EBioMedicine 45 (2019) 70–80

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Research paper

Development and validation of radiomic signatures of head and neck squamous cell carcinoma molecular features and subtypes

Chao Huang [a,b,c,1], Murilo Cintra [c,d,e,1], Kevin Brennan [c], Mu Zhou [c], A. Dimitrios Colevas [f], Nancy Fischbein [d], Shankuan Zhu [a,b,*], Olivier Gevaert [c,g,**]

[a] Chronic Disease Research Institute, School of Public Health, and Women's Hospital, School of Medicine, Zhejiang University, Zhejiang, Hangzhou, China
[b] Department of Nutrition and Food Hygiene, School of Public Health, Zhejiang University, Zhejiang, Hangzhou, China
[c] Department of Medicine, Stanford Center for Biomedical Informatics Research (BMIR), USA
[d] Department of Radiology, Stanford University, USA
[e] Ribeirão Preto Medical School, University of São Paulo, Brazil
[f] Division of Oncology, Department of Medicine, Stanford University, USA
[g] Department of Biomedical Data Science, Stanford University, USA

ARTICLE INFO

Article history:
Received 19 March 2019
Received in revised form 18 June 2019
Accepted 18 June 2019
Available online 27 June 2019

Keywords:
Head and neck squamous cell carcinoma
Radiomics
DNA methylation
Genomics
Transcriptomics

ABSTRACT

Background: Radiomics-based non-invasive biomarkers are promising to facilitate the translation of therapeutically related molecular subtypes for treatment allocation of patients with head and neck squamous cell carcinoma (HNSCC).
Methods: We included 113 HNSCC patients from The Cancer Genome Atlas (TCGA-HNSCC) project. Molecular phenotypes analyzed were RNA-defined HPV status, five DNA methylation subtypes, four gene expression subtypes and five somatic gene mutations. A total of 540 quantitative image features were extracted from pretreatment CT scans. Features were selected and used in a regularized logistic regression model to build binary classifiers for each molecular subtype. Models were evaluated using the average area under the Receiver Operator Characteristic curve (AUC) of a stratified 10-fold cross-validation procedure repeated 10 times. Next, an HPV model was trained with the TCGA-HNSCC, and tested on a Stanford cohort (N = 53).
Findings: Our results show that quantitative image features are capable of distinguishing several molecular phenotypes. We obtained significant predictive performance for RNA-defined HPV+ (AUC = 0.73), DNA methylation subtypes MethylMix HPV+ (AUC = 0.79), non-CIMP-atypical (AUC = 0.77) and Stem-like-Smoking (AUC = 0.71), and mutation of NSD1 (AUC = 0.73). We externally validated the HPV prediction model (AUC = 0.76) on the Stanford cohort. When compared to clinical models, radiomic models were superior to subtypes such as NOTCH1 mutation and DNA methylation subtype non-CIMP-atypical while were inferior for DNA methylation subtype CIMP-atypical and NSD1 mutation.
Interpretation: Our study demonstrates that radiomics can potentially serve as a non-invasive tool to identify treatment-relevant subtypes of HNSCC, opening up the possibility for patient stratification, treatment allocation and inclusion in clinical trials.
Fund: Dr. Gevaert reports grants from National Institute of Dental & Craniofacial Research (NIDCR) U01 DE025188, grants from National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (NIBIB), R01 EB020527, grants from National Cancer Institute (NCI), U01 CA217851, during the conduct of the study; Dr. Huang and Dr. Zhu report grants from China Scholarship Council (Grant NO:201606320087), grants from China Medical Board Collaborating Program (Grant NO:15-216), the Cyrus Tang Foundation, and the Zhejiang University Education Foundation during the conduct of the study; Dr. Cintra reports grants from São Paulo State Foundation for Teaching and Research (FAPESP), during the conduct of the study.

* Correspondence to: S. Zhu, Chronic Disease Research Institute, Department of Nutrition and Food Hygiene, School of Public Health, and Women's Hospital, School of Medicine, Zhejiang University, 866 Yu-hang-tang Road, Zhejiang, Hangzhou 310058, China.
** Correspondence to: O. Gevaert, Stanford Center for Biomedical Informatics Research (BMIR), Department of Medicine, and Department of Biomedical Data Science, Stanford University, 1265 Welch Road, Stanford, CA 94305-5479, USA.
E-mail addresses: zsk@zju.edu.cn (S. Zhu), ogevaert@stanford.edu (O. Gevaert).
[1] co-first authors.

https://doi.org/10.1016/j.ebiom.2019.06.034

Research in context

Evidence before this study

Head and neck squamous cell carcinoma (HNSCC) is a common heterogeneous malignancy that is typically diagnosed using imaging. Existing molecular phenotypes such as HPV status are crucial in the diagnosis of HNSCC and determines treatment. In addition, large genomic studies have shown the existence of novel molecular subtypes that have the potential to be important for treatment. Quantitative analysis of radiographic images of HNSCC patients, also known as radiomic analysis, has shown that prognostic is correlated with quantitative image features. However, to date, only a few studies applied radiomic analysis to predict HPV status in HNSCC, while, to the best of our knowledge, there have not been any radiomic reports studying molecular subtypes.

Added value of the study

We present a comprehensive radiomic analysis of HNSCC patients from a multi-institutional cohort and associate them with both existing, and novel molecular phenotypes and subtypes. We show that radiomics signatures exist that mirror molecular subtypes including a DNA methylation defined HPV + HNSCC subtype. Next, we validate the DNA methylation HPV + model in an independent cohort.

Implications of all the available evidence

Our results demonstrate that radiomic analysis has potential as a non-invasive tool of molecular subtypes potentially facilitating treatment allocation of HNSCC patients. Our work opens up the possibility to identify molecular subtypes of HNSCC non-invasively opening up the possibility for patient stratification, treatment allocation and inclusion in clinical trials in a more rapid manner.

1. Introduction

Head and neck squamous cell carcinoma (HNSCC) is a common heterogeneous malignancy, accounting for >500,000 new cases every year globally and encompassing cancers arising in oral cavity, oropharynx, nasopharynx, hypopharynx, and larynx [1,2]. Conventional treatment of HNSCC includes surgery, radiotherapy, and chemotherapy, employed individually or in combination depending on TNM stage and primary site [3]. Responses to clinical treatments vary greatly among HNSCC patients and remain disappointing, especially in advanced stage disease [4]. Furthermore, chemotherapy and radiation often confer significant toxicity [3]. It is therefore crucial to identify biomarkers with which to select the best treatment strategy for each patient.

More than a decade ago, several reports showed a causal link between the rise of HNSCC cancer and human papilloma virus infection (HPV) [5–7]. Compared to HPV negative (HPV-), HPV positive (HPV+) HNSCC has been identified with better prognosis and response to therapies including immunotherapy. Current efforts are focused on developing treatment strategies specifically for this subtype. For example HPV+ oropharyngeal cancers are now being considered for de-intensifying treatment strategies [8]. Accurate classification of this subtype would enable targeted neoadjuvant treatment, and would enable more accurate diagnosis that is predictive of response to standard and emerging therapies, and of prognosis.

In parallel, large-scale genomic profiling studies, such as The Cancer Genome Atlas (TCGA) project, have been conducted and identified several distinct molecular subtypes and clinically-relevant molecular driver events [9,10], providing insight into different etiologies and sources of molecular heterogeneity in HNSCC. The gene expression subtypes proposed by TCGA group are atypical, classical, mesenchymal and basal [9]. In addition, among the most recognized common mutations by these studies are NOTCH1, TP53, CDKN2A, PIK3CA and NSD1 [9–19].

More recently, we introduced clinically relevant DNA methylation subtypes [15,17,20] using a novel bioinformatics algorithm called MethylMix [14,21]. These subtypes include an HPV subtype, a CIMP atypical subtype that is defined by infiltration of M1 macrophages [15], an NSD1 subtype that as immune cold and has been subsequently experimentally validated [17,22], a non-CIMP-atypical subtype and a stem cell subtype associated with smoking. These proposed DNA methylation subtypes have implications for treatment, but have not been translated to diagnostic assays that are amenable to clinical use.

Taken together, the molecular subtypes mentioned above are showing potential in assisting effective clinical management of HNSCCs but have not yet been translated to clinical use [11]. This is partly due to lack of clinically-applicable tools to identify the molecular properties relevant tof patient-specific selection of treatment because current molecular assays are invasive and has lower financial cost. Non-invasive biomarkers offer the possibility of facilitating such translation.

Medical imaging is usually acquired as routine practice for cancer patients and is promising to develop non-invasive biomarkers for the molecular characterization as radiographical images are believed to be representations of pathophysiology driven by underlying molecular and tissue-level changes [23]. Radiomics is an emerging field concerned with the high-throughput extraction of innumerable quantitative imaging features, complementing and accelerating the advancement to mine the medical images [23,24]. Radiomic features are showing great potential for predicting diagnosis [25–27], survival, treatment response [28–34], and molecular properties of tumors [24,30,32,35].

In head and neck cancer, computed tomography (CT), magnetic resonance (MR) and positron emission tomography (PET) imaging are routinely used [36]. Radiomic features extracted from these various imaging modalities have demonstrated to be useful during the management of HNSCC patients. For instance, Bogowicz et al. showed the good discriminative power of both CT and PET radiomics for local tumor control modeling in HNSCC [37]. In this study, we focus on radiomics analysis based on CT scans. Prediction of clinically-relevant HNSCC subtypes using pretreatment CT images provides an early source of information that can expedite and support the design of treatment strategies and clinical decision-making process, and may prepare clinicians and patients for the appropriate course of treatment. As CT images are routinely generated, imaging-based biomarkers can be rapidly translated for use in clinical practice and clinical trials. Here, we present a radiomic analysis of pretreatment CT images and multi-omics profiles, and show that we can predict molecular phenotypes from CT images. We also validate a radiomic model for HPV status in an independent cohort.

2. Materials and methods

2.1. Patients

In this study, we analyzed two HNSCC datasets (TCGA-HNSCC for model development and Stanford-HNSCC for validation). The former is part of The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma project (TCGA-HNSCC) [38]. The genomic, demographic and clinical data were acquired from the Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov/), while CT scans of matched TCGA-HNSCC patients (N = 163) were downloaded from The Cancer Imaging Archive (TCIA) (https://wiki.cancerimagingarchive.net/) in July 2017. Analysis of this dataset was conducted consistent with TCGA and TCIA IRB approvals and agreements, and did not require additional institutional review board (IRB) approval. The Stanford-HNSCC

data was obtained from Stanford Hospital with IRB approval from Stanford University for this retrospective study. Data were de-identified and informed consent from patients was waived.

For both datasets, inclusion criteria were patients with pre-treatment contrast-enhanced CT scans and with histologically confirmed HNSCC. Patients with CT images obscured by dental artifacts (N = 7), patients without a pre-treatment CT (N = 35) or without a pre-treatment CT with contrast (N = 8) were excluded. For the Stanford-HNSCC cohort, we only used cases for testing the HPV prediction model where HPV testing was done, patients without defined HPV status were excluded from this analysis.

2.2. Molecular phenotypes

We analyzed the following molecular phenotypes that were available for the TCGA-HNSCC cohort: HPV status, gene expression subtypes (transcriptomics-based) [9], DNA methylation subtypes obtained by the MethylMix algorithm [15,17,20] (methylation-based), and a set of frequently observed somatic mutations (genomics-based). For the Stanford-HNSCC cohort HPV status defined by p16 immunohistochemistry was available and used as surrogate indicator of HPV infection.

HPV status for the TCGA-HNSCC cohort was determined via VirusSeq, a computational tool measuring the presence of strain-specific HPV RNA sequences detectable in whole-transcriptome sequencing (RNA-seq) data [39,40]. We denoted it as RNA-defined HPV + in this work.

The four gene expression based subtypes were previously reported by TCGA group as atypical, classical, mesenchymal and basal [9,16]. The atypical subtype is defined as lacking either EGFR amplification or deletion of 9p. The classical subtype has canonical genomic alterations relevant to squamous cell carcinoma, e.g. focal amplification of both EGFR and CCND1, amplification of 3q, and deletion of 3p and 9p. The mesenchymal subtype is defined by expression of genes involved in the epithelial to mesenchymal transition pathway, and finally the basal subtype is characterized by expression of genes that are expressed in basal epithelial cells [9,16].

The DNA methylation subtypes are defined by consensus clustering after running MethylMix algorithm [14,15,17]. This resulted in five DNA methylation subtypes based on DNA methylation aberrations defined as non-CIMP-atypical, NSD1-Smoking, CIMP-atypical, MethylMix HPV+ and stem-like-smoking [14,15,17].

In addition, a set of five common somatic mutations were analyzed with data downloaded from GDC Data Portal (https://portal.gdc.cancer.gov/): translocation-associated Notch homolog 1 (NOTCH1), tumor protein p53 (TP53), cyclin dependent kinase inhibitor 2A (CDKN2A), phosphoinositide-3-kinase catalytic alpha polypeptide (PIK3CA), and nuclear receptor binding SET domain protein 1 (NSD1) [10,12,17].

2.3. Tumor volume segmentation

Before feature extraction, for each patient, a head and neck radiologist (M.C., >10 years of experience) manually outlined the regions of interest (ROIs, defined as the gross tumor volume), using the DICOM viewer Horos (https://horosproject.org) with the pencil tool on each axial slice. ROIs were saved with all pixels within the tumor area set to one and pixels of the background set to zero. During the process of segmentation, all artifacts and other tissues around the tumor, as well as airways, were maximally avoided. Only tumor tissues - solid or necrotic - were included in the ROIs. The ROIs were re-segmented by the same radiologist blinded to the previous segmentations three months after the first segmentation. Divergence in the two segmentations for each tumor volume were resolved by discussing with other experts.

2.4. Quantitative image features extraction

Three-dimensional tumor segmentations were first resampled to isometric voxels of size 1 by 1 by 1 mm^3. A total of 540 quantitative image features were extracted from the region of interest (ROI) of gross tumor volume. These quantitative image features are categorized into five groups of features: (1) first-order features, (2) shape and size features, (3) global histogram features, (4) textural features and (5) filter based features. Group 1 (first-order features) are calculated directly from the intensity values, including features such as minimum, maximum, mean and variance. Group 2 (shape and size features) quantify the 3D shape and size of the tumors, including volume size and surface area. Group 3 (global histogram features) are a group of statistical features based on histograms of the intensity values. Group 4 (textural features) are calculated based on the spatial relationships between voxels and are further subdivided into Gray-Level Co-Occurrence Matrix (GLCM), gray level run length matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM) and Neighboring Gray Tone Difference Matrix (NGTDM) features. Finally, group 5 (filter-based features) were obtained by applying wavelet transformations to CT images to extract filter-transformed first-order features and textural features. The filter-based features are able to capture higher-level abstraction of the imaging objects.

Description of all features and the corresponding algorithms are provided in Data Supplement. All feature implementation was done and reported according to previous work and the current Image Biomarkers Standardization Initiative (IBSI) guidelines [41–44]. The feature extraction pipeline was implemented using in-house developed pipeline in Matlab 2018a (MathWorks, Natick, MA, USA). The feature extraction pipeline was made publicly available at https://github.com/gevaertlab/radiomics_pipeline.

2.5. Model construction

We assessed the feature robustness for potential bias because of variations in imaging acquisition and delineation using the image perturbation strategy introduced by Zwanenburg et al. [45] The perturbation chain applied in this work is a combination of rotation, volume adaption, and contour randomization, which was one of the four highly recommended approaches. A total of 30 perturbation settings were investigated conforming to the proposal. Intraclass correlation coefficient (ICC) and the 95% confidence interval (CI) were computed for each radiomic feature. Features with 95% CI equal to or >0.9 were considered robust. The robustness analysis was implemented with an in-house pipeline written in Python according to Zwanenburg et al. [45] Besides, to minimize the potential noise derived from differences in imaging protocols across sites, we harmonized the radiomic features before modeling with "ComBat" [46–48]. The code was available at: https://github.com/Jfortin1/ComBatHarmonization. The ComBat model was first introduced in genomic studies to correct batch effects in microarray data [49].

After removing near-zero-variance features, we selected the most discriminative features by implementing the minimum redundancy maximum relevance (mRMR) algorithm. mRMR calculates mutual information (MI) between feature and outcome, and ranks the features by minimizing the average MI between the features and maximizing the MI between each feature and the outcome variable [50,51]. mRMR has been shown to be a stable algorithm to efficiently select a non-redundant highly relevant and complementary set of features in a low sample-to-dimensionality ratio scenario [50,52].

Multivariate logistic regression with least absolute shrinkage and selection operator (LASSO) penalty was applied to build binary classifiers based on the retained quantitative image features. LASSO is a widely applied method for the regression of high-dimensional data. One significant advantage is that LASSO shrinks the regression coefficients of the

irrelevant features to zero depending on a suitable regularization term λ, thus only a small amount of features with non-zero coefficients are retained, which makes it easier to interpret the model. Specifically, for multi-class subtypes, DNA methylation subtypes, and gene expression subtypes, it was reduced to multiple binary classification tasks – in this "one-vs-all" approach, one classifier was built to predict each subtype against all others. All quantitative image features were *Z*-score standardized.

To further prove the usefulness of the radiomic features, we compared radiomic models to clinical models based on only clinical features (i.e. age, gender, smoking status, primary tumor site and TNM staging), as well as the combination of radiomics and clinical data defined as radiomic+clinical models. To build a radiomic+clinical model, a radiomics signature for each patient was first computed by multiplying the selected features with their respective coefficients from the best radiomic model based on LASSO. The radiomics signature was then treated as an independent feature representing the radiomics alongside the clinical features. Both clinical models and radiomic+clinical models were fitted using binomial logistic regression with backward selection of the features based on Akaike's information criterion [53].

2.6. Model selection and evaluation

We trained classifiers for the four modes of molecular phenotypes within TCGA-HNSCC. To minimize the true error estimate bias, we applied the nested stratified 10 times repeated 10-fold cross-validation (CV) (nested stratified 10 × 10-CV) for model selection and assessment, where the inner 10 × 10-CV loop was used for model developing and the outer 10 × 10-CV loop was for testing. Here, "stratified" means, for each target subtype, different class levels in each fold are represented in the same proportion as that in the full dataset. The improvements attributed to the stratified cross-validation was evidenced by Witten et al. [54]. Furthermore, to alleviate data imbalance all models were learned by incorporating the weights of the classes to give lower weight to the majority class and higher weight to the minority class. The performance metric was the area under Receiver Operating Characteristic (ROC) curve (AUC). AUCs from all the testing runs were averaged and reported. The most predictive models were determined by maximizing the CV AUCs. Fig. 1 outlines the complete data analysis workflow using nested stratified 10 × 10-CV.

2.7. External validation of the HPV prediction model

One model was trained to predict HPV status using the stratified 10 × 10-CV in TCGA-HNSCC and was externally tested by Stanford-HNSCC. The ROC curve for the results was plotted. The ROC curve of the model trained on TCGA-HNSCC was further used to determine the best cut-off values based on Youden index, afterwards specificity, sensitivity and accuracy were calculated and reported. External validation was not performed for other subtypes because no data was available with both CT images and the genomics, transcriptomics and DNA methylation data to derive these subtypes labels.

Next, we used calibration analysis to visualize calibration of the models build on TCGA-HNSCC and the Stanford-HNSCC cohorts. We compared the radiomic, the clinical and the radiomic + clinical models.

In spite of lacking molecular data for external validation, the best models for somatic mutations, DNA methylation and gene expression subtypes are also derived from the TCGA-HNSCC the same as the HPV prediction model for validation by other investigators.

We also tested if any interactions were present between radiomic features and clinical features using Variance Inflation Factor (VIF) analysis [55] We calculated VIFs for the clinical features and/or radiomic signatures when combining these features into the above clinical models or radiomic + clinical models.

3. Results

3.1. Patient characteristics and quantitative image features

Based on our inclusion and exclusion criteria, this work included 113 patients from TCGA-HNSCC and 53 patients from Stanford-HNSCC cohorts (Table 1). There was no significant difference in the clinical characteristics between the two cohorts except for anatomic site of primary tumor, smoking status, clinical T and N stage. HPV positive rate is significantly higher in Stanford-HNSCC (P-value < 0.001, Chi-square test). Before building models, we analyzed the relationships between all molecular phenotypes (Fig. S1). One remarkable finding is that the MethylMix HPV+ distribution coincides strongly with the distribution of RNA-defined HPV status, which also corresponds with our results (see below). Next, 540 quantitative image features were extracted from the segmented gross tumor regions. As a result of feature robustness analysis, 491 out of 540 features met the robustness criteria and were kept for further analysis. For each binary task, we collected together the features retained in the optimal model trained from each iteration of the outer CV loop of the nested cross-validation. In total, a subset of 279 quantitative image features contributed to significant radiomic models for all tasks. Next, we counted the image feature frequencies of being selected through the outer CV loop for each feature, showing that the most important features belong to the shape and size features group (Fig. S2).

3.2. Radiomic signature of HPV

We first examined if quantitative image features could discriminate RNA-defined HPV+ and HPV- patients [39,40]. Our radiomic models demonstrated a significant ability to distinguish HPV+ from HPV- status (AUC = 0.73, Fig. 2a). To ensure the robustness of our radiomic model classification of HPV, we also tested whether it could classify the previously-reported DNA methylation MethylMix HPV+ subtype [14,15,21], characterized by a signature of abnormal methylation that is observed in both HPV positive HNSCC and cervical cancer [56]. Compared with the RNA-based measure for HPV status, the radiomic models achieved higher performance in discriminating MethylMix HPV+ from the remaining patients (AUC = 0.79, Fig. 2a). Next, we developed models using clinical data for both RNA-defined HPV+ and MethylMix HPV+ resulting in higher AUC values of 0.86 and 0.90 respectively. This can be attributed primarily to primary tumor site (pharynx vs. other sites) whereas the radiomic models are tumor site agnostic. Models combining radiomics signature with clinical features did not improve the clinical models.

Next, we investigated which feature groups are most important in the above two radiomic models, showing that for both models the dominant features are wavelet features followed by shape and size features and first order features (Fig. S3). To further interpret the radiomic models of HPV, we show two examples of patients with a tumor in the oropharynx: one HPV+ case correctly predicted and one HPV- case correctly predicted, according to both RNA-defined HPV status and DNA methylation subtype MethylMix HPV+ status (Fig. 2). The HPV+ tumor is a solid enhancing lesion at the right base of tongue (Fig. 2b). The HPV- tumor appears radiologically to be more aggressive, with signs of pre-vertebral fascia invasion and retropharyngeal fat blurring located at the posterior wall of inferior oropharynx and extending to posterior wall of hypopharynx (Fig. 2c).

3.3. Prediction of other DNA methylation subtypes

Similar to MethylMix HPV+, we evaluated predictive modeling of each of the other four DNA methylation subtypes. All these classifiers attained good performance (Fig. 3a, Fig. S4), with the Non-CIMP-Atypical model resulting in the best performance (AUC = 0.77), followed by Stem-like-Smoking (AUC = 0.71), CIMP-Atypical (AUC =

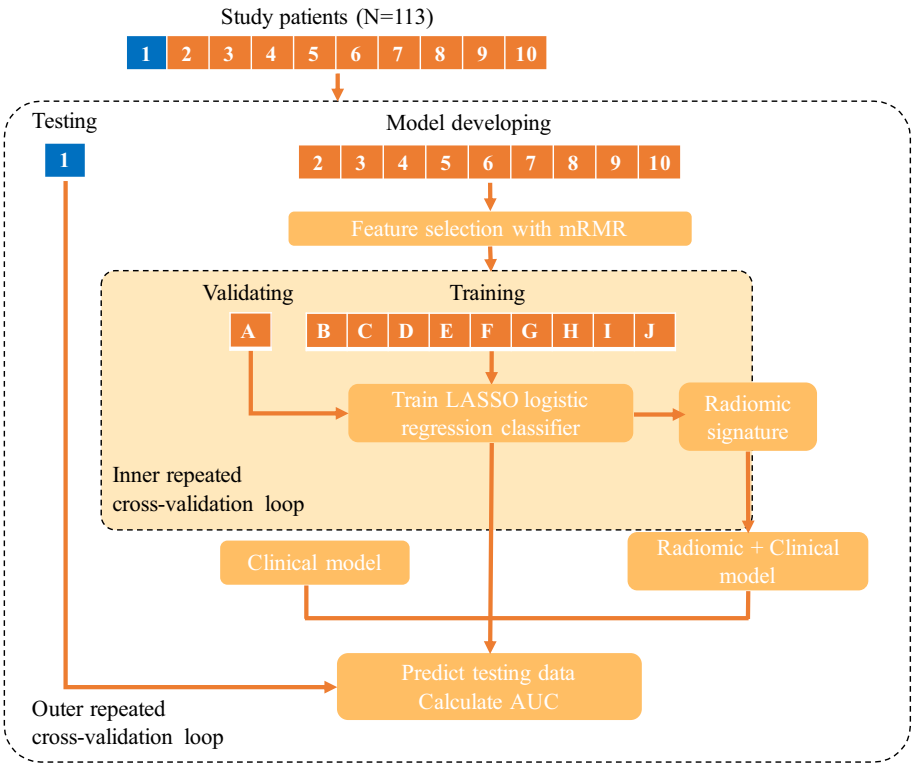

Fig. 1. Data analysis framework embedded with nested stratified repeated cross-validation. The inner loop is used to train and select out the optimal binary classifier based only on quantitative image features, while the outer loop is used to generate different resampling splits to evaluate the optimal models' generalization performance.

0.68) and NSD1-Smoking (AUC = 0.68). Next, models using only clinical data showed higher performance for CIMP-Atypical (AUC = 0.75) and NSD1-Smoking (AUC = 0.73) compared to the radiomic models, while clinical models for Non-CIMP-Atypical (AUC = 0.53) and Stem-like-Smoking (AUC = 0.66) were inferior to the corresponding radiomic models. Combining radiomic signature and clinical features did not improve the results for the DNA methylation subtypes significantly.

Next, we explored the contributing features of radiomic models for the non-CIMP-atypical and Stem-like-Smoking subtypes (Fig. S4). For both subtypes, the dominant feature group was wavelet features. For non-CIMP-atypical, the global histogram feature group was the second most important. While for Stem-like-Smoking, textural features ranked the second.

3.4. Prediction of gene expression subtypes

Next, we investigated the prediction of the previously defined gene expression subtypes: atypical, classical, mesenchymal and basal (Fig. 3b). Radiomic models were capable of predicting the subtypes atypical (AUC = 0.65), classical (AUC = 0.69), mesenchymal (AUC = 0.67) and basal (AUC = 0.67). Compared to the radiomic models, the clinical models were better for atypical (AUC = 0.80) and basal (AUC = 0.78), while the ones for classical (AUC = 0.65) and mesenchymal (AUC = 0.65) were not. Combining radiomic features and clinical features did not show better results than both radiomic models and clinical models for gene expression subtypes.

3.5. Prediction of somatic gene mutations

Next, we investigated if radiomic models are able to predict somatic mutations in the most commonly mutated genes in HNSCC: NOTCH1, TP53, CDKN2A, PIK3CA, and NSD1. The most significant performance was obtained for NSD1 (AUC = 0.73). (Fig. 3c, Fig. S5). For NSD1, the top contributing radiomic features were dominated by wavelet features, followed by the global histogram features, textural features, and shape and size features. Weaker performances were observed for predicting NOTCH1 (AUC = 0.66), TP53 (AUC = 0.65), CDKN2A (AUC = 0.66) and PIK3CA (AUC = 0.69).

Compared to radiomic models, the clinical model for NOTCH1 (AUC = 0.58) was inferior while that for TP53 (AUC = 0.76) increased. The clinical models for other three mutations were comparable to the radiomic models. Models integrating radiomic signature and clinical features did not show great improvement except for NOTCH1 (AUC = 0.69).

3.6. External validation of the HPV radiomic signature

Next, we acquired an independent cohort (Stanford-HNSCC). Therefore, external validation was further conducted by first training models on the TCGA-HNSCC cohort, and then testing the models on the Stanford-HNSCC cohort. The established models were as below (all output is logits):

(1) Radiomic model (the same model to extract the radiomic signature): y = −0.48–0.39*firstOrder_InterquartileRange + 0.35*waveletHHH_firstOrder_Kurtosis − 0.68*waveletLLL_firstOrder_Maximum + 0.25*shapeSize_Solidity − 0.43*firstOrder_QuartileCoefficientDispersion.
(2) Clinical model: y = −5.23 + 3.32*Gender (Female:0, Male:1) + 0.82*Anatomic_site1(Larynx:0, Oral cavity:1) + 5.08*Anatomic_site2(Larynx:0, Pharynx:1).
(3) Radiomic + Clinical model: y = −4.41 + 0.61*radiomic_signature + 2.65*Gender(Female:0, Male:1) + 0.81*Anatomic_site1(Larynx:0, Oral cavity:1) + 4.50*Anatomic_site2(Larynx:0, Pharynx:1).

Although we could not validate them due to lack of molecular data for the validation cohort, the best models for somatic mutations, DNA methylation and gene expression subtypes are also provided in Supplementary Table 1 for validation by other investigators.

Table 1
Basic patient characteristics.

		TCGA-HNSCC (N = 113)	Stanford-HNSCC (N = 53)	P-value
Clinical characteristics	Age, mean ± SD, years	60.1 ± 11.1	63.3 ± 10.3	0.095
	Sex (%)			0.802
	Female	27 (23.9)	11 (20.8)	
	Male	86 (76.1)	42 (79.2)	
	Anatomic site (%)			<0.001
	Larynx	29 (25.7)	4 (7.5)	
	Oral-cavity	64 (56.6)	12 (22.6)	
	Pharynx	20 (17.7)	37 (69.8)	
	Smoking (%)			<0.001
	Non-smoker	40 (35.4)	41 (77.4)	
	Smoker	73 (64.6)	12 (22.6)	
	Clinical T stage (%)			0.002
	T1	7 (6.2)	7 (13.2)	
	T2	23 (20.4)	21 (39.6)	
	T3	38 (33.6)	7 (13.2)	
	T4	45 (39.8)	18 (34.0)	
	Clinical N stage (%)			0.002
	N0	49/112 (43.8)	11 (20.8)	
	N1	17/112 (15.2)	6 (11.3)	
	N2	41/112 (36.6)	31 (58.5)	
	N3	5/112 (4.5)	5 (9.4)	
	Unknown	1 (0.9)	0	
	Clinical M stage (%)			0.169
	M0	112/113 (99.1)	48 (90.6)	
	M1	1/113 (0.9)	3 (5.7)	
	Unknown	0	2 (3.8)	
Molecular subtypes	HPV infection (%)			<0.001
		RNA-defined	p16-defined	
	−	92 (81.4)	14 (26.4)	
	+	21 (18.6)	39 (73.6)	
	DNA methylation subtypes (%)			NA
	Non-CIMP-Atypical	26 (23.0)	NA	
	NSD1-Smoking	21 (18.6)		
	CIMP-Atypical	27 (23.9)		
	HPV+	17 (15.0)		
	Stem-like-Smoking	22 (19.5)		
	Gene expression subtypes (%)		NA	NA
	Atypical	28/83 (33.7)		
	Classical	28/83 (33.7)		
	Mesenchymal	10/83 (12.0)		
	Basal	17/83 (20.5)		
	Unknown	30 (26.5)		
	NOTCH1 (%)		NA	NA
	−	88/109 (80.7)		
	+	21/109 (19.3)		
	Unknown	4 (3.5)		
	TP53 (%)		NA	NA
	−	34/109 (31.2)		
	+	75/109 (68.8)		
	Unknown	4 (3.5)		
	CDKN2A (%)		NA	NA
	−	83/109 (76.1)		
	+	26/109 (23.9)		
	Unknown	4 (3.5)		
	PIK3CA (%)		NA	NA
	−	89/109 (81.7)		
	+	20/109 (18.3)		
	Unknown	4 (3.5)		
	NSD1 (%)		NA	NA
	−	94/109 (86.2)		
	+	15/109 (13.8)		
	Unknown	4 (3.5)		

Note: To compare the differences in clinical characteristics between the two datasets, two-sample *t* test was used for age, while Chi-square or Fisher exact tests, as appropriate, were applied for categorical variables.
Definitions: Smoking: Non-smoker = never-smoker or former-smoker quitted >15 years before diagnosis; Smoker = current-smoker or former-smoker quitted <15 years before Diagnosis.

We also tested if any interactions were present between radiomic features and clinical features using VIF analysis. However, all VIFs were below 2.0 for clinical features and/or radiomic signatures which were combined to train the clinical models or radiomic+clinical models, suggesting absence of severe multi-collinearity between these features (Supplementary Table 2).

The external validation of the radiomic model was good (AUC = 0.76, Figs. 4, 5a). The cut-offs of probability were obtained based on Youden's index and were 0.42, 0.38, 0.47 for the radiomic model, clinical model, and radiomic+clinical models, respectively. The specificity, sensitivity and accuracy were 0.74, 0.79 and 0.75, respectively. The clinical model was also verified (Fig. 4, AUC = 0.86) with specificity, sensitivity

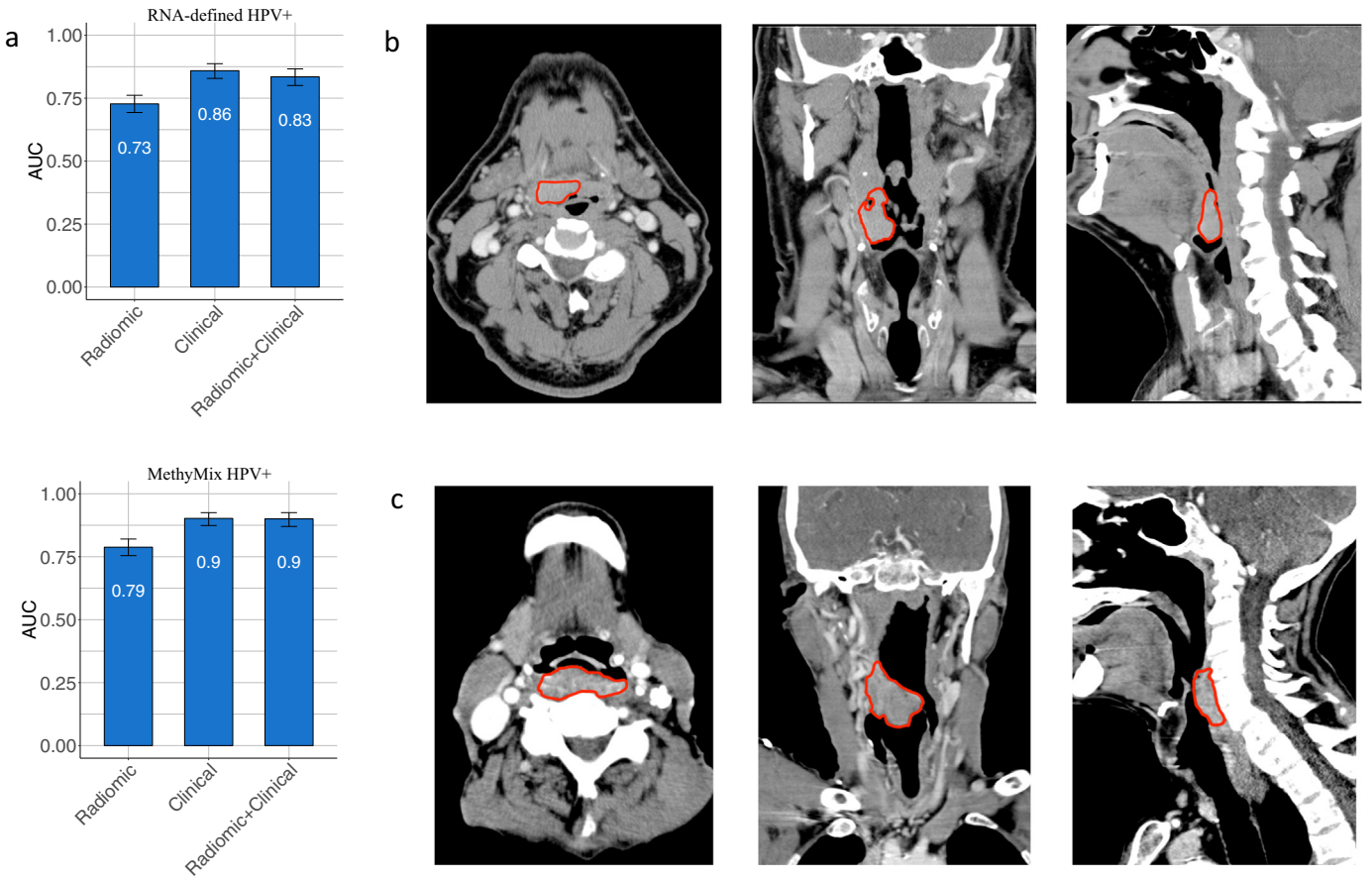

Fig. 2. Prediction of HPV infection. (a) Prediction of HPV infection status and the DNA methylation HPV subtype. Each bar plot depicts the average AUC and the error bars represent the 95% Confidence Interval (CI) with 1000 bootstrapped resamples of the AUCs from the outer cross-validation loop. (b) Selected CT image of patient TCGA-CR-5250 who had an HPV positive tumor located in oropharynx. (images order: axial, coronal, sagittal). (c) Selected CT image of patient TCGA-BA-A4IF who had an HPV negative tumor located in oropharynx. (images order: axial, coronal, sagittal).

and accuracy as 0.87, 0.79 and 0.85. When combining radiomic signature with clinical features, the performance significantly increased compared to the radiomic model (Figs. 4 and 5a, AUC = 0.88) (*P*-value = 0.032, Delong's test). The respective specificity, sensitivity and accuracy were 0.82, 0.79 and 0.81.

We also studied model calibration on the Stanford-HNSCC cohort, and compared with the TCGA-HNSCC cohort. This analysis showed that model calibration for the Stanford-HNSCC radiomic model was poor compared to the TCGA-HNSCC (Fig. S6). In addition, for both cohorts the radiomic+clinical model showed the best calibration.

Consistent with previous studies, we found that the primary site is a strong indicator of HPV infection in HNSCC. To investigate whether the radiomic model is agnostic to primary site information, we further tested the models with a subset of Stanford-HNSCC patients whose primary tumor sites were pharynx. This showed that the radiomic model still has comparable performance (AUC = 0.79) while the performance of the clinical model (AUC = 0.61) deteriorates significantly, furthermore combining the radiomic signature and clinical features (AUC = 0.77) retained comparable performance regarding to the radiomic model (Fig. 5b).

4. Discussion

In this study, we extracted quantitative image features from HNSCC and present a comprehensive framework to develop and evaluate machine-learning classifiers of these features to identify distinct molecular phenotypes derived from genomics, transcriptomics and DNA methylation data. Our results show that molecular phenotypes of HNSCC tumors, especially HPV infection status, can be predicted from quantitative features extracted from CT images. This radiomic signature of HPV can be used to develop non-invasive tools for diagnosing HNSCC patients.

HPV-positive cases are increasingly prevalent, usually have better survival and benefit from specifically tailored treatment protocols [4,13,57]. Therefore, HPV testing is important for the clinical management for HNSCC patients. Currently, there's no standard clinical marker or screening test for HPV-associated HNSCC. The most commonly used screening tests are in situ hybridization to detect HPV DNA and immunochemical p16 staining based on tumor biopsy [58]. P16 immunostaining, the surrogate biomarker that is most widely used to determine HPV status in clinic, is not sufficiently specific to ensure safe de-escalation of treatment in clinical practice, because false positive results could lead to undertreatment of aggressive HPV negative HNSCC [59]. Moreover, P16 overexpression fails to detect HPV in subanatomical regions apart from the oropharynx [60]. Our results suggest that a combined clinical +radiomic model provides an alternative strategy to determine HPV status that is convenient, non-invasive and has lower financial cost.

The radiomic model to predict HPV was observed to be significantly better than other molecular subtypes and somatic mutations. Moreover, the HPV radiomic signature was validated in an independent Stanford-HNSCC cohort, even though the HPV detection method between both cohorts are different, p16 immunohistochemistry in the Stanford-HNSCC cohorts vs. RNA-seq in the TCGA-HNSCC cohort. In addition, the two cohorts have very different prevalence of HPV+ cases (18.6% in TCGA vs. 73.6% in Stanford, Table 1). This suggests our radiomic model for HPV prediction is robust to the method used to determine HPV infection and the prevalence of HPV+ in the HNSCC population. When compared with the clinical model, the radiomic model was not

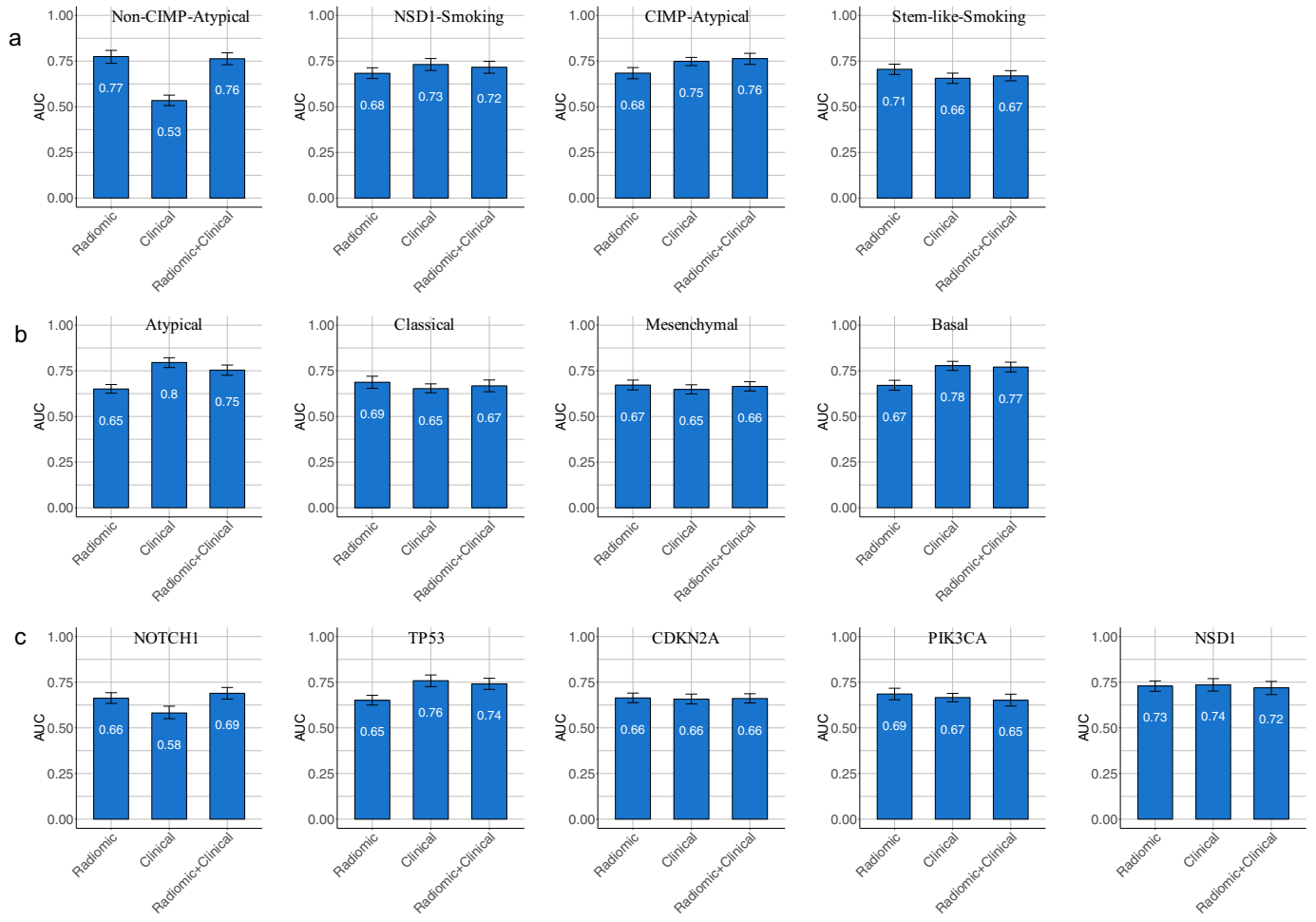

Fig. 3. Prediction of non-HPV related DNA methylation subtypes (a), gene expression subtypes (b), and five somatic gene mutations (c). Each bar plot depicts the average AUC and the error bars represent the 95% Confidence Interval (CI) with 1000 bootstrapped resamples of the AUCs from the outer cross-validation loop.

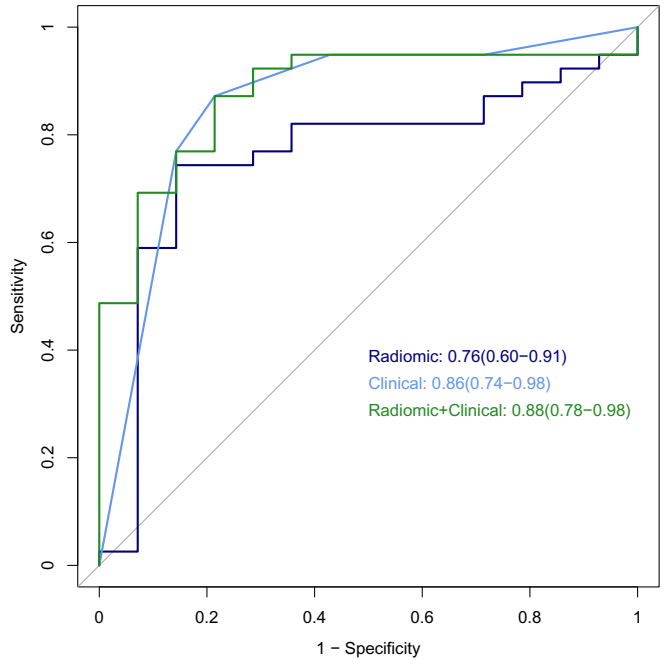

Fig. 4. ROC curves for validating models on the independent Stanford cohort for the Radiomic, Clinical and Radiomic+Clinical models.

as good in the whole HNSCC population likely due to confounding with tumor location, however in a subset analysis of tumors in the pharynx, radiomic features improved the performance. This suggests that radiomic models for HPV prediction are agnostic to primary site information, and as such might capture the molecular heterogeneity.

We also observed that the radiomic signature for the DNA methylation subtype MethylMix HPV+ has slightly higher performance compared to RNA-defined HPV status. Note that all MethylMix HPV+ cases are also RNA-defined HPV+ but only 81% of RNA-defined HPV+ patients are MethylMix HPV+ (Fig. S1), suggesting that RNA-defined HPV+ cases are not capturing all HPV+ cases and are therefore less accurately predicted by quantitative image features.

A few previous studies have reported the capability of quantitative CT imaging features to identify HPV infection among HNSCCs [61–63]. Our work differs from these previous studies as follows. Firstly, instead of randomly splitting a dataset into training and test set, we applied nested stratified cross-validation framework to train and evaluate the models, resulting in more robust estimates of the performance. We also tested the radiomic model for HPV status in an independent dataset. Secondly, we also compared the radiomic model to a model built on clinical features, suggesting that the radiomic model is superior when tumor location is removed by focusing only on pharynx tumors, showing that a clinical-only model is suboptimal. Next, Zhu et al. used the same source dataset as us [62], however they report inferior results. One possible explanation is that we also extracted wavelet features which we show are crucial for prediction (Figs. S3-S6). Another reason could be that the nested 10 times 10 stratified cross-validation is more

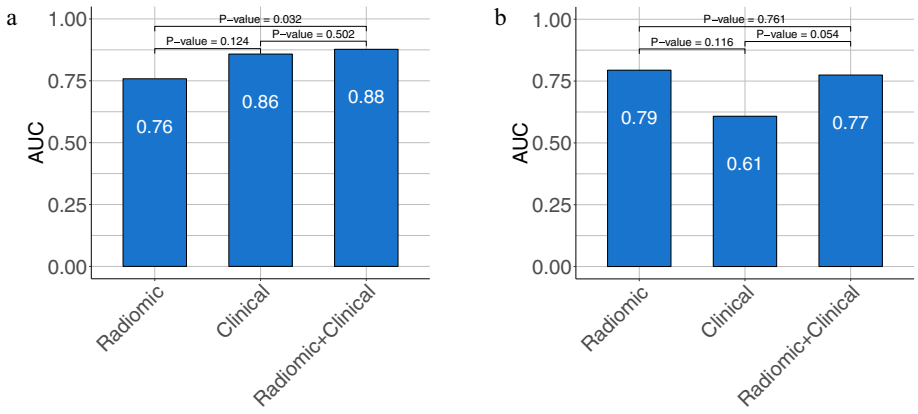

Fig. 5. External validation on an independent Stanford cohort (a) and on a subset of Stanford cohort patients whose primary tumor site were located in the pharynx (b). The *P*-values derived from Delong's test between each two AUCs shown in each pane.

effective for this sample size scenario. Lastly, in our training datasets, HPV status was determined with a more accurate method, RNA-seq, instead of immunohistochemical p16 staining which is sensitive but not specific enough [58]. This is backed up by the evidence that our HPV radiomic model trained using the TCGA-HNSCC cohort, where RNA-seq was applied to determine HPV infections, could still be validated in an independent Stanford-HNSCC cohort, even though the HPV detection method was p16. There are also a few studies working on HPV prediction in HNSCC with imaging modalities other than CT. For example, Vallières et al. identified a set of FDG-PET features that can determine HPV status in HNSCC [64]. Next, Nakahira et al. showed that apparent diffusion coefficients from MR could be used to predict HPV in patients with oropharyngeal squamous cell carcinoma [65].

The gene expression atypical subtype is another subtype related to HPV infection. In this work, radiomic models for atypical showed an acceptable performance. The subtype atypical contains all MethylMix HPV + cases and 86% of the RNA-defined HPV+ cases, and is thus heavily linked with HPV infection. But, atypical contains approximately 50% of samples that are not HPV+ by any definition (MethylMix HPV+ or RNA-defined HPV+) suggesting that this transcriptomic subgroup has its own radiomic signature characterizing this subtype. Besides enrichment of HPV+, atypical cases are characterized by lack of chromosome 7 amplifications and mutations in the helical domain of PIK3CA.

As for non-HPV-related subtypes, our results show that quantitative image features have moderate ability to identify the DNA methylation subtypes including the non-CIMP-Atypical, and the mutation NSD1. The CIMP, 'CpG island methylator phenotype', is a driver for cancer developments related to epigenetic silence of tumor suppressor genes [15]. The non-CIMP-atypical subtype was mostly found with CASP8 and NOTCH1 mutations. It is interesting that radiomic models for DNA methylation subtypes non-CIMP-atypical and Stem-like-Smoking, gene expression subtype classical, and mutation NOTCH1 performed better than clinical models, suggesting the CT imaging features are more representative of the tumors and the inter-tumor and intra-tumor characteristics can be much better captured by the imaging than semantic clinical features for these subtypes.

Next, we explored the top contributing quantitative image features from the radiomic signatures for each of the molecular phenotypes. Across all prediction tasks, shape and size features, global histogram features, and wavelet filter-based features are mostly selected feature groups in the radiomic signatures (Fig. S2). However, the most predictive and robust features varied across different molecular phenotypes, and could source radiomics features from any feature group. Taken together, the extracted quantitative image features could capture significant information to individualize various molecular phenotypes, thus helping to explain how the predictive models could work.

We admit that there are several limitations residing in this study. Firstly, there could be potential bias in quantitative image features because of the limited number of experts involved in tumor volume segmentation or variability in CT image acquisition protocols for both the two datasets included in this work. Previously Zhao et al. reported that quantitative image features were reproducible over a broad range of CT scan settings [66]. To minimize such noise, we only included contrast-enhanced images resampled to isovolumetric voxels, and we applied *Z*-score standardization to the quantitative image features. Besides, we tested the features robustness with image perturbation approach proposed by Zwanenburg et al. [45]. The resulted robust features were further used for the analysis. Features harmonization was also applied to harmonize the radiomic features of images collected from different sites. Secondly, we externally validated the prediction model for HPV infection to evaluate if a model built on the TCGA cohort is able to predict the same phenotype in a Stanford cohort, however, such external validations were not performed for other molecular subtypes for the reason of unavailability of the tissue. In addition, we did a calibration analysis for both the TCGA-HNSCC and Stanford-HNSCC analyses, showing that the Stanford-HNSCC radiomic model was not well calibrated compared to the TCGA-HNSCC radiomic model (Fig. S6). In addition, the combined clinical+radiomic model was better calibrated than models only using clinical or radiomics features, suggesting that integrating clinical and radiomic data does not lead to the best models. Nevertheless, our results show good validation in terms of predictive performance of the HPV radiomic signature, suggesting that also other predictive models of gene expression subtypes, DNA methylation subtypes, and somatic mutations have the potential to generalize to new cohorts.

In conclusion, our findings highlight the feasibility of non-invasive molecular phenotyping systems to subtype HNSCCs for both established biomarkers such as HPV+, and proposed transcriptomic and DNA methylation subtypes beyond HPV status. With the movement towards stratified treatment strategies for HPV+ versus HPV- HNSCC, our radiomic model could potentially be used to predict HPV status at the time of diagnosis, allowing more informed planning of surgery and treatment. More generally, this radiomic analysis opens up non-invasive assignment of HPV+ patients prior to surgery for inclusion in clinical trials, for de-intensification of treatment, for neoadjuvant therapies and for checkpoint immunotherapies that show greater response rates in HPV+ cases. However, it is noteworthy that more and larger cohorts are warranted to verify and further enhance the findings of our study before translating the models as diagnostic tools in clinical practice.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.06.034.

Funding sources

Dr. Gevaert reports grants from National Institute of Dental & Craniofacial Research (NIDCR) U01 DE025188, grants from National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (NIBIB), R01 EB020527, grants from National Cancer Institute (NCI), U01 CA217851, during the conduct of the study; Dr. Huang and Dr. Zhu report grants from China Scholarship Council (Grant NO: 201606320087), grants from China Medical Board Collaborating Program (Grant NO: 15-216), the Cyrus Tang Foundation, and the Zhejiang University Education Foundation during the conduct of the study; Dr. Cintra reports grants from São Paulo State Foundation for Teaching and Research (FAPESP), during the conduct of the study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of Competing Interests

The authors declare no potential conflicts of interest.

Author contributions

Study concept & design: Olivier Gevaert; data acquisition or data analysis/interpretation, Chao Huang, Murilo Cintra, Shankuan Zhu, Olivier Gevaert; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, Chao Huang, Murilo Bicudo Cintra; manuscript editing, all authors.

Acknowledgments

The results published here are, in part, based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. We thank Vonn Walter and Neil Hayes from the TCGA Network for sharing the four gene expression subtypes for the TCGA-HNSCC cohort.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7–30. https://doi.org/10.3322/caac.21442.
[2] Belcher R, Hayes K, Fedewa S, et al. Current treatment of head and neck squamous cell cancer. J Surg Oncol 2014;110:551–74 2014/07/24 https://doi.org/10.1002/jso.23724.
[3] Marur S, Forastiere AA. Head and neck squamous cell carcinoma: update on epidemiology, diagnosis, and treatment. Mayo Clin Proc 2016;91:386–96 2016/03/06 https://doi.org/10.1016/j.mayocp.2015.12.017.
[4] 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. https://doi.org/10.1056/NEJMoa0912217.
[5] Gillison ML, Koch WM, Capone RB, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst 2000;92:709–20 [2000/05/04].
[6] Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol 2011;29:4294–301 2011/10/05 https://doi.org/10.1200/JCO.2011.36.4596.
[7] D'Souza G, Kreimer AR, Viscidi R, et al. Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med 2007;356:1944–56 2007/05/15 https://doi.org/10.1056/NEJMoa065497.
[8] Mirghani H, Blanchard P. Treatment de-escalation for HPV-driven oropharyngeal cancer: where do we stand? Clin Transl Radiat Oncol 2018;8:4–11. https://doi.org/10.1016/j.ctro.2017.10.005.
[9] Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517:576–82. https://doi.org/10.1038/nature14129.
[10] Stransky N. The mutational landscape of head and neck squamous cell. Science 2011;1208130:333.
[11] Hammerman PS, Hayes DN, Grandis JR. Therapeutic insights from genomic studies of head and neck squamous cell carcinomas. Cancer Discov 2015;5:239–44. https://doi.org/10.1158/2159-8290.CD-14-1205.
[12] Cho J, Johnson DE, Grandis JR. Therapeutic implications of the genetic landscape of head and neck cancer. Semin Radiat Oncol 2018;28:2–11. https://doi.org/10.1016/j.semradonc.2017.08.005.
[13] Fakhry C, Westra WH, Li S, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 2008;100:261–9. https://doi.org/10.1093/jnci/djn011.
[14] Gevaert O. MethylMix: an R package for identifying DNA methylation-driven genes. Bioinformatics 2015;31:1839–41. https://doi.org/10.1093/bioinformatics/btv020.
[15] Brennan K, Koenig JL, Gentles AJ, et al. Identification of an atypical etiological head and neck squamous carcinoma subtype featuring the CpG island methylator phenotype. EBioMedicine 2017;17:223–36. https://doi.org/10.1016/j.ebiom.2017.02.025.
[16] Walter V, Yin X, Wilkerson MD, et al. Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. PLoS One 2013;8:e56823. https://doi.org/10.1371/journal.pone.0056823.
[17] Brennan K, Shin JH, Tay JK, et al. NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. Sci Rep 2017;7:17064. https://doi.org/10.1038/s41598-017-17298-x.
[18] Champion M, Brennan K, Croonenborghs T, et al. Module analysis captures pancancer genetically and epigenetically deregulated cancer driver genes for smoking and antiviral response. EBioMedicine 2018;27:156–66 2018/01/15 https://doi.org/10.1016/j.ebiom.2017.11.028.
[19] Gevaert O, Villalobos V, Sikic BI, et al. Identification of ovarian cancer driver genes by using module network integration of multi-omics data. Interface Focus 2013;3:20130013. https://doi.org/10.1098/rsfs.2013.0013.
[20] Gevaert O, Tibshirani R, Plevritis SK. Pancancer analysis of DNA methylation-driven genes using MethylMix. Genome Biol 2015;16:17. https://doi.org/10.1186/s13059-014-0579-8.
[21] Cedoz PL, Prunello M, Brennan K, et al. MethylMix 2.0: an R package for identifying DNA methylation genes. Bioinformatics 2018. https://doi.org/10.1093/bioinformatics/bty156 2018/04/19.
[22] Papillon-Cavanagh S, Lu C, Gayden T, et al. Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. Nat Genet 2017;49:180–5. https://doi.org/10.1038/ng.3757.
[23] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–77 10.1148/radiol.2015151169.
[24] Zhou M, Scott J, Chaudhury B, et al. Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. AJNR Am J Neuroradiol 2018;39:208–16 2017/10/07 https://doi.org/10.3174/ajnr.A5391.
[25] Iv M, Zhou M, Shpanskaya K, et al. MR imaging-based radiomic signatures of distinct molecular subgroups of medulloblastoma. AJNR Am J Neuroradiol 2019;40. https://doi.org/10.3174/ajnr.A5899 154-161. 2018/12/14.
[26] Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. Eur Respir J 2019;53. https://doi.org/10.1183/13993003.00986-2018 2019/01/13.
[27] Han Y, Xie Z, Zang Y, et al. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an MRI-based radiomics signature in lower-grade gliomas. J Neurooncol 2018;140. https://doi.org/10.1007/s11060-018-2953-y 297-306. 2018/08/12.
[28] Gevaert O, Xu J, Hoang CD, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. Radiology 2012;264:387–96 2012/06/23. DOI: radiol.12111607 [pii] https://doi.org/10.1148/radiol.12111607.
[29] Nair VS, Gevaert O, Davidzon G, et al. Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. Cancer Res 2012;72:3725–34 2012/06/20. DOI: 0008-5472.CAN-11-3943 [pii] https://doi.org/10.1158/0008-5472.CAN-11-3943.
[30] Gevaert O, Mitchell LA, Achrol AS, et al. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology 2014:131731. https://doi.org/10.1148/radiol.14131731.
[31] Zhou M, Leung A, Echegaray S, et al. Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications. Radiology 2018;286:307–15 2017/07/21 https://doi.org/10.1148/radiol.2017161845.
[32] Itakura H, Achrol AS, Mitchell LA, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. Sci Transl Med 2015;7:303ra138. https://doi.org/10.1126/scitranslmed.aaa7582.
[33] Joye I, Debucquoy A, Deroose CM, et al. Quantitative imaging outperforms molecular markers when predicting response to chemoradiotherapy for rectal cancer. Radiother Oncol 2017. https://doi.org/10.1016/j.radonc.2017.06.013.
[34] Bulens P, Couwenberg A, Haustermans K, et al. Development and validation of an MRI-based model to predict response to chemoradiotherapy for rectal cancer. Radiother Oncol 2018. https://doi.org/10.1016/j.radonc.2018.01.008 2018/02/06.
[35] Nicolasjilwan M, Hu Y, Yan C, et al. Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. J Neuroradiol 2015;42:212–21. https://doi.org/10.1016/j.neurad.2014.02.006.
[36] Wong AJ, Kanwar A, Mohamed AS, et al. Radiomics in head and neck cancer: from exploration to application. Transl Cancer Res 2016;5:371–82. https://doi.org/10.21037/tcr.2016.07.18.
[37] Bogowicz M, Riesterer O, Stark LS, et al. Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. Acta Oncol 2017;56:1531–6. https://doi.org/10.1080/0284186X.2017.1346382.
[38] Zuley ML, Jarosz R, Kirk S, Lee Y, Colen R, Garcia K, et al. Radiology data from the cancer genome atlas head-neck squamous cell carcinoma [TCGA-HNSC] collection. Cancer Imaging Archive 2016. https://doi.org/10.7937/K9/TCIA.2016.LXKQ47MS.
[39] Chakravarthy A, Henderson S, Thirdborough SM, et al. Human papillomavirus drives tumor development throughout the head and neck: improved prognosis is associated with an immune response largely restricted to the oropharynx. J Clin Oncol 2016;34:4132–41. https://doi.org/10.1200/JCO.2016.68.2955.
[40] Chen Y, Yao H, Thompson EJ, et al. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. Bioinformatics 2013;29:266–7 2012/11/20 https://doi.org/10.1093/bioinformatics/bts665.

[41] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006. https://doi.org/10.1038/ncomms5006.
[42] Vallieres M, Kay-Rivest E, Perrin LJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep 2017;7:10117. https://doi.org/10.1038/s41598-017-10371-5.
[43] Vallieres M, Zwanenburg A, Badic B, et al. Responsible radiomics research for faster clinical translation. J Nucl Med 2018;59:189–93. https://doi.org/10.2967/jnumed.117.200501.
[44] Zwanenburg A, Leger S, Vallières M, et al. Image biomarker standardisation initiative arXiv preprint arXiv ; 2016:161207003.
[45] Zwanenburg A, Leger S, Agolli L, et al. Assessing robustness of radiomic features by image perturbation. Sci Rep 2019;9:614.
[46] Fortin J-P, Parker D, Tunç B, et al. Harmonization of multi-site diffusion tensor imaging data. Neuroimage 2017;161:149–70.
[47] Fortin J-P, Cullen N, Sheline YI, et al. Harmonization of cortical thickness measurements across scanners and sites. NeuroImage 2018;167:104–20.
[48] Lucia F, Visvikis D, Vallières M, et al. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging 2018:1–14.
[49] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118–27.
[50] Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol 2005;3:185–205.
[51] De Jay N, Papillon-Cavanagh S, Olsen C, et al. mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics 2013;29:2365–8. https://doi.org/10.1093/bioinformatics/btt383.
[52] Parmar C, Grossmann P, Bussink J, et al. Machine learning methods for quantitative radiomic biomarkers. Sci Rep 2015;5:13087. https://doi.org/10.1038/srep13087.
[53] Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC Med 2015;13:1.
[54] Witten IH, Frank E, Hall MA, et al. Data mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2016.
[55] Fox J, Monette G. Generalized collinearity diagnostics. J Am Stat Assoc 1992;87:178–83.
[56] Campbell JD, Yau C, Bowlby R, et al. Genomic, pathway network, and immunologic features distinguishing squamous carcinomas. Cell Rep 2018;23:194–212 e196 https://doi.org/10.1016/j.celrep.2018.03.063.
[57] Chera BS, Amdur RJ, Tepper J, et al. Phase 2 trial of de-intensified chemoradiation therapy for favorable-risk human papillomavirus–associated oropharyngeal squamous cell carcinoma. Int J Radiat Oncol Biol Phys 2015;93:976–85.
[58] Ambulos Jr NP, Schumaker LM, Mathias TJ, et al. Next-generation sequencing-based HPV genotyping assay validated in formalin-fixed, paraffin-embedded oropharyngeal and cervical cancer specimens. J Biomol Tech 2016;27:46–52. https://doi.org/10.7171/jbt.16-2702-004.
[59] Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. Nat Rev Cancer 2018;18:269–82 2018/03/03 https://doi.org/10.1038/nrc.2018.11.
[60] Maschio F, Lejuste P, Ilankovan V. Evolution in the management of oropharyngeal squamous cell carcinoma: systematic review of outcomes over the last 25 years. Br J Oral Maxillofac Surg 2019. https://doi.org/10.1016/j.bjoms.2018.12.006 2019/01/23.
[61] Bogowicz M, Riesterer O, Ikenberg K, et al. Computed tomography radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma. Int J Radiat Oncol Biol Phys 2017;99:921–8. https://doi.org/10.1016/j.ijrobp.2017.06.002.
[62] Zhu Y, Mohamed ASR, Lai SY, Yang S, Kanwar A, Wei L, Kamal M, Sengupta S, Elhalawani H, Skinner H, Mackin DS, Shiao J, Messer J, Wong A, Ding Y, Zhang L, Court L, Ji Y, Fuller CD. Imaging-Genomic Study of Head and Neck Squamous Cell Carcinoma: Associations Between Radiomic Phenotypes and Genomic Mechanisms via Integration of The Cancer Genome Atlas and The Cancer Imaging Archive. JCO Clin Cancer Inform 2019 Feb;3:1–9. https://doi.org/10.1200/CCI.18.00073.
[63] Leijenaar RT, Bogowicz M, Jochems A, et al. Development and validation of a radiomic signature to predict HPV (p16) status from standard CT imaging: a multicenter study. Br J Radiol 2018;91:20170498. https://doi.org/10.1259/bjr.20170498.
[64] Vallieres M, Kumar A, Sultanem K, et al. FDG-PET image-derived features can determine HPV status in head-and-neck cancer. Int J Radiat Oncol Biol Phys 2013;87:S467.
[65] Nakahira M, Saito N, Yamaguchi H, et al. Use of quantitative diffusion-weighted magnetic resonance imaging to predict human papilloma virus status in patients with oropharyngeal squamous cell carcinoma. Eur Arch Otorhinolaryngol 2014;271:1219–25.
[66] Zhao B, Tan Y, Tsai W-Y, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep 2016;6:23428.