ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL ARTICLE

A multiomic framework for predicting laryngo-esophageal dysfunction following induction chemotherapy in hypopharyngeal-laryngeal carcinoma

D. Mattavelli[1,2†], M. Compagnoni[3†], S. Calza[4†], M. Ravanelli[2,5], M. Plana[6], G. Wichmann[7], A. Paderno[8,9], Y. Zhan[3], D. Smussi[2,10], P. Rondi[2,5], A. Pinacoli[1,2], A. Medda[3], L. Lorini[11], K. Khelik[12], E. Hovig[12,13], F. A. Maffini[14], M. Kalogerini[15], A. Alberti[10], D. Farina[2,5], A. Dietz[7], R. Mesia[16], C. Piazza[1,2], S. Chiocca[3‡] & P. Bossi[8,11‡*], on behalf of the PRESERVE Consortium

[1]Unit of Otorhinolaryngology—Head and Neck Surgery, ASST Spedali Civili di Brescia, Brescia; [2]Department of Medical and Surgical Specialties, Radiological Sciences, and Public Health, University of Brescia, Brescia; [3]Department of Experimental Oncology, IEO, European Institute of Oncology, IRCCS, Milan; [4]Unit of Biostatistics and Bioinformatics, Department of Molecular and Translational Medicine, University of Brescia, Brescia; [5]Unit of Radiology, ASST Spedali Civili di Brescia, Brescia, Italy; [6]Medical Oncology Department, Institut Català d'Oncologia (ICO-Hospitalet), IDIBELL, Hospitalet de Llobregat, Barcelona, Spain; [7]Clinic of Otolaryngology, Head and Neck Surgery, Department of Head Medicine and Oral Health, University Hospital Leipzig, Leipzig, Germany; [8]Department of Biomedical Sciences, Humanitas University, Milan; [9]Otorhinolaryngology Head & Neck Surgery Unit, IRCCS Humanitas Research Hospital, Rozzano, Milan; [10]Unit of Medical Oncology, ASST Spedali Civili di Brescia, Brescia; [11]Medical Oncology and Hematology Unit, IRCCS Humanitas Research Hospital, Milan, Italy; [12]Centre for Bioinformatics, Department of Informatics, University of Oslo, Oslo; [13]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway; [14]Department of Surgical Pathology, IEO, European Institute of Oncology, IRCCS, Milan, Italy; [15]Athens Technology Center S.A., Athens, Greece; [16]Institut Català d'Oncologia (ICO-Badalona), B-ARGO Group, Care Program, IGTP, Barcelona, Spain

Available online xxx

Background: Pre-treatment predictors of laryngeal preservation (LP) and survival in advanced laryngeal-hypopharyngeal squamous-cell carcinoma (LHSCC) represent an unmet clinical need.

Materials and methods: A multicentric, international, retrospective series of LHSCC patients undergoing induction chemotherapy (IC) within an LP protocol was analyzed. The primary objective was to develop a predictive model by exploiting multiomics data (clinical, genomics, radiomics). Endpoints were laryngo-esophageal dysfunction (LED), response to IC, overall survival (OS), and progression-free survival (PFS). Patients were divided into three groups: group A (no LED); group B (responders to IC with LED); group C (non-responders to IC with LED). Several algorithms (support vector machine, random forest, C5.0, k-nearest neighbors, XGBoost, and naive Bayes) were run and compared in terms of multiclass area under the curve (AUC) score and classification error.

Results: One hundred and ninety-one LHSCC patients were included (median age 60 years, 72% laryngeal, 80% T1-T3, and 58% N+). Responders to IC were 85%, while 66% suffered from LED. The 5-year PFS and OS were 58.4% and 64.7%, respectively. When comparing the three predictive models (clinical, clinical + genomics, clinical + radiomics), the addition of genomics provided the highest AUC. Then, we selected a 64-gene signature and 6 clinical variables (comorbidities, primary site, smoking, T category, N category, performance status) to build up the PRESERVE model. It showed a classification error of 28.9% and an AUC of 87.4%. Risks of major misclassification were low (group A to C, 1.13%; group C to A, 7.38%). Decision analysis confirmed the efficiency of the model.

Conclusions: The PRESERVE model proved to be efficient and accurate in predicting LED and response to IC in LHSCC. External validation is needed before clinical application.

Key words: laryngeal cancer, squamous-cell carcinoma, organ preservation protocols, predictive model, laryngo-esophageal dysfunction, prediction

INTRODUCTION

Worldwide, ~255 000 new laryngeal and hypopharyngeal cancer cases are diagnosed annually.[1] Because of the paucity of symptoms, these patients often present with advanced disease (stage III and IV), resulting in a 5-year relative survival <60%.[2] For years, the gold standard of treatment for locally advanced laryngeal or hypopharyngeal squamous-cell carcinoma (LA-LHSCC) has been surgical removal of the organ [total laryngectomy (TL)].

*Correspondence to: Prof. Paolo Bossi, Medical Oncology and Hematology Unit, IRCCS Humanitas Research Hospital, V. Manzoni 56, Rozzano 20089, Milan, Italy. Tel: +39-3200720438

E-mail: paolo.bossi@hunimed.eu (P. Bossi).

†These authors share first authorship.

‡These authors share last authorship.

However, TL may negatively influence the patient's psychosocial well-being and quality of life.[3] Research is committed to finding new therapeutic strategies for laryngeal preservation (LP). In LP protocols, two different outcomes have equal importance and must be fully achieved to define the treatment as successful: the cure of the cancer and the preservation of a functional organ, i.e. able to swallow (no feeding tube dependence) and breathe (no need of permanent tracheotomy).

The main non-surgical LP approaches foresee different combinations of radiation with systemic treatments.[4] To date, completed trials have shown that long-term results with the highest overall survival (OS) benefit were achieved by using induction chemotherapy (IC) followed by radiotherapy (RT), with LP rates ranging between 47% and 70%.[5-7] Therefore, one out of two to three patients does not benefit from an LP strategy, but instead suffers from the toxicities of the treatment without obtaining functional and/or oncological benefit. Until now, except for limited clinical and radiological characteristics, we lack biomarkers for IC response supporting decision making upfront at the start of LP therapeutic approaches. Validated clinical/biomolecular signatures of the responsive disease would allow selecting patients for a functional LP strategy before administering IC, as well as unveiling the heterogeneous genetic background of individual patients. They could also offer the rationale for enriched trials with new drugs for patients with no expected benefit from a standard IC regimen.

The multiomic algorithm we present in this article results from the analysis of a multicentric, international, retrospective series of LA-LHSCC patients treated in LP protocols including IC. The aim of this study was to define the predictors of laryngo-esophageal dysfunction (LED) and response to IC.

MATERIALS AND METHODS

Patients' series

A multicenter, international European consortium was established within the PRESERVE project among the following centers: 'ASST Spedali Civili di Brescia', Brescia, Italy; 'Istituto Europeo di Oncologia (IEO)', Milan, Italy; 'Institut Català d'Oncologia', Hospitalet de Llobregat, Barcelona, Spain; 'University of Leipzig', Leipzig, Germany.

Retrospective data were retrieved from the following prospective studies and center-specific retrospective series: (i) GSTCC Ita (phase II-III randomized trial)[8,9]; (ii) Interceptor (phase III randomized trial)[10]; (iii) GONO dataset (Italian retrospective consecutive series, unpublished); (iv) Spanish cohort (retrospective consecutive series, unpublished); (v) DeLOS-II (phase II randomized trial).[11]

Main inclusion criteria were as follows: (i) LA-LHSCC; (ii) American Joint Committee on Cancer seventh edition stage III-IV (excluding T4b category and stage IVc); (iii) organ preservation protocol including IC followed by (chemo)-radiation; (iv) availability of clinical data and tumor formalin-fixed paraffin-embedded (FFPE) specimens. The availability of radiological imaging data for radiomics analysis was preferential but did not represent an exclusion criterion.

The full description of the treatment protocols is provided in Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2025.105933. The response to IC was evaluated in each center according to the specific study protocol (Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2025.105933). For the purpose of this study, we did not review all the radiological examinations, while we retained the evaluations carried out in each center that guided the subsequent treatment plan.

The PRESERVE project was approved by the ethics committee of the leading center ('Comitato Etico di Brescia', Italy, NP 4554) and ratified by all the others. All data were pseudonymized. All patients or their legal guardians were required to sign an informed consent; when this was impossible (deceased patient or lost to follow-up), permission was obtained from the National Privacy Authority to use personal data.

Study objectives and endpoints

The primary objective was to develop a multimodal signature predictive of LED and response to IC in LA-LHSCC by exploiting clinical and multiomics data. LED was described by the presence of any of the following after 6 months from the end of treatments: tracheotomy; feeding tube dependence (percutaneous gastrostomy or nasogastric tube); inhalation with repeated episodes of *ab ingestis* pneumonia.

Secondary objectives were: (i) to define alternative therapeutic targets to be exploited in patients not fully responsive to conventional IC; and (ii) to describe the oncological and functional outcomes, and prognosticators, in a series of LA-LHSCC patients treated with organ preservation protocols.

Endpoints were as follows: (i) LED; (ii) response to IC [complete response (CR) versus partial response (PR) versus stable disease (SD) versus progression of disease (PD)] assessed at radiological imaging [computed tomography (CT) or magnetic resonance] according to RECIST 1.1; (iii) OS and progression-free survival (PFS).

Gene expression analysis

The complete workflow is depicted in Supplementary Figure S1, available at https://doi.org/10.1016/j.esmoop.2025.105933.

Histological preparation. The IEO Pathology Division (Milan, Italy) processed FFPE samples, cutting eight slides each. Five-millimeter-thick FFPE consecutive sections were positioned on polyethylene naphthalate membrane (PEN) slides (Leica Microsystems, cat. 11600289) previously UV photoactivated in a UV cross-linker for 30 min and on Superfrost™ slides (TermoFisher, cat.22037246, Waltham, MA) for laser microdissection (LMD) and immunohistochemistry, respectively.

Hematoxylin—eosin (H&E)-stained slides cut from FFPE sections were reviewed by a pathologist (FAM) who annotated the presence of tumor tissue, normal tissue, debris, and lymphocyte-infiltrating tissue with different colors. H&E slides were scanned using NanoZoomer S60 (Hamamatsu Photonics) at ×20 magnification. Annotated areas containing tumor tissue were matched and aligned to the consecutive sections mounted on specific slides for microdissection.

Laser microdissection. FFPE sections mounted on PEN slides were deparaffinized with two changes of xylene and then partially rehydrated in an ethanol gradient up to 75% ethanol (EtOH). Sections were then counterstained for 30 s with freshly prepared cresyl violet (0.8% cresyl violet in 75% EtOH and 4 mM Tris—HCl, pH 8.0) (Merck Life Science S.R.L., cat. C5042), washed in 75%-50% EtOH, and air-dried completely before proceeding to the microdissection. Cresyl violet binds to RNA and prevents its degradation. Single or multiple areas were drawn to collect tumoral tissue using a UV-based LMD7 system (Leica Microsystems, RRID: SCR_020232) at ×5-10 magnification. Whenever possible, multiple samples of morphologically distinct areas were collected from each patient. Microdissected areas were collected by gravity into 1.5 ml Eppendorf caps and immediately processed for RNA extraction.

RNA extraction. FFPE RNA extraction was completed using the MAgMAX™ FFPE RNA Ultra Kit (Applied Biosystems, cat. A31881) following the manufacturer's protocol with minor modifications. Briefly, microdissected areas were digested with a protease solution for 1 h at 55°C, followed by 1 h at 90°C. Then, the supernatant was used for the extraction using an automated extractor (KingFisher™ Flex Magnetic Particle Processor 96DW - TermoFisher, cat.5400630). Changes from the user guide concerned an increased time for three different steps (the sample's binding to magnetic beads, the sample's rebinding to magnetic beads, and elution). The RNA was quantified using Qubit™ RNA High Sensitivity (HS) (ThermoFisher, cat. Q32855) and its quality, including RNA integrity number and distribution value 200, was checked using Agilent RNA 6000 Pico (Agilent Technologies, cat. 5067-1513) on an Agilent 2100 Bioanalyzer instrument.

FFPE samples' library preparation. FFPE samples were processed using the SMART-seq Stranded kit (cat. 634444, protocol version 051018; Takara Bio USA, Inc.) starting from 10 ng of total RNA according to the manufacturer's protocol with minor modifications. Briefly, the fragmentation step was omitted, and RNA molecules were copied into first-strand complementary DNA (cDNA) by reverse transcription using random primers and a template-switching oligo. Sequencing adapters and indexes were added to single-stranded cDNA with five cycles of PCR. cDNA originating from ribosomal RNA was depleted using scZapR in the presence of the mammalian-specific scR-Probes following the manual instructions. The resulting ribo-depleted library fragments were amplified with 14 cycles of PCR, purified with AMPure beads, and profiled for size distribution on the Agilent 2100 Bioanalyzer with DNA HS reagent kits (Agilent Technologies, cat. 5067-4626). Libraries were quantified by Qubit™ DNA HS run on the Agilent 2100 Bioanalyzer instrument and sequenced on an Illumina NovaSeq 6000 platform to obtain 50 bp paired-end reads.

Bioinformatics analysis. RNA sequencing (RNAseq) samples were analyzed using nf-core/rnaseq v3.11.2 with default parameters and GRCh38 genome build. FFPE samples with <1 million uniquely mapped reads were filtered out. Differential gene expression analysis was carried out using DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html).

Transcript counts were computed using Salmon.[12] Salmon estimates transcript abundances by analyzing the alignments from STAR and calculating how likely each read is to originate from a given transcript, considering factors such as fragment length distribution, GC content, and other biases. Using these probabilities, Salmon applies an optimization algorithm to assign read counts across transcripts, providing highly accurate expression levels for each one in the sample.

Radiomics analysis

Radiomics analysis was conducted on contrast-enhanced CT images acquired at baseline (before IC) during the venous phase (70-90 s post-contrast injection) using a voltage of 120 kV and a modulated tube current. The images were reconstructed with filtered back projection, utilizing a soft-tissue kernel and a slice thickness ranging from 1.5 to 2.5 mm.

A double-thresholding technique was applied to exclude voxels with attenuation values outside the range of 0-300 Hounsfield units (HU). Following thresholding, the entire tumor volume was manually contoured by an expert head and neck radiologist. Additionally, the peritumoral region was segmented by expanding the tumor's outer boundaries by 5 mm.

Pre-processing included resampling voxel size to $1 \times 1 \times 1\ mm^3$ and discretizing the images with fixed bins of 15 HU. Feature extraction adhered to an IBSI-compliant protocol and encompassed shape descriptors, first-order statistics, and texture features derived from the original images, as well as from images processed with Laplacian of Gaussian filters (widths of 2, 3, and 4 mm) and wavelet decomposition.

In total, 1130 features were extracted from the tumor volumes and peritumoral regions. Image segmentation was carried out using 3D Slicer (www.slicer.org), and feature extraction was conducted with the Python library PyRadiomics.

Statistics

Patients were classified based on their follow-up status into event and censored groups for survival analysis. An event

was defined as death from any cause for OS and as disease progression for PFS.

Patients were stratified according to larynx functionality and treatment response into three groups: group A, patients without LED irrespective of tumor response; group B, patients with LED and CR or PR; group C, patients experiencing LED without any response to IC.

The Kaplan–Meier method was used to calculate the probability of survival, and the log-rank test was used to assess the statistical significance between groups.

Gene expression processing and filtering. To remove genes with low expression, we filtered out genes with <10 counts in all three outcome levels.

To reduce the dimensionality of the input matrix for classification algorithms, genes were filtered selecting those with a larger difference among the outcome levels. To do so, we fitted negative binomial generalized models using both edgeR[13] and DESeq2[14] and selected the genes with a false discovery rate <0.1 in any contrast among groups in either procedure. This produced a list of 682 genes for downstream classification models.

Expression counts were then normalized with respect to library size and regularized using a variance stabilizing transformation[15] and expressed on a log2 scale.

Algorithm benchmarking. Several algorithms were run and compared in terms of multiclass area under the curve (AUC) score[16] and classification error. The algorithms evaluated were support vector machine (SVM; both with linear and radial kernel), random forest, C5.0,[17] k-nearest neighbors, XGBoost, and naive Bayes. All algorithms were run using default settings. A repeated (B = 50) five-fold cross-validation procedure was adopted to estimate the model's performances.

The benchmarking procedure was carried out both considering the clinical features only and with the addition of expression data.

RESULTS

Demographics, clinical variables, and oncological outcomes

Detailed clinical features of the whole cohort are reported in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105933.

The study cohort consisted of 191 patients, with a median age of 60 years (range 38-78 years). Most patients were male (88%). Primary tumors were predominantly located in the larynx (72%), with 84 (44%) cases of supraglottic and 53 (28%) cases of glottic cancers. They were classified as T1-T3 (80%) and N1-3 (58%). Comorbidities, measured by the Charlson Comorbidity Index (CCI) ≥5, were present in 35% of the cohort. Current, previous, and never all-tobacco smokers were 63%, 33%, and 4%, respectively. Alcohol consumption data were available for 135 patients, of whom 73% had a drinking history. Regarding IC response and subsequent treatment approach, 162 (85%) patients achieved PR (109, 57%) or CR (53, 28%) and therefore continued the organ preservation strategy. The remaining 29 (15%) patients had SD (22, 12%) or PD (7, 3%).

The mean follow-up duration was 57 months. Disease relapse was observed in 31% of patients, with a median disease-free interval of 10 months (range 3-55 months). One hundred and twenty-six (66%) patients suffered from LED.

At the last update, 96 (50%) patients were dead, and 40 (21%) deaths were cancer related. The remaining 95 (50%) patients were alive: 90 (47%) were disease-free while 5 (2.6%) were under treatment for disease relapse.

At 3 and 5 years, PFS was 68.4% and 58.4%, and OS was 75.9% and 64.7%, respectively (Supplementary Figure S2, available at https://doi.org/10.1016/j.esmoop.2025.105933).

Several factors were significantly associated with LED (Table 1). Protective factors included female sex [odds ratio (OR) 0.28, 95% confidence interval (CI) 0.11-0.68, $P = 0.006$], higher prognostic nutritional index values (OR 0.90, 95% CI 0.83-0.97, $P = 0.012$), and achieving PR or CR (PR/CR) compared with PD or SD (PD/SD) (OR 0.05, 95% CI 0.00-0.27, $P = 0.005$). Likewise, pre-treatment preserved laryngeal function (mobile vocal cord) was associated with

Table 1. Clinical factors associated with laryngo-esophageal dysfunction

Variables	*n*	OR	95% CI	*P* value
Sex	191			
Male		—	—	
Female		0.28	0.11-0.68	0.006
Response	191			
PD/SD		—	—	
PR/CR		0.05	0.00-0.27	0.005
Primary site	191			
Hypopharynx		—	—	
Larynx		0.85	0.43-1.65	0.64
Smoking	191			
Never		—	—	
Former/current		2.68	1.44-5.03	0.002
PS	184			
0		—	—	
1		2.25	1.22-4.22	0.010
T category	191			
<4		—	—	
≥4		2.33	1.04-5.74	0.050
N category	191			
0		—	—	
Others		0.89	0.48-1.63	0.70
0-1		—	—	
Others		0.96	0.53-1.76	0.90
CCI	190			
<5		—	—	
≥5		1.82	0.95-3.59	0.075
Vocal cord	151			
Fixed		—	—	
Partially fixed		0.53	0.21-1.36	0.19
Mobile		0.33	0.14-0.73	0.008
Basal Hb	103	1.01	0.79-1.30	0.92
PNI	85	0.90	0.83-0.97	0.012

The table reports the potential predictors of LED tested according to bivariate logistic regression.

CCI, Charlson Comorbidity Index; CI, confidence interval; Hb, hemoglobin; LED, laryngo-esophageal dysfunction; OR, odds ratio; PD/SD, progression of disease/stable disease; PNI, prognostic nutritional index; PR/CR, partial response/complete response; PS, performance status.

better LED outcomes (OR 0.33, 95% CI 0.14-0.73, $P = 0.008$). Conversely, negative prognosticators for an increased risk of LED included smoking (considering both current and former smokers: OR 2.68, 95% CI 1.44-5.03, $P = 0.002$), ECOG performance status (PS) 1 (OR 2.25, 95% CI 1.22-4.22, $P = 0.010$), and T4 category (OR 2.33, 95% CI 1.04-5.74, $P = 0.050$).

Gene expression analysis

A total of 191 patient samples were set and processed.

To identify differences between samples from responsive and resistant patients, we conducted a differential gene expression analysis on the RNAseq data. The samples were divided into four subgroups based on the patients' responses to IC: CR ($n = 53$), PR ($n = 109$), SD ($n = 22$), and PD ($n = 7$). Each subgroup was further divided according to the functional outcome as functional (no LED) versus non-functional (with LED). The resulting heatmap is shown in Figure 1A. Principal component analysis was run to illustrate the distribution of samples according to functionality (Figure 1B) and IC response (Figure 1C). While for functionality we identified a slight separation (Figure 1B), for IC response no significant partition between the subgroups was evident, as they occupy overlapping regions in the plot. The levels of variance between samples were low, around 9% and 22%, underscoring the high similarity across the samples. Additionally, the samples did not cluster according to IC response classification.

To identify suitable therapies for enhancing platinum-based IC efficacy, we investigated the immune 'hot' or 'cold' phenotype in patient samples by analyzing the expression levels of 27 immune-related genes identified by Foy et al.[18] as markers of immune system activation. In this classification, patients with a 'hot' phenotype (HOT score >0) are more likely to respond to immune checkpoint inhibitors (anti-programmed cell death protein 1/anti-programmed death-ligand 1). Our data show that most patients with SD or PD—classified as IC resistant (R)—exhibited a 'cold' phenotype, characterized by minimal immune cell presence in the tumor and likely low immune activation post-immunotherapy. Interestingly, patients with CR, regardless of post-IC LED, also displayed an average cold phenotype, similar to R patients. Conversely, the highest HOT scores were observed in patients with PR and preserved organ function post-IC (Figure 2).

Performance of the predictive algorithm

A classification model for outcome prediction was developed trying to include clinical, genomic, and radiomics data. Patients were stratified according to larynx functionality and treatment response into three groups: group A, patients without LED irrespective of tumor response; group B, patients with LED and CR or PR; group C, patients experiencing LED without any response to IC. The total number of patients pertaining to each group was as follows: group A = 65, group B = 98, and group C = 28.

The predictive algorithm was firstly based on clinical data, and then its performance was compared with models with the addition of genomics or radiomics information.

The clinical characteristics accounted for in the prediction models were as follows: CCI (<5 versus ≥ 5), primary site (hypopharynx versus larynx), smoking (never versus previous, current), T (<4 versus 4), N (0 versus other), PS (0 versus 1).

The three models (clinical, clinical + genomics, clinical + radiomics) were tested with several algorithms and compared in Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2025.105933, and Figure 3.

All algorithms evaluated on the clinical data showed comparable performances, with an AUC of ∼60%. The addition of genomics data significantly improved the predictive power of the model, with the AUC approximately between 60% and 85%. The inclusion of radiomics data forced the reduction of the sample to 106 patients, who were the ones having both clinical and radiological data. The efficiency of this model dropped to ∼50%. Consequently, we did not test a model including all three sources of information.

Then, the SVM-linear algorithm with weights inversely proportional to class prevalence was applied to select the most informative genes and build a reproducible predictive model. Sixty-four genes were identified. The resulting model (64 genes and the abovementioned 6 clinical features, hereafter the 'PRESERVE model') showed a classification error of 28.91% and an AUC of 87.40%. Table 2 shows the predictive accuracy of the PRESERVE model. Interestingly, the risk of extreme misclassification (from group A to C or vice versa) was very low. In fact, the probability that a predicted group A was in reality a group C (i.e. a patient erroneously addressed to IC) was as low as 1.13%. Conversely, the probability of a shift from group C to A (i.e. a candidate to TL who could have benefitted from LP protocol) was 7.4%.

Instead, the misclassification from group C to B (i.e. a candidate to TL who could have benefitted from IC at the expense of some LED) was higher (18.44%).

The receiver operating characteristic curves for the final PRESERVE model (64-gene signature + 6 clinical variables with SVM-linear) are presented in Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2025.105933, with the description of AUCs for a one-versus-all comparison using averaged predicted probability of class assignment from the final model, using repeated five-fold cross-validation (B = 50).

The performance of the PRESERVE model was maintained in the optimization analysis (classification error of 30.95% and AUC of 85.5%, Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2025.105933). Decision analysis confirmed that the PRESERVE model can be an efficient clinical decision support system (Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2025.105933).

DISCUSSION

In this study, we present a predictive model based on clinical and genomics data to anticipate LED and response

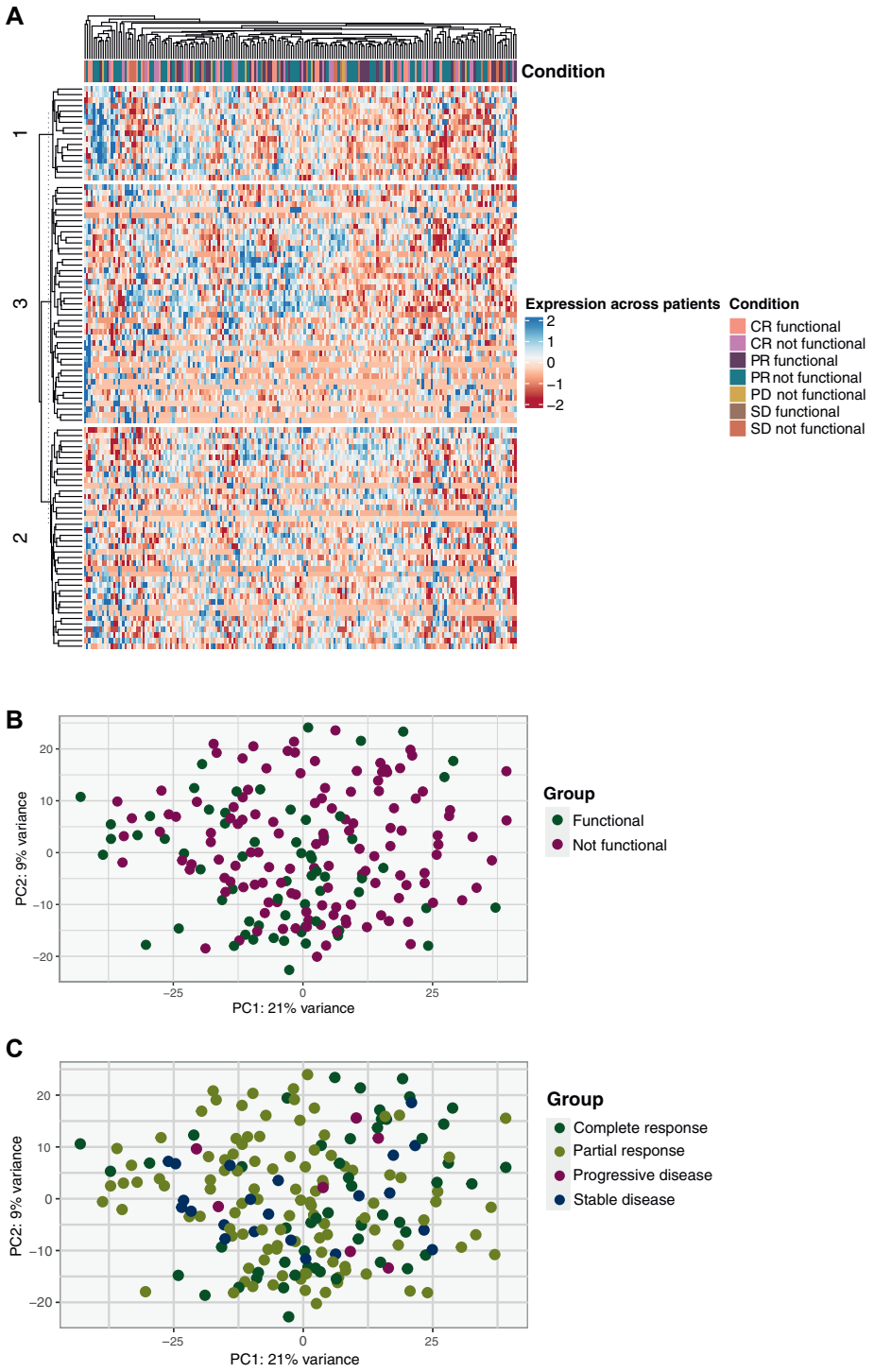

Figure 1. Differential gene expression analysis according to IC response and functional outcome. (A) RNA from patients in the retrospective study was analyzed based on organ functionality post-IC, classifying them into groups according to their IC response: complete response ($n = 53$), partial response ($n = 109$), stable disease ($n = 22$), and progressive disease ($n = 7$). RNA sequencing was carried out, and the heatmap displays z-scores of gene expression between these groups. Each column represents a patient sample, and each row corresponds to a gene. (B) Principal component analysis (PCA) plot illustrating the distribution of samples across the first two principal components. Samples are color-coded based on organ functionality: functional ($n = 65$) and non-functional ($n = 126$). (C) PCA plot illustrating the distribution of samples across the first two principal components. Samples are color-coded by IC response: complete response ($n = 53$), partial response ($n = 109$), stable disease ($n = 22$), and progressive disease ($n = 7$).

to IC in an LP protocol to treat LA-LHSCC, as built on a large retrospective, multicenter series of patients. The power and accuracy of the algorithm are remarkable, and the likelihood of extreme classification errors is low.

The success of an LP protocol is complex and composite, since both oncological and functional outcomes have equal importance and must be achieved. Tumor clearance leaving a non-functional larynx is to be regarded as a failure. In

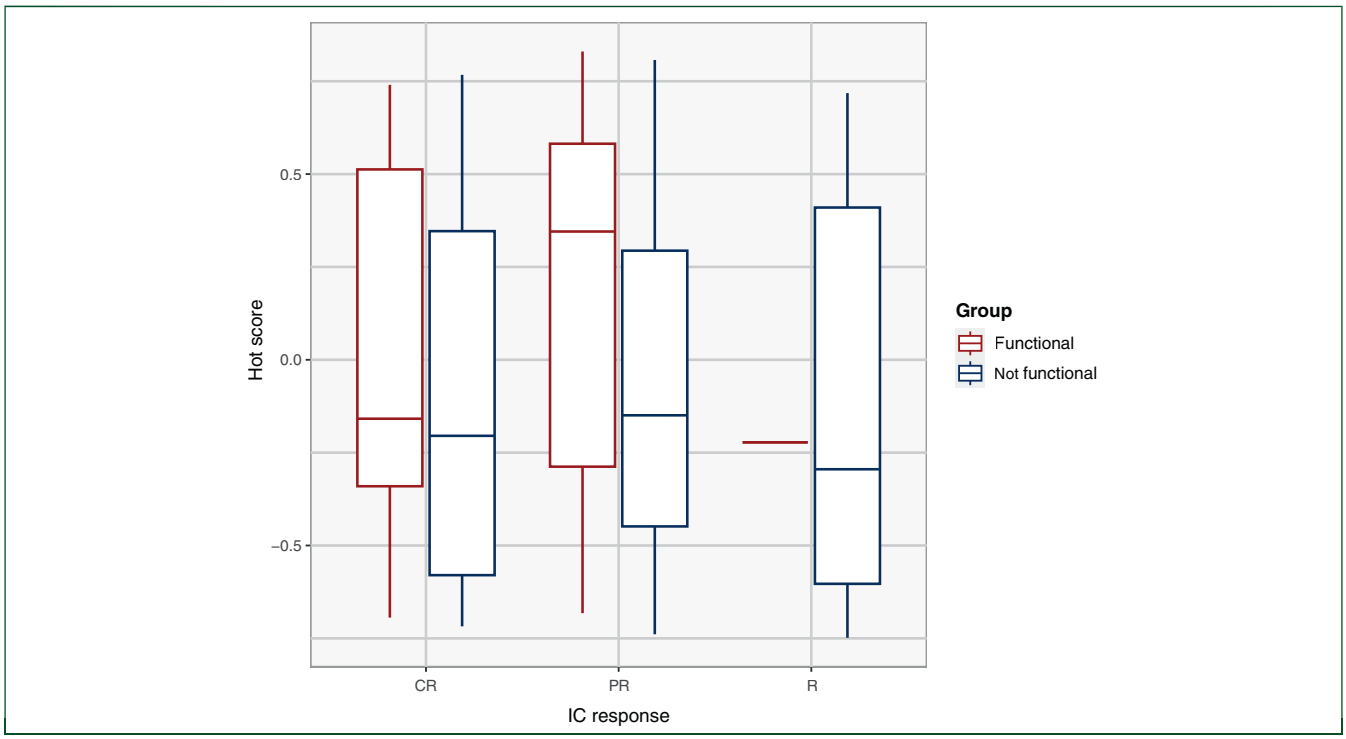

Figure 2. Box plot depicting gene expression changes in HOT score-related genes (Foy et al., 2022) among patients stratified by response to IC and LED. CR, complete response; functional, no LED; IC, induction chemotherapy; LED, laryngo-esophageal dysfunction; not functional, presence of LED; PR, partial response; R, resistant.

fact, the patient is exposed to a highly toxic treatment strategy, with possible delayed morbidity and treatment-related mortality,[7] with a lower probability of disease control in the long term, and a higher risk of complications in case salvage surgery is needed and feasible. If these disadvantages are not counterbalanced by achieving the preservation of breathing and swallowing, with the related positive impact on the quality of life, the cost—benefit balance is negative.

In this regard, anticipation of treatment response and LED in LP protocols is paramount. Obviously, in case SD/PD or relevant loss of functionality could be predicted, the treatment plan could shift toward the classical paradigm [TL and post-operative (chemo)RT] providing the patient with the highest chances of cure at the lowest possible 'biological' expense. Likewise, in the gray zones (i.e. good chances of response with high risk of LED), more informative counseling and shared decision making would be possible.

The first LP protocols were introduced more than 30 years ago,[5] and from the beginning, the issue of identifying biomarkers that predict chemoresistant and radioresistant LA-LHSCC was put on top of the priorities. Nevertheless, till now patient selection has been only roughly standardized, and it is still a matter of debate. Currently, the choice of

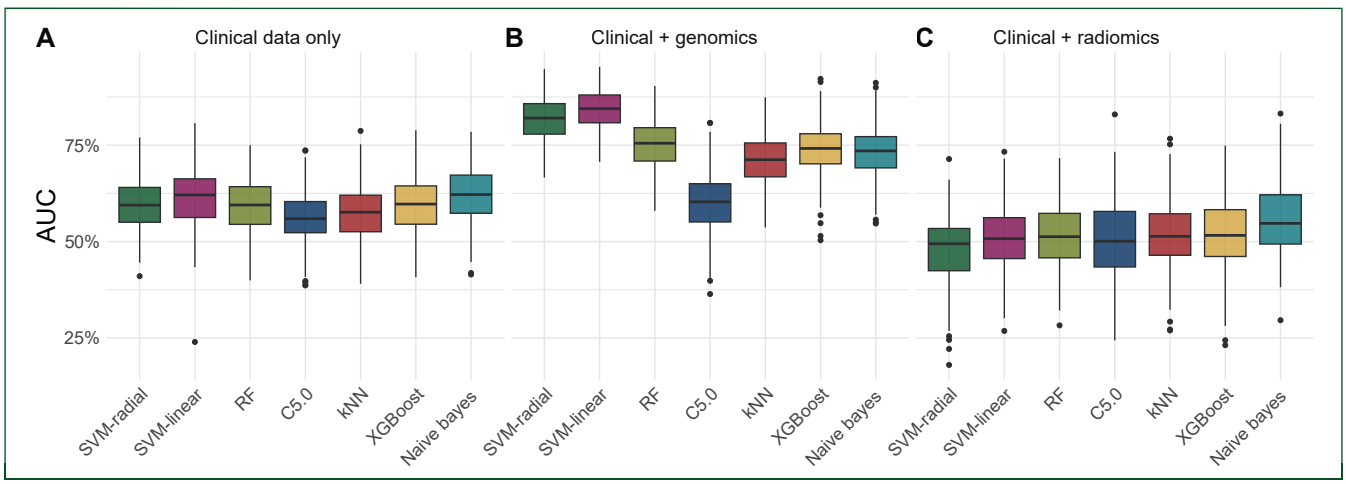

Figure 3. Graphical comparison of the predictive efficiency [measured with the 'area under the curve' (AUC)] of the three models. (A) Predictive efficiency of Clinical data only model. (B) Predictive efficiency of Clinical data + genomics data model. (C) Predictive efficiency of Clinical data + radiomics data model. kNN, k-nearest neighbors; RF, random forest; SVM, support vector machine.

Table 2. Predictive accuracy of the PRESERVE predictive model

Prediction	Real outcome (%)		
	Group A	Group B	Group C
Group A	71.00	27.87	1.13
Group B	20.78	70.47	8.75
Group C	7.38	18.44	74.18

The contingency table shows the predictive accuracy of the PRESERVE predictive model in terms of classification into groups A (no LED), B (IC responders with LED), and C (IC non-responders with LED).
IC, induction chemotherapy; LED, laryngo-esophageal dysfunction.

whether to embark or not on LP protocols is mostly based on physician's expertise, patient's willingness, and clinical features. The latter are quite simple, including PS, age, smoking habit, comorbidities, tumor subsite, tumor burden, nutritional status, tumor metabolic activity, and pre-treatment larynx functionality. These factors are widely supported by published evidence,[7,19-21] and most of them have been confirmed in our analysis. However, although this clinical assessment can be efficient in the identification of the bad candidate, it is largely inadequate to pinpoint relevant toxicity and/or absence of response in apparently good ones.

Different strategies to predict the success of LP protocols have been proposed over the years. Urba et al. demonstrated that tumor response after a single cycle of IC is effective to select patients who could benefit from the prosecution of the LP protocol.[22] This approach is aimed at minimizing unnecessary toxicity and delay in surgical treatment in non-responders. Recently, a score based on four features (number of clinically positive nodes, volume of the residual primary tumor, total residual cancer volume, and Standardized Uptake Value in a Positron Emission Tomography scan) was proposed to improve patient selection after the first IC cycle.[23] Nevertheless, this approach is suboptimal because it still requires the indiscriminate administration of IC, which may imply a risk of toxicity, possible loss of treatment opportunities in non-responders, and dispersive allocation of resources.

In the last decades, research on LP protocol in LHSCC has mostly focused on treatment strategy, i.e. different combinations of drugs or different sequences of treatments.[24-26] Currently, several clinical trials are exploring the possible role of immunotherapy in enhancing LP protocols.[27] However, it is evident that reliable predictive biomarkers for patient selection at the time of diagnosis are lacking, and further advancements in this field should tackle this unmet clinical need. In this setting, precision medicine should mainly translate into early identification of the best candidate for a given LP protocol.

Our series is one of the largest ever published including IC-treated LA-LHSCC investigated within a multiomics framework. We have demonstrated that the addition of a 64-gene signature to clinical data remarkably improved the prediction power of both LED and treatment response. However, the apparent low informative power of radiomics should be weighted cautiously, since the size of the series was largely underpowered to run radiomics analysis.

The PRESERVE model showed promising values of AUC (87.40%) and classification error (28.91%). The latter deserves further considerations. From a clinical standpoint, in a predictive model, extreme errors (responders versus non-responders, or LED versus no LED) are detrimental, since they may lead to erroneous choices in the decision-making process, with possible catastrophic consequences for the patients. As an example, classifying a CR/PR as a SD/PD may induce the clinician to suggest TL to a patient who could benefit from an LP protocol. In this regard, our model works well. In fact, the risks for major errors in decision making (i.e. propose IC in a non-responder, or TL in someone who could have been a candidate for IC) were relatively low (∼1% and 7%, respectively). Moreover, one should consider that this model has been developed in a disadvantageous setting, i.e. a series of patients a priori selected for IC. Therefore, they were considered fit for this treatment according to current standards, while very bad candidates were already excluded. In the real world, this model could work even better. Nevertheless, 'minor' misclassification should not be overlooked. For example, about one out of four responders with LED (27.87%) were anticipated by the model to be without LED (Table 2). This would have exposed a relevant proportion of patients to an unexpected laryngeal toxicity. This finding compels further refinements to the model by improving its predictive accuracy and by introducing data about the time to LED. In fact, in our model, the use of a binary variable for LED (yes/not) hindered the graduation of its severity.

Lastly, our analysis has provided some insights on possible personalized treatments. By applying a previously published HOT score to differentiate immune-hot and immune-cold tumors in head and neck,[18] we demonstrated that patients experiencing PR without LED showed the highest values, which is related to a higher chance of response to immunotherapy. This finding suggests genomic profiling being a possible clue to identify a subgroup of patients in whom IC could be rationally intensified by combining it with immunotherapy to improve tumor response.

Our study has some limitations. The sample size is still limited for statistical analysis of multiomics data. In particular, the low number of patients with available radiological studies hindered the inclusion of radiomics, potentially weakening the performance of the model. The retrospective nature of the series implies the risks for selection bias and data inaccuracy. However, most patients were retrieved from prospectively annotated databases accrued in randomized controlled trials. IC scheme and radiological re-assessments varied across the different series that contributed to our study, which may introduce heterogeneity and some degree of inconsistency. Lastly, in a proportion of patients the time to LED was absent, which hindered the use of laryngo-esophageal-free survival as a study endpoint.

Future refinements of the PRESERVE model should prioritize the validation and fine-tuning on a large,

independent cohort of patients. Moreover, the upcoming introduction of immunotherapy also in the neoadjuvant/induction setting may require relevant adaptation of the model. Firstly, it should be tested against a cohort of patients treated with induction/neoadjuvant immunotherapy to verify its validity. Then, it is foreseeable that some biomarkers related to immune response, such as PD-L1 combined positive score, tumor-infiltrating lymphocytes density, and specific gene signatures, may be included in the model to improve its accuracy in this setting.

Conclusions

The PRESERVE model showed remarkable predictive accuracy and low likelihood of extreme classification errors. External validation in an independent cohort is warranted.

ACKNOWLEDGEMENTS

The authors thank all the patients who participated in the trials and their families as well as all physicians and clinical staff involved in conducting this project and the previously executed investigations. The authors sincerely thank the following colleagues for their invaluable support to this project: Consiglia Paduano (Project Manager), Medical Oncology Unit, ASST Spedali Civili of Brescia, Brescia, Italy; Vittorio Baggio, Department of Radiation Oncology, Treviso Regional Hospital, Treviso, Italy; Maria Grazia Ghi, Oncology Unit 2, Veneto Institute of Oncology (IRCCS), Padua, Italy; Martina Violati, Department of Medical Oncology, San Paolo Hospital, Milan, Italy; Nerina Denaro, Medical Oncology, Santa Croce e Carle General Hospital, Cuneo, Italy; Ilaria Imarisio, Unit of Oncology, Department of Oncology, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy; Giuseppe Azzarello, Oncology and Haematology Unit, AULSS 3 Serenissima—Mirano Hospital, Venice, Italy; Silvia Mezi, Department of Radiological, Oncological, and Anatomo-Pathological Science 'Sapienza', University of Rome, Rome, Italy; Luca Galli, Medical Oncology Unit 2, Santa Chiara Hospital, AOU Pisana, Pisa, Italy; Fabio Trippa, Oncology Department, Radiation Oncology Centre, Santa Maria Hospital, Terni, Italy; and Alicia Tosoni, Nervous System Medical Oncology Department, IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy.

FUNDING

This work was supported by Fondazione Regionale per la Ricerca Biomedica (Regione Lombardia), Project ERAPERMED2020—283 under the frame of ERA PerMed [grant number 779282] and by the Sächsische Aufbaubank (SAB WI413, grant number 100610496). SCa was supported by research grants from the Italian Ministry of University (PRIN project no. 2022SYXEHJ_002). PB was supported by research grants from the Italian Ministry of University (PRIN 2020 grant; project code 20209KY3Y7).

DISCLOSURE

RM declares support for travel to meetings from MSD, Merck KgA; payment for lectures including service on speakers' bureaus from MSD, Merck, Roche; advisory board with compensation from Merck, MSD, Roche, Boehringer, Segean, Pfizer, Regeneron, Genmab. MP has received honoraria and funding for travel and accommodations from Eisai Europe. PB reported participation to advisory board or conference honoraria for Merck, Sanofi-Regeneron, Merck Sharp & Dohme, Daiichi-Sankyo, GlaxoSmithKline, Merus, Sun Pharma, Pfizer, Genmab, Bicara, Angelini, Nutricia. All other authors have declared no conflicts of interest.

DATA SHARING

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
2. Botta L, Capocaccia R, Bernasconi A, et al. Estimating cure and risk of death from other causes of cancer patients: EUROCARE-6 data on head & neck, colorectal, and breast cancers. *Eur J Cancer*. 2024;208:114187.
3. Perry A, Casey E, Cotton S. Quality of life after total laryngectomy: functioning, psychological well-being and self-efficacy. *Int J Lang Commun Disord*. 2015;50(4):467-475.
4. Bonomi MR, Blakaj A, Blakaj D. Organ preservation for advanced larynx cancer: a review of chemotherapy and radiation combination strategies. *Oral Oncol*. 2018;86:301-306.
5. Department of Veterans Affairs Laryngeal Cancer Study Group, Wolf GT, Fisher SG, Hong WK, et al. Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. *N Engl J Med*. 1991;324(24):1685-1690.
6. Lefebvre JL, Chevalier D, Luboinski B, Kirkpatrick A, Collette L, Sahmoud T. Larynx preservation in pyriform sinus cancer: preliminary results of a European Organization for Research and Treatment of Cancer phase III trial. EORTC Head and Neck Cancer Cooperative Group. *J Natl Cancer Inst*. 1996;88(13):890-899.
7. Forastiere AA, Zhang Q, Weber RS, et al. Long-term results of RTOG 91-11: a comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer. *J Clin Oncol*. 2013;31(7):845-852.
8. Paccagnella A, Orlando A, Marchiori C, et al. Phase III trial of initial chemotherapy in stage III or IV head and neck cancers: a study by the Gruppo di Studio sui Tumori della Testa e del Collo. *J Natl Cancer Inst*. 1994;86(4):265-272.
9. Paccagnella A, Ghi MG, Loreggian L, et al. Concomitant chemoradiotherapy versus induction docetaxel, cisplatin and 5 fluorouracil (TPF) followed by concomitant chemoradiotherapy in locally advanced head and neck cancer: a phase II randomized study. *Ann Oncol*. 2010;21(7):1515-1522.
10. Merlano MC, Denaro N, Vecchio S, et al. Phase III randomized study of induction chemotherapy followed by definitive radiotherapy + cetuximab versus chemoradiotherapy in squamous cell carcinoma of head and neck: the INTERCEPTOR-GONO study (NCT00999700). *Oncology*. 2020;98(11):763-770.
11. Dietz A, Wichmann G, Kuhnt T, et al. Induction chemotherapy (IC) followed by radiotherapy (RT) versus cetuximab plus IC and RT in advanced laryngeal/hypopharyngeal cancer resectable only by total laryngectomy—final results of the larynx organ preservation trial DeLOS-II. *Ann Oncol*. 2018;29(10):2105-2114.
12. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-419.

13. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40(10):4288-4297.
14. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
15. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.
16. Fawcett T. Using rule sets to maximize ROC performance. In: *Proceedings of the 2001 IEEE International Conference on Data Mining*. USA: IEEE Computer Society; 2001. p. 131-138.
17. Kuhn M, Johnson K. *Applied Predictive Modeling*. New York: Springer; 2013.
18. Foy JP, Karabajakian A, Ortiz-Cuaran S, et al. Immunologically active phenotype by gene expression profiling is associated with clinical benefit from PD-1/PD-L1 inhibitors in real-world head and neck and lung cancer patients. *Eur J Cancer*. 2022;174:287-298.
19. Bradford CR, Wolf GT, Carey TE, et al. Predictive markers for response to chemotherapy, organ preservation, and survival in patients with advanced laryngeal carcinoma. *Otolaryngol Head Neck Surg*. 1999;121(5):534-538.
20. Forastiere AA, Goepfert H, Maor M, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med*. 2003;349(22):2091-2098.
21. Semrau S, Haderlein M, Schmidt D, et al. Single-cycle induction chemotherapy followed by chemoradiotherapy or surgery in patients with head and neck cancer: what are the best predictors of remission and prognosis? *Cancer*. 2015;121(8):1214-1222.
22. Urba S, Wolf G, Eisbruch A, et al. Single-cycle induction chemotherapy selects patients with advanced laryngeal cancer for combined chemoradiation: a new treatment paradigm. *J Clin Oncol*. 2006;24(4):593-598.
23. Wichmann G, Krüger A, Boehm A, et al. Induction chemotherapy followed by radiotherapy for larynx preservation in advanced laryngeal and hypopharyngeal cancer: outcome prediction after one cycle induction chemotherapy by a score based on clinical evaluation, computed tomography-based volumetry and ^{18}F-FDG-PET/CT. *Eur J Cancer*. 2017;72:144-155.
24. Lefebvre JL, Pointreau Y, Rolland F, et al. Induction chemotherapy followed by either chemoradiotherapy or bioradiotherapy for larynx preservation: the TREMPLIN randomized phase II study. *J Clin Oncol*. 2013;31(7):853-859.
25. Haddad R, O'Neill A, Rabinowits G, et al. Induction chemotherapy followed by concurrent chemoradiotherapy (sequential chemoradiotherapy) versus concurrent chemoradiotherapy alone in locally advanced head and neck cancer (PARADIGM): a randomised phase 3 trial. *Lancet Oncol*. 2013;14(3):257-264.
26. Pointreau Y, Garaud P, Chapet S, et al. Randomized trial of induction chemotherapy with cisplatin and 5-fluorouracil with or without docetaxel for larynx preservation. *J Natl Cancer Inst*. 2009;101(7):498-506.
27. Mattavelli D, Wichmann G, Smussi D, et al. Is precision medicine the solution to improve organ preservation in laryngeal/hypopharyngeal cancer? A position paper by the Preserve Research Group. *Front Oncol*. 2024;14:1433333.