposes, the risk score of each patient in the testing sets (D2-D6) is computed using the feature coefficients found in training. Then, each patient is classified as either high risk, if his or her computed risk score is higher than the cutoff defined in training, or low risk otherwise.

Kaplan-Meier survival analysis with the log-rank test was used to examine the differences of time-to-event data (DFS and OS) between low-risk patient groups categorized by the OP-TIL risk classifier. Multivariable Cox regression analysis with Firth's Penalized Likelihood (26) was employed to examine the prognostic ability of the OP-TIL classifier when controlling the effects of clinical and pathological parameters. *P* values were 2-sided, and all values .05 or less were considered statistically significant.

## Results

### Clinicopathologic Features of the Patient Cohorts

Among the 439 eligible patients, the median age was 57 years, 41.7% had T1 disease, 74.9% had N1 disease, and the median follow-up was 61 months. Of the patients, 81.5% achieved a 2-year DFS and 88.4% a 2-year OS. Table 2 shows a summary of clinical and pathological features of cohorts D1-D6, individually.

### Prognostic Ability of OP-TIL

The most prognostic features, identified by the least absolute shrinkage and selection operator, for DFS and OS were related to spatial arrangement features of TILs and non-TILs, to closeness of TILs (Table 1, set 6), and to density of non-TILs (Supplementary Table 1, available online.). Then, a linear combination of the top feature values and their respective coefficients generated a risk score for each patient. Patients with lower OP-TIL risk scores tended to have more TIL clusters intermixed with non-TIL clusters, the closeness/compactness among TILs was more variable, and the density of non-TIL clusters was smaller compared with patients with high OP-TIL risk scores. Qualitative analysis, performed using the t-stochastic neighbor-embedding algorithm (27), suggests that OP-TIL features are resistant to batch effects (Supplementary Figure 2, available online). Analysis of the correlation between OP-TIL risk scores for DFS and OS is included in the Supplementary Methods (available online).

Figure 3 illustrates the Kaplan-Meier plots for OP-TIL applied to different subgroups of patients in the validation set (D2-D6) using DFS as endpoint. A total of 88 (36.6%) patients with less than 10 pack-year smoking history was classified as OP-TIL high risk (2-year DFS = 81.8%; 5-year DFS = 72.7%) and 152 (63.3%) patients as low risk (2-year DFS = 96.1%; 5-year DFS = 90.8%). Similarly, 120 (34.7%) patients with less than 30 pack-year smoking history were classified as high risk (2-year DFS = 82.5%; 5-year DFS = 74.2%) and 225 (65.2%) as low risk (2-year DFS = 94.2%; 5-year DFS = 88.4%).

TIL was prognostic of DFS for patients with overall stage I and with less than 10 and 30 pack-year smoking history. The hazard ratios were 3.47 (95% confidence interval [CI] = 1.79 to 6.72; *P* < .001) and 2.56 (95% CI = 1.52 to 4.32; *P* < .001), respectively. Furthermore, multivariable survival analysis showed that OP-TIL was prognostic independent of age, T and N stages, treatment, and smoking history (number of pack-year) for DFS (Table 3) with a hazard ratio of 2.27 (95% CI = 1.32 to 3.94; *P* = .003).

Although DFS was the main focus of this study because that was the metric employed in NRG-HN002 to demonstrate the safety of de-escalation, we also assessed the performance of OP-TIL for OS. OP-TIL was prognostic for patients with overall stage I and less than 30 pack-year with a hazard ratio of 2.34 (95% CI = 1.08 to 5.07; *P* = .03) (Supplementary Figure 3, available online). Multivariable analysis illustrated that OP-TIL was also prognostic independent of age, T and N stages, treatment, and smoking for OS (Supplementary Table 2, available online).

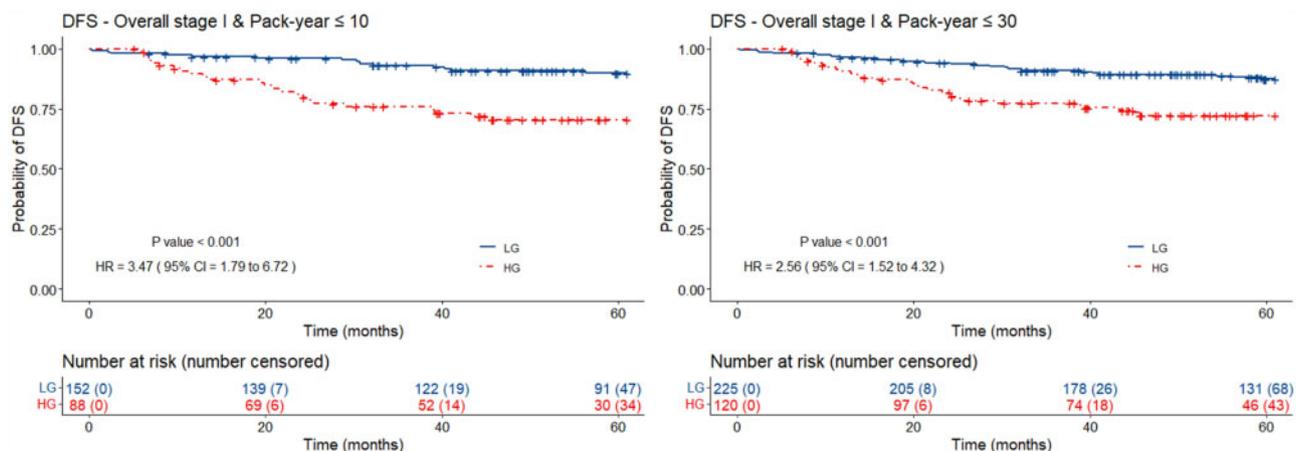
## Discussion

Although definitive chemoradiation therapy is often curative for patients with HPV-associated OPSCC, the resultant toxicity can affect the quality of life of patients meaningfully (6). This has motivated clinicians to find patients who are at low risk of recurrence and death and whose treatment could be de-escalated to reduce the secondary acute and late effects of chemoradiation therapy without affecting the cure rates (7). Recently, Yom et al. (6) published the results of NRG-HN002, in which low-risk HPV-associated OPSCC patients were randomly assigned to either of 2 de-escalated treatment arms: 1) dose-reduced RT with weekly cisplatin or 2) accelerated RT (60 Gy) alone. Patients in the chemoradiation therapy arm met the target 2-year DFS of 85% or higher, where those in the RT-alone

**Table 2.** Summary of clinical and pathological features of the studied HPV-associated OPSCC cohorts

Variable	Cohort					
	D1	D2	D3	D4	D5	D6
Total patients, No. (%)	94 (21.4)	51 (11.6)	45 (10.3)	66 (15.0)	123 (28.0)	60 (13.7)
Median follow-up, mo	57.65	52.14	45.8	83.67	65.79	70.59
Age, No. (%), y						
> 55	67 (71.3)	28 (54.9)	23 (51.1)	30 (45.5)	80 (65.0)	29 (48.3)
≤ 55	27 (28.7)	23 (45.1)	22 (48.9)	36 (54.6)	43 (35.0)	31 (51.7)
Race, No. (%)						
Caucasian	—	48 (94.1)	43 (95.6)	62 (93.9)	117 (95.1)	58 (96.7)
Non-Caucasian	—	3 (5.9)	2 (4.4)	4 (6.1)	6 (4.9)	2 (3.3)
No data	94 (100)	—	—	—	—	—
Sex, No. (%)						
Male	93 (98.9)	45 (88.2)	37 (82.2)	57 (86.4)	110 (89.4)	54 (90.0)
Female	1 (1.1)	6 (11.8)	8 (17.8)	9 (13.6)	13 (10.6)	6 (10.0)
Smoking pack-years, No. (%)						
0	17 (18.1)	21 (41.2)	23 (51.1)	23 (34.9)	59 (48.0)	39 (65.0)
1–10	10 (10.6)	18 (35.3)	14 (31.1)	13 (19.7)	24 (19.5)	6 (10.0)
11–30	24 (25.5)	12 (23.5)	8 (17.8)	30 (45.5)	40 (32.5)	15 (25.0)
No data	2 (2.1)	—	—	—	—	—
Overall stage by AJCC 8th edition, No. (%)						
I	16 (17.0)	51 (100)	45 (100)	66 (100)	123 (100)	60 (100)
II/III/IV	78 (83.0)	—	—	—	—	—
T stage, No. (%)						
T1/T2	55 (58.5)	51 (100)	45 (100)	66 (100)	123 (100)	55 (91.7)
T3/T4	39 (41.5)	—	—	—	—	—
No data	—	—	—	—	—	5 (8.3)
N stage, No. (%)						
N0/N1	23 (24.5)	51 (100)	45 (100)	66 (100)	123 (100)	58 (96.7)
N2/N3	71 (75.5)	—	—	—	—	—
No data	—	—	—	—	—	2 (3.3)
Treatment, No. (%)						
Surgery + adjuvant therapy	3 (3.2)	15 (29.4)	30 (66.7)	66 (100)	—	25 (41.7)
Surgery alone	—	1 (2.0)	13 (28.9)	—	—	6 (10.0)
Primary chemoradiation	91 (96.8)	35 (68.6)	2 (4.4)	—	123 (100)	26 (43.3)
No data	—	—	—	—	—	3 (5.0)

<sup>a</sup>AJCC = American Joint Committee on Cancer; HPV = human papillomavirus; OPSCC = oropharyngeal squamous cell carcinoma.



**Figure 3.** Kaplan-Meier plots for the DFS OP-TIL classifier applied to patients in the validation set (D2-D6) with overall stage I [AJCC 8th ed. (13)], and with less than 30 pack-year of smoking history. Patients with less than 10- and 30 pack-year classified by OP-TIL as high risk (dashed line) are approximately 3 and 2 times, respectively, more likely to develop disease recurrence and/or die. P values were 2-sided and computed using the log-rank test. AJCC = American Joint Committee on Cancer; CI = confidence interval; DFS = disease-free survival; HG = high-risk group; HR = hazard ratio; LG = low-risk group.

**Table 3.** Univariable and multivariable survival analyses for disease-free survival including all comers ( $\leq 30$  pack-year smoking history) in the testing sets (D2-D6)<sup>a</sup>

Variable	Univariable		Multivariable	
	HR (95% CI)	p <sup>b</sup>	HR (95% CI)	p <sup>b</sup>
Age ( $\geq 55$ vs $< 55$ years) <sup>c</sup>	1.63 (0.96 to 2.75)	.10	1.05 (1.02 to 1.09)	.001
Smoking ( $\geq 10$ vs $< 10$ pack-years)	1.18 (0.68 to 2.03)	.55	1.00 (0.97 to 1.02)	.76
T stage (T1 vs T2)	2.52 (1.49 to 4.25)	.001	2.28 (1.27 to 4.27)	.005
N stage (N0 vs N1)	1.14 (0.38 to 3.41)	.83	1.32 (0.50 to 4.87)	.61
Treatment (surgery + AT vs others)	0.94 (0.55 to 1.61)	.82	1.07 (0.61 to 1.83)	.81
OP-TIL (low vs high risk)	2.56 (1.52 to 4.32)	$< .001$	2.27 (1.32 to 3.94)	.003

<sup>a</sup>For univariable analysis, age and smoking were dichotomized, whereas for multivariable, they were used continuously. AT = adjuvant therapy; CI = confidence interval; HR = hazard ratio.

<sup>b</sup>P values were 2-sided and computed using the log-rank test.

<sup>c</sup>The cutoff for age was set to 55 years, as suggested by Thompson et al. (28).

arm did not. The former regimen is being tested in a confirmatory phase III standard-of-care setting 3-arm trial, NRG-HN005. This study plans to accrue more than 700 participants by February 2025, the results of which will not be known for another several years thereafter (7). It is important to mention that a subset of the patients of such NRG-HN002 trial developed local-regional failure and/or distant metastasis, showing that the definition of low risk employed was not granular enough for identifying patients who can or cannot benefit from therapy de-escalation. It is important to note, however, that HPV-associated OPSCC biology is not homogeneous and that several recent studies have identified (29) distinct biological mechanisms that are associated with an increased risk of treatment failure and/or recurrence. Interestingly, multivariable analysis showed that T stage is also prognostic of DFS, suggesting the association between tumor size and aggressiveness, as has been reported for multiple cancer types (30).

In this work, we present OP-TIL, an imaging biomarker, to characterize the spatial architecture patterns of TILs and surrounding nucleated cells in H&E images of HPV-associated OPSCC patients. OP-TIL was found to be associated with DFS in low-risk patients (overall stage I [AJCC 8th ed.] and with less than 10 packs-year of smoking). Previous work (8,31-33) has shown the importance of the spatial location of immune cells for predicting patients in various cancer types such as lung, breast, prostate, and ovary. However, this work is the first demonstration of computationally derived spatial patterns of TIL architecture as a prognostic marker in HPV-associated OPSCC patients, a disease that presents with complex biologic and oncologic patterns due to the presence of transcriptionally active high-risk HPV (34). In this multi-institutional study, OP-TIL was able to identify patients who, according to current risk definitions (low-stage disease and never or light smokers), could be considered for therapy de-escalation but, because of their unfavorable biology, are likely to have a poor prognosis. OP-TIL was trained using a dataset containing 94 OPSCC patients. Although the size of this training dataset was smaller than the other validation datasets, it contained a larger number of patients who experienced death (64.9%) and recurrence (31.9%). We hypothesize that for training robust generalizable models, a large dataset is not the only condition, as we are carrying out time-to-event analysis, the percentage of events (or censoring) is crucial.

OP-TIL is not the first biomarker to be studied for this application; alternative risks measurements have been explored, including liquid (eg, saliva) and tissue-based biomarkers that can screen tumors through genomic and proteomic molecular

targets (35,36). These methods, however, have high costs, restricted availability, and unclear risk stratification, which limits clinical applicability (37,38). Other features, such as tumor cell anaplasia and multinucleation on histopathologic examination, are usually very focal, when present, so may not even be sampled in a small biopsy sample, which is frequently all the material that is available for diagnosis prior to treatment. When actually present in the tissue sample, they are subject to the inherent quantitative subjectivity and inconsistency of human review (37). Previous work (9) has shown TILs to have a great potential to stratify patients into high-risk and low-risk groups for low-stage HPV-associated OPSCC. Unfortunately, TIL counting is cumbersome, time-consuming, and subject to intra- and interpathologist variability (31). In contrast, OP-TIL is easy to quantify and universally available because of the availability of high-resolution digital slide scanning. This can be done locally or, if an institution does not have its own scanner, a representative H&E slide could easily be shipped to a facility for digitization. OP-TIL represents a potentially powerful, low-cost, and easy-to-scale imaging-based biomarker that may be able to select out the patients from the so-called low-risk HPV-associated OPSCC cohort who are nevertheless destined to recur. This can spare them from ill-advised de-escalated therapy and thereby enhance the outcomes for the remaining truly low-risk patients for whom de-escalation is more appropriate.

A paper that is related to the work presented here is that by Kemnade et al. (10). In that study, the authors used immunohistochemistry to demonstrate that high density of CD4 and CD8 TILs is associated with prognosis in OPSCC (both HPV-associated and HPV-independent) patients with extensive tobacco exposure (median pack-years = 40). They found that tobacco exposure was correlated with decreased CD8 infiltration in HPV-associated OPSCC. Although the study by Kemnade et al. (10) has shown the prognostic relevance of TIL density, OP-TIL features demonstrate the added value of analyzing the spatial interplay and co-localization of TILs and surrounding nuclei for prognosticating DFS in low-risk OPSCC and specifically on H&E-stained slides, not immunostains.

This study has limitations. We used p16 as a surrogate marker of high-risk HPV status; although p16 immunohistochemistry is a thoroughly proven prognostic marker in OPSCC, it is not a perfect surrogate of high-risk HPV status. Our findings using standard H&E samples suggest that interplay between immune and nonimmune cells plays a role in tumor biology and prognosis, but in this work, we did not subtype either the immune or nonimmune cells. Automatic identification of subtypes of TILs (eg, CD4, CD8, and CD20) as well as other cellular

subgroups (eg, cancerous cells, macrophages, fibroblasts, among others) could meaningfully enrich this type of work in the future. Additionally, experimental results showed that OP-TIL was prognostic for DFS but not for OS when applied to patients with less than 10 pack-year of smoking. A possible reason for this is the low number of deaths experienced by light or never smoker HPV-positive OPSCC patients with overall stage I. However, OP-TIL was able to risk stratify patients for DFS and OS with a less than 30 pack-year smoking history. Determination of an adequate cutoff for tobacco exposure is still an open question as different studies (6,10-12) have used different threshold values over the last decade. In this study, a single slide was used for each patient because this approach applies to both patients who underwent biopsy only and those who underwent surgical resection. However, future work may benefit from analysis of all tumor-containing slides (including lymph node metastases) and identifying differences in the algorithm performance when using 1 vs multiple slides. A comprehensive analysis of the implications of the threshold value employed to build TIL and non-TIL clusters for characterizing patients could meaningfully enrich future work on quantitative image analysis. This study was carried out using retrospectively collected data, always subject to known and unknowable sources of bias. As such, validation studies on specimens from cooperative group clinical trials (RTOG 0129, 0522) are in process. We also plan to explore the predictive capability of OP-TIL by performing validation studies on clinical trial datasets in which alternative treatment regimens were used (eg, RT + cetuximab in RTOG 1016 and RT alone in NRG-HN002).

In summary, we have developed and validated a computational TIL-based biomarker, OP-TIL, that exploits features related to the spatial architectural patterns of TILs and non-TILs using standard H&E samples on both biopsies and resection specimens alike. This approach was able to identify low-risk HPV-associated OPSCC patients enriched for poor outcomes in whom treatment de-escalation should likely be avoided. If validation studies confirm these preliminary observations, OP-TIL could be a biomarker beyond clinical stage and tobacco exposure to help clinically optimize patient selection for treatment de-escalation.

## Funding

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (under award numbers 1U24CA199374-01, R01CA202752-01A1, R01CA208236-01A1, R01CA216579-01A1, R01CA220581-01A1, 1U01 CA239055-01, R01CA249992-01A1, R01CA257612-01A1, 1U01CA239055-01, 1U01CA248226-01, 1U54CA254566-01); the National Heart, Lung and Blood Institute 1R01HL15127701A1; the National Institute for Biomedical Imaging and Bioengineering 1R43EB028736-01; the National Center for Research Resources under award number 1 C06 RR12463-01; the VA Merit Review Award IBX004121A from the United States Department of Veterans Affairs Biomedical Laboratory Research and Development Service; the Department of Defense (DOD) Breast Cancer Research Program Breakthrough Level 1 Award W81XWH-19-1-0668; the DOD Prostate Cancer Idea Development Award (W81XWH-15-1-0558); the DOD Lung Cancer Investigator-Initiated Translational Research Award (W81XWH-18-1-0440); the DOD Peer Reviewed Cancer

Research Program (W81XWH-16-1-0329); the Ohio Third Frontier Technology Validation Fund; the Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering; the Clinical and Translational Science Award Program (CTSA) at Case Western Reserve University; the National Cancer Institute Cancer Center Support Grant P30CA125123; the Career Development Award IK2 CX001953 from the United States Department of Veterans Affairs Clinical Sciences Research and Development Program; and the Computational Genomic Epidemiology of Cancer Program at Case Comprehensive Cancer Center (T32CA094186).

## Notes

**Role of the funders:** The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

**Disclosures:** Dr Madabhushi is an equity holder in Elucid Bioimaging and in Inspirata Inc. In addition, he has served as a scientific advisory board member for Inspirata Inc, AstraZeneca, Bristol Myers Squibb and Merck. Currently, he serves on the advisory board of Aiforia Inc. He also has sponsored research agreements with Philips, AstraZeneca, Boehringer-Ingelheim, and Bristol Myers Squibb. His technology has been licensed to Elucid Bioimaging. He is also involved in a NIH U24 grant with PathCore Inc, and 3 different R01 grants with Inspirata Inc. Dr Koefman is a consultant for Merck and Regeneron, he has sponsored research agreements with Bristol Myers Squibb and Merck, and he receives honoraria from UpToDate. No potential conflicts of interest were disclosed by the other authors.

**Author contributions:** Conceptualization: GC, PT, CK, CL, KB, JSL, AM; Data curation: GC, PT, CK, KB, JSL; Formal analysis: GC, PT, CK, PF, AM; Funding acquisition: JSL, AM; Investigation: GC, PT, MMe. KAE, MMo, JSL; Methodology: GC, PT, CK, CL, VS, SAK, JSL, AM; Project administration: KB, CB, JSL, AM; Resources: CL, CB, KB, KY, DC, DJA, LDRT, JAB, FF, WT, PC, VS, SAK, JSL, AM; Software: GC, CK, CL; Supervision: VS, SAK, JSL, AM; Validation: GC, PT, CK; Visualization: GC, PT; Writing—original draft preparation: GC, PT; Writing—review and editing: All.

**Disclaimers:** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the US Department of Veterans Affairs, the DOD, or the US government.

**Prior presentations:** Poster presentations at the United States and Canadian Academy of Pathology (USCAP) Annual Meeting 2020, ASCO Annual Meeting 2020, and USCAP Annual Meeting 2021.

## Data Availability

The data underlying this article were provided by the involved institutions (ie, Michael E. DeBakey Veterans Affairs Medical Center, Johns Hopkins University, Washington University in St. Louis, Southern California Permanente Medical Group, Cleveland Clinic, and Vanderbilt University Medical Center) under licence / by permission. Data will be shared on request to

the corresponding author with permission of the involved institutions.

## References

- Lewis JS, Beadle B, Bishop JA, et al. Human papillomavirus testing in head and neck carcinomas: guideline from the College of American Pathologists. *Arch Pathol Lab Med*. 2018;142(5):559–597. doi:10.5858/arpa.2017-0286-CP
- Viens LJ, Henley SJ, Watson M, et al. Human papillomavirus-associated cancers—United States, 2008–2012. *MMWR Morb Mortal Wkly Rep*. 2016. 2016; 65(26):661–666. doi:10.15585/mmwr.mm6526a1
- Ward MJ, Thirdborough SM, Mellows T, et al. Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. *Br J Cancer*. 2014;110(2):489–500.
- Chen TC, Wu CT, Ko JY, et al. Clinical characteristics and treatment outcome of oropharyngeal squamous cell carcinoma in an endemic betel quid region. *Sci Rep*. 2020;10(1):526.
- Nichols AC, Lang P, Prisman E, et al. Treatment de-escalation for HPV-associated oropharyngeal squamous cell carcinoma with radiotherapy vs. trans-oral surgery (ORATOR2): study protocol for a randomized phase II trial. *BMC Cancer*. 2020;20(1):125.
- Yom SS, Torres-Saavedra P, Caudell JJ, et al. Reduced-dose radiation therapy for HPV-associated oropharyngeal carcinoma (NRG Oncology HN002). *J Clin Oncol*. 2021;39(9):956–965. doi:10.1200/J Clin Oncol.20.03128.
- Chundury A, Kim S. Radiation dose de-escalation in HPV-positive oropharynx cancer: when will it be an acceptable standard of care? *J Clin Oncol*. 2021; 39(9):947–949. doi:10.1200/J Clin Oncol.21.00017.
- Saltz J, Gupta R, Hou L, et al.; for the Cancer Genome Atlas Research Network. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep*. 2018;23(1): 181–193.e7. doi:10.1016/j.celrep.2018.03.086.
- Faraji F, Fung N, Zaidi M, et al. Tumor-infiltrating lymphocyte quantification stratifies early-stage human papillomavirus oropharynx cancer prognosis. *Laryngoscope*. 2020;130(4):930–938. doi:10.1002/lary.28044.
- Kemnade JO, Elhalawani H, Castro P, et al. CD8 infiltration is associated with disease control and tobacco exposure in intermediate-risk oropharyngeal cancer. *Sci Rep*. 2020;10(1):243. doi:10.1038/s41598-019-57111-5.
- Elhalawani H, Mohamed ASR, Elgohari B, et al. Tobacco exposure as a major modifier of oncologic outcomes in human papillomavirus (HPV) associated oropharyngeal squamous cell carcinoma. *BMC Cancer*. 2020;20(1):912. doi: 10.1186/s12885-020-07427-7.
- Vawda N, Banerjee RN, Debenham BJ. Impact of smoking on outcomes of HPV-related oropharyngeal cancer treated with primary radiation or surgery. *Int J Radiat Oncol Biol Phys*. 2019;103(5):1125–1131. doi: 10.1016/j.ijrobp.2018.11.046.
- Amin MB, Edge SB; for the American Joint Committee on Cancer, American Cancer Society. *AJCC Cancer Staging Manual* 8th ed. Chicago, IL: American Joint Committee on Cancer, Springer; 2017:113–122.
- Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistoQC: an open-source quality control tool for digital pathology slides. *J Clin Oncol Clin Cancer Inform*. 2019;3:1–7. doi:10.1200/cci.18.00157.
- Macenko M, Niethammer M, Marron JS, et al. A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Institute of Electrical and Electronics Engineers (IEEE); 2009:1107–1110.
- Salgado R, Denkert C, Demaria S, et al.; for the International TILs Working Group 2014. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol*. 2015;26(2):259–271. doi:10.1093/annonc/mdu450.
- Hendry S, Salgado R, Gevaert T, et al. Assessing tumor-infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in melanoma, gastrointestinal tract carcinomas, non-small cell lung carcinoma and mesothelioma, endometrial and ovarian carcinomas, squamous cell carcinoma of the head and neck, genitourinary carcinomas, and primary brain tumors. *Adv Anat Pathol*. 2017;24(6):311–335. doi:10.1097/PAP.0000000000000161.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444. doi:10.1038/nature14539.
- Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J Pathol Inform*. 2016; 7(1):29. doi:10.4103/2153-3539.186902.
- Mahmood F, Borders D, Chen R, et al. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans Med Imag*. 2019;39(11):3257–3267.
- Corredor G, Wang X, Lu C, Velcheti V, Romero E, Madabhushi A. A watershed and feature-based approach for automated detection of lymphocytes on lung cancer images. In: Tomaszewski JE, Gurcan MN, eds. *SPIE Medical Imaging*. Houston, TX: International Society for Optics and Photonics; 2018.
- Corredor G, Whitney J, Arias V, Madabhushi A, Romero E. Training a cell-level classifier for detecting basal-cell carcinoma by combining human visual attention maps with low-level handcrafted features. *J Med Imaging (Bellingham)*. 2017;4(2):021105–021105.
- Gough A, Stern AM, Maier J, et al. biologically relevant heterogeneity: metrics and practical insights. *SLAS Discov Adv Discov*. 2017;22(3):213–237. doi: 10.1177/2472555216682725.
- Cox DR. Regression models and life-tables. *J R Stat Soc Ser B Methodol*. 1972; 34(2):187–202. doi:10.1111/j.2517-6161.1972.tb00899.x.
- Tibshirani R. The LASSO method for variable selection in the Cox model. *Statist Med*. 1997;16(4):385–395. doi:10.1002/(sici)1097-0258(19970228)16: 4<385::aid-sim380>3.0.co;2-3.
- Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993; 80(1):27–38. doi:10.1093/biomet/80.1.27.
- Li W, Cerise JE, Yang Y, Han H. Application of t-SNE to human genetic data. *J Bioinform Comput Biol*. 2017;15(4):1750017. doi:10.1142/S0219720017500172.
- Thompson LDR, Burchette R, Iganaj S, Bhattasali O. Oropharyngeal squamous cell carcinoma in 390 patients: analysis of clinical and histological criteria which significantly impact outcome. *Head Neck Pathol*. 2020;14(3): 666–688. doi:10.1007/s12105-019-01096-0.
- Facompre ND, Rajagopalan P, Sahu V, et al. Identifying predictors of HPV-related head and neck squamous cell carcinoma progression and survival through patient-derived models. *Int J Cancer*. 2020;147(11):3236–3249. doi: 10.1002/ijc.33125.
- Moore C, Flynn MB, Greenberg RA. Evaluation of size in prognosis of oral cancer. *Cancer*. 1986;58(1):158–162. doi:10.1002/1097-0142(19860701)58:1<158:: aid-cnrc2820580127>3.0.co;2-b.
- Corredor G, Wang X, Zhou Y, et al. Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non-small cell lung cancer. *Clin Cancer Res*. 2019;25(5):1526–1534.
- Abduljabbar K, Raza SEA, Rosenthal R, et al.; for the TRACERx Consortium. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med*. 2020;26(7):1054–1062. doi:10.1038/s41591-020-0900-x.
- Azarianpour S, Corredor G, Bera K, et al. Computer extracted features related to the spatial arrangement of tumor-infiltrating lymphocytes predict overall survival in epithelial ovarian cancer. In: Tomaszewski JE, Ward AD, eds. *Medical Imaging 2020: Digital Pathology*. Houston, TX: SPIE; 2020:25. doi: 10.1117/12.2550188.
- Sandulache VC, Wilde DC, Sturgis EM, Chiao EY, Sikora AG. A hidden epidemic of “intermediate risk” oropharynx cancer. *Laryngoscope Investig Otolaryngol*. 2019;4(6):617–623. doi:10.1002/lio2.316.
- Wang Z, Li F, Rufo J, et al. Acoustofluidic salivary exosome isolation. *J Mol Diagn*. 2020;22(1):50–59. doi:10.1016/j.jmoldx.2019.08.004.
- Gleber-Netto FO, Rao X, Guo T, et al. Variations in HPV function are associated with survival in squamous cell carcinoma. *JCI Insight*. 2019;4(1):e124762. doi:10.1172/jci.insight.124762.
- Lewis JS, Ali S, Luo J, Thorstad WL, Madabhushi A. A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. *Am J Surg Pathol*. 2014;38(1):128–137.
- Bernstein JM, Bernstein CR, West CML, Homer JJ. Molecular and cellular processes underlying the hallmarks of head and neck cancer. *Eur Arch Otorhinolaryngol*. 2013;270(10):2585–2593.