European Radiology (2021) 31:3306–3314
https://doi.org/10.1007/s00330-020-07439-8

GASTROINTESTINAL

Addition of HER2 and CD44 to ^{18}F-FDG PET–based clinico-radiomic models enhances prediction of neoadjuvant chemoradiotherapy response in esophageal cancer

Roelof J. Beukinga[1] · Da Wang[2,3] · Arend Karrenbeld[4] · Willemieke P. M. Dijksterhuis[2] · Hette Faber[3,5] · Johannes G. M. Burgerhof[6] · Véronique E. M. Mul[5] · Riemer H. J. A. Slart[1,7] · Robert P. Coppes[3,5] · John Th. M. Plukker[2]

Received: 12 May 2020 / Revised: 21 September 2020 / Accepted: 21 October 2020 / Published online: 5 November 2020

Abstract

Objectives To assess the complementary value of human epidermal growth factor receptor 2 (HER2)-related biological tumor markers to clinico-radiomic models in predicting complete response to neoadjuvant chemoradiotherapy (NCRT) in esophageal cancer patients.

Methods Expression of HER2 was assessed by immunohistochemistry in pre-treatment tumor biopsies of 96 patients with locally advanced esophageal cancer. Five other potentially active HER2-related biological tumor markers in esophageal cancer were examined in a sub-analysis on 43 patients. Patients received at least four of the five cycles of chemotherapy and full radiotherapy regimen followed by esophagectomy. Three reference clinico-radiomic models based on ^{18}F-FDG PET were constructed to predict pathologic response, which was categorized into complete versus incomplete (Mandard tumor regression grade 1 vs. 2–5). The complementary value of the biological tumor markers was evaluated by internal validation through bootstrapping.

Results Pathologic examination revealed 21 (22%) complete and 75 (78%) incomplete responders. HER2 and cluster of differentiation 44 (CD44), analyzed in the sub-analysis, were univariably associated with pathologic response. Incorporation of HER2 and CD44 into the reference models improved the overall performance (R^2s of 0.221, 0.270, and 0.225) and discrimination AUCs of 0.759, 0.857, and 0.816. All models exhibited moderate to good calibration. The remaining studied biological tumor markers did not yield model improvement.

Conclusions Incorporation of HER2 and CD44 into clinico-radiomic prediction models improved NCRT response prediction in esophageal cancer. These biological tumor markers are promising in initial response evaluation.

Hette Faber is deceased. This paper is dedicated to his memory.

Roelof J. Beukinga and Da Wang equally contributed to this work and are in control over the presented data.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-020-07439-8.

✉ Roelof J. Beukinga
r.j.beukinga@umcg.nl

[1] Medical Imaging Center, Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands

[2] Department of Surgical Oncology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[3] Department of Biomedical Sciences of Cells and Systems, Section Molecular Cell Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[4] Department of Pathology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[5] Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[6] Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[7] Faculty of Science and Technology, Department of Biomedical Photonic Imaging, University of Twente, Enschede, The Netherlands

Key Points

- *A multimodality approach, integrating independent genomic and radiomic information, is promising to improve prediction of γpCR in patients with esophageal cancer.*
- *HER2 and CD44 are potential biological tumor markers in the initial work-up of patients with esophageal cancer.*
- *Prediction models combining ^{18}F-FDG PET radiomic features with HER2 and CD44 may be useful in the decision to omit surgery after neoadjuvant chemoradiotherapy in patients with esophageal cancer.*

Keywords Positron emission tomography · Radiomics · Oncogene protein HER-2 · CD44 antigen · Esophageal cancer

Abbreviations

^{18}F-FDG PET	^{18}F-Fluorodeoxyglucose positron emission tomography
CD44	Cluster of differentiation 44
CROSS	Chemoradiotherapy for Oesophageal Cancer followed by Surgery Study
CT	Computed tomography
CXCR4	C-X-C chemokine receptor type 4
HER2	Human epidermal growth factor receptor 2
HIF1α	Hypoxia-inducible factor 1-alpha
LRT	Likelihood-ratio test
NCRT	Neoadjuvant chemoradiotherapy
PTCH1	Protein patched homolog 1
SHH	Sonic Hedgehog
TRG	Tumor regression grade

Introduction

Neoadjuvant chemoradiotherapy (NCRT) followed by surgery is the preferred treatment for locally advanced ($T_1/N_{1–3}/M_0$; $T_{2–4a}/N_{0–3}/M_0$) curative-intended resectable esophageal cancer. Pathologic complete response to NCRT (γpCR) is achieved in approximately 29% of these patients [1]. For complete responders, a "wait-and-see" policy instead of esophagectomy may lead to equivalent results and avoids surgical morbidity and mortality. In evaluating tumor response, clinicians usually report features extracted from pre- and posttreatment imaging such as overall tumor volume measured on computed tomography (CT) and maximum standardized uptake value measured on ^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET). However, these features have relatively unsatisfactory predictive values; and hence, no clinical relevant conclusions can be drawn from these features only [2].

These features neglect spatial information, while intratumoral heterogeneity is associated with higher levels of tumor aggressiveness and impaired response to NCRT. Spatial variation in either gene expression profiles or environmental stressors leads to multiple distinct subclonal populations with different patterns of oxygen consumption and glucose metabolism [3–5]. Intratumoral heterogeneity can be quantified using a sophisticated method called radiomics, which extracts a large number of quantitative imaging features from medical images to capture the tumor phenotype and its microenvironment. These features quantify image intensity, shape, and texture and may be of added value to traditionally clinico-pathological reports. Numerous recent studies have demonstrated that CT and/or ^{18}F-FDG PET radiomics can outperform conventional radiological measurements in the prediction of γpCR in esophageal cancer patients [6–15]. Nevertheless, the use of radiomic features as quantitative image biomarkers requires further optimization and improvement in order to achieve routine clinical application.

To improve radiomics-based prediction models, incorporation of molecular targets involved in treatment resistance mechanisms is promising [16]. Studies on molecular targets are usually based on a single biopsy; therefore, they do not fully capture the spectrum of resistance clones within the individual patients' specific tumor. Imaging features may therefore be complementary as they capture independent information from the entire tumor burden [17–19]. An important biological tumor marker which has gained predictive characteristics in esophageal cancer is human epidermal growth factor receptor 2 (HER2) [20, 21]. Activation of the HER2 proto-oncogene initiates signaling pathways leading to proliferation, inhibition of apoptosis, and tumor progression [22, 23]. Other potentially active biological tumor markers predicting resistance to NCRT are cluster of differentiation 44 (CD44) and the Hedgehog pathway markers receptor protein patched homolog 1 (PTCH1) and ligand Sonic Hedgehog (SHH) [24]. Moreover, transcription factor hypoxia-inducible factor 1-alpha (HIF1α) and the biological interaction between HER2 and CD44 may lead to upregulation of C-X-C chemokine receptor type 4 (CXCR4), which promotes tumor progression and NCRT resistance in gastroesophageal cancer [25, 26].

The aim of this study was to assess the complementary value of HER2 and its associated biological tumor markers to ^{18}F-FDG PET–based clinico-radiomic prediction models to predict γpCR in esophageal cancer patients.

Patients and methods

Patients

This retrospective study was granted by the Local Institutional Review Board and obtaining informed consent was waived according to the legal regulations of our University Hospital. Patients were eligible for inclusion if they had histologically confirmed locally advanced ($T_1/N_{1-3}/M_0$; $T_{2-4a}/N_{0-3}/M_0$) esophageal cancer (according to the seventh tumor–node–metastasis classification system) and if sufficient amounts of pre-treatment biopsy material were available [27]. Moreover, patients were only included if they had a baseline ^{18}F-FDG PET/CT scan to perform the radiomics analysis and received at least four of the five cycles of chemotherapy and full concomitant radiotherapy, followed by esophagectomy with curative intent in our hospital. The enrolled 96 patients were treated between March 2010 and June 2018 (group 1). In only 43 of these 96 patients (group 2) the additional required pre-treatment biopsy material was available to perform analyses of CD44, HIF1α, PTCH1, and SHH (Fig. 1).

Treatment and pathology

NCRT was given according to the Chemoradiotherapy for Oesophageal Cancer followed by Surgery Study (CROSS) regimen, including carboplatin (AUC of 2 mg min mL^{-1}) and paclitaxel (50 mg/m^2) with concurrent radiotherapy (41.4 Gy in 23 fractions) [1]. A curative intended transthoracic open or minimal invasive esophagectomy with mediastinal and upper abdominal lymphadenectomy was performed 6–8 weeks after completion of NCRT. The primary outcome was pathologic response according to the Mandard tumor regression grade (TRG) [28]. This grading system classifies the ratio of residual vital tumor cells and the degree of NCRT-induced fibrosis and defines γpCR; γpT0N0 (Mandard TRG 1) as no residual vital tumor cells and non-γpCR (Mandard TRG 5) as no tumor regression at all.

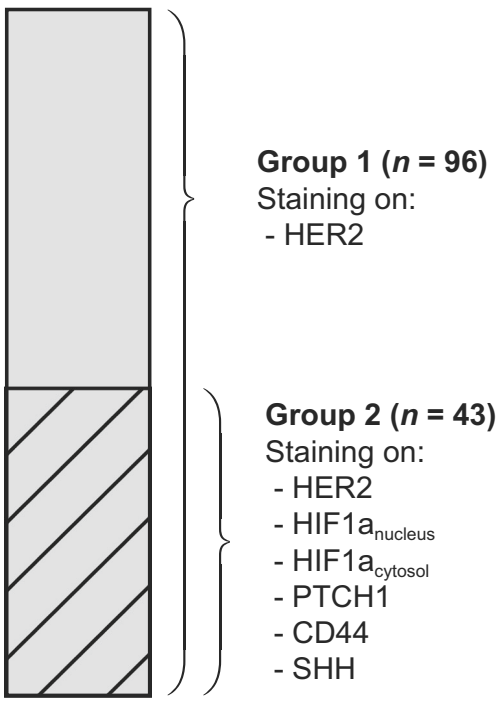

Fig. 1 Illustration of patient groups. HER2 analysis was performed in 96 patients (group 1). In 43 of these 96 patients, sufficient pre-treatment biopsy was available to perform additional analyses (group 2)

^{18}F-FDG PET/CT

Pre-treatment ^{18}F-FDG PET/CT (Biograph mCT-64 PET/CT; Siemens) scans were acquired in radiation treatment planning position. Patients were instructed to fast except for the consumption of water for at least 6 h before scanning. Images were acquired 60 min after the intravenous injection of 3 MBq/kg ^{18}F-FDG. ^{18}F-FDG PET images were obtained within 2–3 min per bed position in three-dimensional setting. Images were reconstructed using a time-of-flight iterative reconstruction method (three iterations; 21 subsets) with point-spread-function correction [29]. Images were corrected for random coincidences, scatter, and attenuation (CT-based), and were smoothed with a Gaussian filter of 6.5 mm in full-width at half-maximum.

Radiomic feature extraction

Tumor volume was delineated manually after reaching consensus between 3 collaborating researchers (RTx Workstation 1.0; Mirada Medical). In-house software was developed with Matlab 2018a (MathWorks) to process baseline ^{18}F-FDG PET images and to extract 101 radiomic features [30]. Quantification of ^{18}F-FDG PET offers spatial information on the rate of metabolism, which is affected by well-known risk factors for tumor NCRT resistance such as hypoxia, necrosis, and cellular proliferation. As image texture depends on image quality, the low-dose CT images were not analyzed. The extracted radiomic features consisted of 19 morphologic features (including the conventional metrics volume and total lesion glycolysis), two local intensity features, 18 statistical features (including the conventional metrics SUV_{max}, SUV_{peak}, and SUV_{mean}), 25 gray-level co-occurrence–based features, 16 gray-level run-length–based features, 16 gray-level size-zone–based features, and five neighborhood gray-tone difference–based features. The extracted radiomic features were all listed in the IBSI reference manual and matched the IBSI benchmark values [30]. The Supplemental Method provides a more detailed description regarding the radiomic feature extraction process.

Immunohistochemistry and scoring of biological tumor markers

Immunohistochemistry staining was performed on 5-μm tissue sections from archival biopsies using primary antibodies against HER2 (1:100, Fremont), CD44 (1:100, Biolegend), HIF1α (1:100, ABCAM), PTCH1 (1:100, ABCAM), and SHH (1:100, ABCAM). De-paraffinized tissue sections were immersed in PBS 2% hydrogen peroxidase to block endogenous peroxidase activity. Antigen retrieval was performed and the sections were incubated overnight at 4 °C with primary antibodies. Tissue sections were then incubated with biotinylated secondary antibodies at 1:300 dilutions for 1 h. The ABC complex was formed using the Vectastain Elite ABC HRP kit (Vector Laboratories). This complex was visualized with SIGMA FAST 3,3'-diaminobenzidine tablets (Sigma-Aldrich). Finally, sections were counterstained with hematoxylin and scored.

Scoring was blinded and carried out by two researchers independently. Discordant cases and random slides of each marker were validated by a pathologist specializing in upper gastrointestinal cancer. If at least 5 clustered tumor cells were stained, HER2 was scored according to standardized methods in a discrete scale of 0, 1+, 2+, and 3+ [22, 31]. Following the guidelines in the immunohistochemical evaluation of HER2, this scale was dichotomized. Tumors with score 0 were considered negative while tumors with score 3+ were considered positive. As tumors with 1+ represent the majority of false-negative results and tumors with 2+ represent the majority of false-positive results, these cases were subjected to fluorescence in situ hybridization to confirm the HER2 status [22, 23, 32]. HIF1α was scored separately as it is expressed either in the cytosol (normoxia) or in the nucleus (hypoxia). CD44, $HIF1\alpha_{cytosol}$, and PTCH1 were scored using the 15-point immuno-reactivity score as described in the Supplemental Methods. $HIF1\alpha_{nucleus}$ and SHH intensity were categorized as either present or absent [33].

Statistical analysis

Statistical analysis was performed with R 3.5.3 open-source software using the regression modeling strategies package (version 5.1-3), available from the Comprehensive R Archive Network (http://www.r-project.org). The outcome variable, categorized as γpCR versus non-γpCR (Mandard TRG 1 vs. 2–5), was modeled using logistic regression. Biological tumor markers with a likelihood-ratio test (LRT) p value < 0.2 in univariable logistic regression analysis were preselected and entered separately to three reference models. The reference models were constructed based on clinical features as listed in Table 1 (model 1), radiomic features (model 2), and both clinical and radiomic features (model 3). Prior to the final selection of radiomic features, the feature space was reduced by agglomerative hierarchical clustering with average linkage to group both radiomic features and patients in clusters based on their Spearman rank correlation coefficient. Clusters were formed when groups of nodes in the dendrogram were < 60% of the maximum linkage. From each feature cluster, a representative radiomic feature was selected based on the lowest univariable LRT p value per cluster. Only representative radiomic features with LRT $p < 0.2$ were subjected to the final feature selection process based on least absolute shrinkage and selection operator, a technique for L1-norm regularization.

The performance of the constructed models was assessed using overall performance, discrimination, and calibration measures. Discrimination describes the ability to discriminate between γpCR and non-γpCR and was measured using the AUC and the discrimination slope. Calibration refers to the agreement between observed outcomes and predictions, and was evaluated with calibration plots. Measures of overall performance are composed of discrimination and calibration characteristics of the model and include the Nagelkerke R^2 and the Brier score [34]. All measures were corrected for model optimism by internal validation with bootstrapping (20,000 repetitions).

Results

Patient characteristics

Table 1 displays the patient and tumor characteristics. Seventy-eight patients (81%) in group 1 and 33 patients (77%) in group 2 received all 5 cycles of chemotherapy, while 18 patients (19%) in group 1 and 10 patients (23%) in group 2 received 4 cycles of chemotherapy. In group 1, 21 patients (22%) and 75 patients (78%) were scored as γpCR and non-γpCR, compared to respectively 9 patients (21%) and 34 patients (79%) in group 2.

Immunohistochemistry scores

HER2 amplification could not be established in two patients in group 1 and one patient in group 2 and was therefore scored as missing. Immunohistochemical staining for all remaining samples was successfully performed. The distributions of the immunohistochemical scores of the six analyzed biological tumor markers are provided in Table 2. HER2 and CD44 were the only significant biological tumor markers at univariable logistic regression analysis (LRT p value = 0.043 and LRT p value = 0.051) and were therefore considered for further analysis. $HIF1\alpha_{nucleus}$, $HIF1\alpha_{cytosol}$, PTCH1, and SHH were not found to be significant (LRT p values of 0.847, 0.367, 0.443, and 0.236 respectively).

Table 1 Patient and tumor characteristics

Characteristics	Group 1		Group 2	
	γpCR (n = 21)	Non-γpCR (n = 75)	γpCR (n = 9)	Non-γpCR (n = 34)
Sex				
Male	14 (18%)	64 (82%)	7 (19%)	29 (81%)
Female	7 (39%)	11 (61%)	2 (29%)	5 (71%)
Age (median (IQR)) (years)	65 (7)	63 (10)	65 (6)	63 (10)
Histology				
Adenocarcinoma	16 (18%)	72 (82%)	8 (%)	34 (81%)
Squamous cell carcinoma	5 (62%)	3 (38%)	1 (100%)	0 (0%)
Tumor location				
Distal esophagus/GEJ	21 (22%)	75 (78%)	9 (21%)	34 (79%)
Tumor length (median (IQR)) (cm)	5.0 (5.3)	6.0 (4.0)	3.0 (6.5)	6.0 (4.0)
Clinical T-stage				
T1 and T2	6 (55%)	5 (45%)	3 (50%)	3 (50%)
T3 and T4a	15 (18%)	70 (82%)	6 (16%)	31 (84%)
Clinical N-stage				
N0 and N1	19 (28%)	49 (72%)	8 (27%)	22 (73%)
N2 and N3	2 (7%)	26 (93%)	1 (8%)	12 (92%)
Number of chemotherapy cycles				
4	4 (22%)	14 (78%)	2 (20%)	8 (80%)
5	17 (22%)	61 (78%)	7 (21%)	26 (79%)
Mandard tumor regression grade				
1	21 (100%)	0 (0%)	9 (100%)	0 (0%)
2	0 (0%)	24 (100%)	0 (0%)	12 (100%)
3	0 (0%)	33 (100%)	0 (0%)	17 (100%)
4	0 (0%)	16 (100%)	0 (0%)	4 (100%)
5	0 (0%)	2 (100%)	0 (0%)	1 (100%)

Abbreviations: *γpCR*, pathologic complete response; *IQR*, interquartile range; *GEJ*, gastroesophageal junction

High HER2 expression was found to be more prevalent in impaired responders. To be more specific, 17 of the 18 HER2-positive patients developed non-γpCR (negative predictive value of 94%) and HER2 was negative in 19 of the 20 patients with γpCR (sensitivity of 95%). Moreover, 18 of the 18 (100%) HER2-positive patients had a clinical $T_{3–4a}$ tumor and 11 of the 76 (14%) HER2-negative patients had a clinical $T_{1–2}$ tumor (p value = 0.12, Fisher's exact test). Supplemental Figure 1 demonstrates representative pictures of low (0) and high (3+) HER2 expression patterns. CD44 amplification was found more frequently in complete responders than in incomplete responders. From 16 CD44-negative patients, 15 developed a non-γpCR (positive predictive value of 94%), while 8 of the 9 patients with γpCR were CD44-positive (sensitivity of 89%). Only 1 of the 16 (6%) CD44-negative patients had a $T_{1–2}$ tumor, while 22 of the 27 (81%) CD44-positive patients had a $T_{3–4a}$ tumor (p value = 0.35, Fisher's exact test). In group 2, all 6 patients who were both HER2-positive and CD44-positive developed a non-γpCR.

Reference model construction

Hierarchical clustering revealed 7 groups of correlated features (Fig. 2). The representative features corresponding to these feature clusters (in Fig. 2, from top to bottom in the tree) were inverse variance, coarseness, Moran's I index, second measure of information correlation, elongation, Geary's C measure, and long-run low gray-level emphasis. The conventional metrics (volume, total lesion glycolysis, SUV_{max}, SUV_{peak}, and SUV_{mean}) were not selected as representative features. Model 1 was constructed based on histology and clinical T-stage, while Geary's C measure and long-run low gray-level emphasis were selected from the representative features for model 2. Geary's C measure is an indicator of spatial autocorrelation for finding repeating metabolic patterns. Long-run low gray-level emphasis is dependent on long sets of aligned voxels with low metabolic activity. Model 3 was the full model incorporating all variables of models 1 and 2.

Table 2 The distribution of the immunohistochemistry scores

Tumor markers	Group 1		Group 2	
	γpCR (n = 21)	Non-γpCR (n = 75)	γpCR (n = 9)	Non-γpCR (n = 34)
HER2				
Negative	19 (90%)	57 (76%)	8 (89%)	26 (76%)
Positive	1 (5%)	17 (23%)	0 (0%)	8 (24%)
Missing	1 (5%)	1 (1%)	1 (11%)	0 (0%)
$HIF1\alpha_{nucleus}$				
Negative			8 (89%)	31 (91%)
Positive			1 (11%)	3 (9%)
$HIF1\alpha_{cytosol}$				
Negative			2 (22%)	13 (38%)
Positive			7 (78%)	21 (62%)
PTCH1				
Negative			2 (22%)	4 (12%)
Positive			7 (78%)	30 (88%)
CD44				
Negative			1 (11%)	15 (44%)
Positive			8 (89%)	19 (56%)
SHH				
Negative			0 (0%)	3 (9%)
Positive			9 (100%)	31 (91%)

Abbreviations: *γpCR*, pathologic complete response; *HER2*, human epidermal growth factor receptor 2; *CD44*, cluster of differentiation 44; *HIF1α*, hypoxia-inducible factor alpha; *PTCH1*, protein patched homolog 1; *SHH*, Sonic Hedgehog

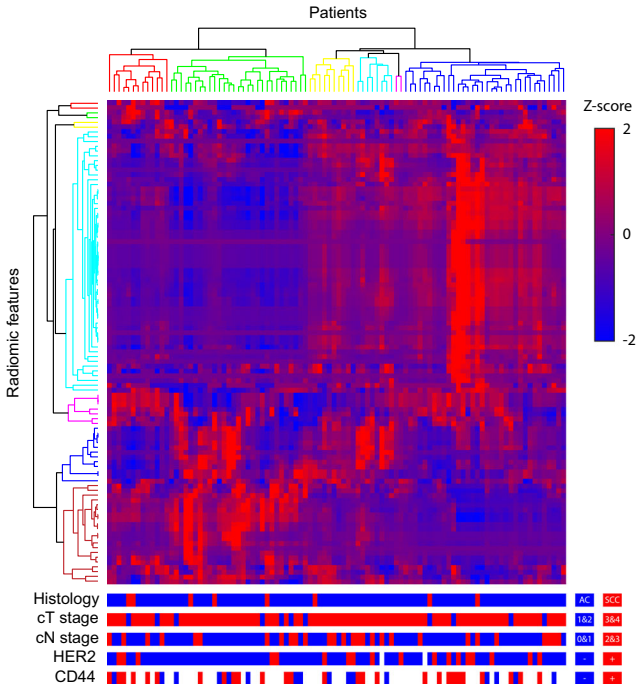

Fig. 2 Heatmap for radiomic feature expression with a *Z*-score. Hierarchical clustering revealed 7 radiomic feature clusters (different tree colors along the y-axis) and 6 patient clusters (different tree colors along the x-axis) with similar radiomic feature expression patterns. Representative radiomic features corresponding to the feature clusters (from top to bottom in the tree) were inverse variance, coarseness, Moran's I index, second measure of information correlation, elongation, Geary's C measure, and long-run low gray-level emphasis. We tested whether clinico-pathological features (histology, clinical T-stage, and clinical N-stage) and biological expression (HER2 and CD44) are distributed equally across different patient clusters. The fact that no association was found suggests independent information of these multimodal features

Complementary value of the biological tumor markers

As no associations were found between the patient clusters and histology, clinical T-stage, clinical N-stage, HER2, and CD44 (χ^2 test; Fig. 2), we assumed that there was no connection between gene expression profiles and imaging phenotypes. HER2 and CD44 were separately added to the abovementioned reference models to evaluate their complementary value. The performance measures of all constructed models are depicted in Table 3. The reference models had the following performance measures: overall model performance (R^2_{M1} = 0.140, R^2_{M2} = 0.103, and R^2_{M3} = 0.173), discrimination (AUC_{M1} = 0.657, AUC_{M2} = 0.654, and AUC_{M3} = 0.685), and calibration ($Intercept_{M1}$ = 0.086 and $Slope_{M1}$ = 0.817; $Intercept_{M2}$ = 0.047 and $Slope_{M2}$ = 0.826; and $Intercept_{M3}$ = 0.035 and $Slope_{M3}$ = 0.895). Although separate incorporation of HER2 or CD44 into the reference models did not improve model performance, incorporating HER2 and CD44 simultaneously yielded substantially improved overall performance (R^2_{M10} = 0.221, R^2_{M11} = 0.270, and R^2_{M12} = 0.225) and discrimination (AUC_{M10} = 0.759, AUC_{M11} = 0.857, and AUC_{M12} = 0.816). However, the calibration plots (Fig. 3) show that model 10 underestimated the probability of achieving a γpCR in low-risk patients, while models 10–12 slightly overestimated the probability in high-risk patients.

Discussion

The value of this study is the combination of biological tumor markers from pre-treatment tumor biopsies and intratumoral ^{18}F-FDG spatial distribution capturing information from the entire tumor burden. The improvement of the reference prediction models by the simultaneous addition of HER2 and CD44 demonstrates the value of this concept. Prediction model 11, composed of ^{18}F-FDG PET–based radiomic features (Geary's C measure and long-run low gray-level emphasis) and biological tumor markers (HER2 and CD44), was the preferred prediction model as it showed the highest observed level of overall performance and discrimination. The model had good ability to differentiate between γpCR and

Table 3 Performance of prediction models with and without biological tumor markers

Model	AIC	R^2	Brier	AUC	DS	Int	Slope
M1	94.7	0.140	0.158	0.657	0.083	0.086	0.817
M2	97.4	0.103	0.163	0.654	0.058	0.047	0.826
M3	92.3	0.173	0.151	0.685	0.105	0.035	0.895
M4 = M1 + HER2	90.9	0.133	0.153	0.700	0.075	0.041	0.894
M5 = M2 + HER2	92.3	0.115	0.155	0.694	0.063	0.033	0.880
M6 = M3 + HER2	89.1	0.162	0.147	0.700	0.094	0.030	0.894
M7 = M1 + CD44	45.1	0.127	0.163	0.739	0.046	0.072	0.758
M8 = M2 + CD44	47.3	0.087	0.174	0.748	0.017	0.064	0.701
M9 = M3 + CD44	46.5	0.114	0.175	0.737	0.016	0.080	0.643
M10 = M1 + HER2 + CD44	40.5	0.221	0.146	0.759	0.106	0.069	0.763
M11 = M2 + HER2 + CD44	41.2	0.270	0.135	0.857	0.134	0.036	0.834
M12 = M3 + HER2 + CD44	42.0	0.225	0.150	0.816	0.073	0.060	0.724

Abbreviations: *M1*, clinical reference model; *M2*, radiomic reference model; *M3*, clinico-radiomic reference model; *HER2*, human epidermal growth factor receptor 2; *CD44*, cluster of differentiation 44; *AIC*, Akaike Information Criterion; R^2, Nagelkerke R^2; *Brier*, Brier score; *AUC*, area under the receiver operating characteristic; *DS*, discrimination slope; *Int*, intercept

non-γpCR, but in some cases it was less useful to predict the actual probability of γpCR due to its fairly well calibration.

We found that HER2 activation was associated with a lower probability of achieving a γpCR. This is consistent with studies which showed the association between overexpression of HER2 with poor survival and therapy response [17, 20, 23]. Another study showed that enrichment of postoperative HER2 levels in surgical resection material compared to pre-treatment biopsies was associated with poor pathologic response [35]. In our study, immunohistochemical activation of CD44 was related to a higher probability of achieving γpCR. CD44 is a cell-surface transmembrane glycoprotein that has been observed to be present in cancer stem cells of esophageal cancer, a subpopulation of cells with the capacity of self-renewal. In vitro, the combination of $CD44^+/CD24^-$ subpopulation of esophageal cancer cells has shown to be more resistant to NCRT [36]. Although the reference prediction models showed similar performance when HER2 and CD44 were added separately, adding both variables simultaneously yielded improved performances. This might be explained by the biological interaction between HER2 and CD44, which leads to upregulation of CXCR4 in gastroesophageal cancer and may enhance NCRT resistance [25, 26]. Contradictory results have been reported regarding the association of HER2 and CD44 expression and T-stage [37, 38]. In this study, we did not observe any association; and therefore, these variables may contribute to the prediction model as independent variables. This was confirmed by the fact that model performance improved when HER2 (model 4) and CD44 (model 7) were separately added to the clinical reference model (model 1), which was constructed based on histology and clinical T-stage. According to the general guidelines, HER2 status should only be verified in patients with advanced esophageal adenocarcinoma who are potential candidates for HER2-targeting [22]. HER2 and CD44 status assessment is currently

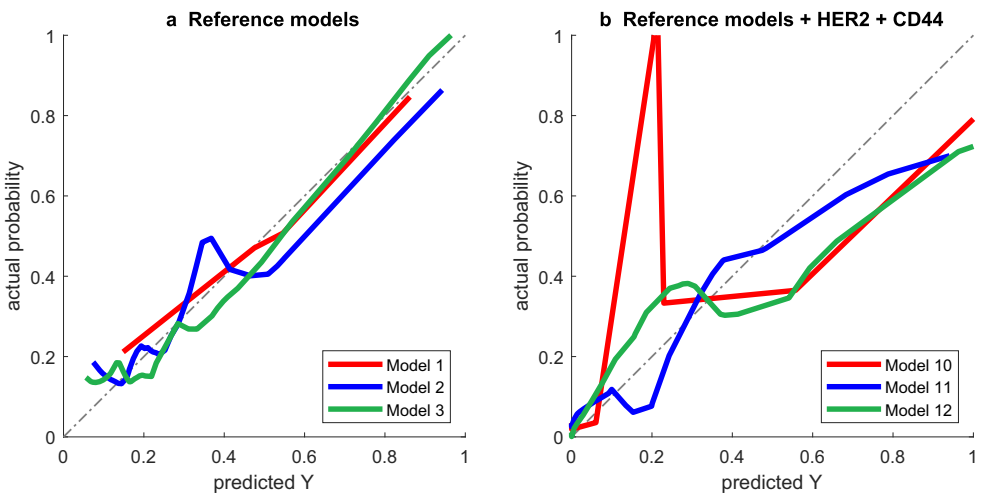

Fig. 3 Calibration plots of reference prediction models 1, 2, and 3, without (**a**) and with (**b**) HER2 and CD44 incorporated

not part of the pre-treatment staging. However, the findings of this study supported by literature suggest a potential value of adding pre-treatment HER2 and CD44 status in NCRT response prediction in locally advanced esophageal cancer.

Although adding HER2 and CD44 complements our reference prediction models, it should be noted that some of the constructed models were based on a relatively small number of patients (n = 43), which increases the risk of model overfitting. The reference prediction models were slightly better calibrated than the preferred prediction model, but had substantially lower overall performance and discrimination. However, well-calibrated models which have poor discrimination do not have any clinical value. Vice versa, the usability of the model is only limited if the model is poorly calibrated at the clinically chosen threshold to omit surgery. To prevent optimistic performance estimates, all measures were corrected for optimism by internal validation. However, as even internal validation may not be indicative of the final model's performance in future settings, the model should be validated in an independent validation cohort. Moreover, the reproducibility of immunohistochemical evaluation and radiomic feature extraction should be established before it can be safely implemented and generalized in a clinical setting. Immunohistochemical protocols may differ among institutes and results depend on the level of experience of the researcher. In this study, inter-rater variation was minimized by the independent evaluation of two researchers. Discordant cases were validated by an experienced pathologist. Moreover, only a few studies have reported on measurement errors (i.e., reliability, reproducibility, or repeatability) of radiomics [39–41]. Although further research is warranted, initial results show that most radiomic features are sensitive to numerous confounding factors including image acquisition, reconstruction protocols, or delineation methods. Standardization in the radiomics extraction workflow therefore remains essential to achieve routine clinical adoption.

This study indicates that HER2 and CD44 in the initial work-up could be useful biological tumor markers in predicting γpCR to NCRT in esophageal cancer. As genomic and ^{18}F-FDG PET–based radiomic features may yield independent complementary information, integration of these multimodal features may improve the performance of prediction models.

Funding The authors state that this work has not received any funding.

Compliance with Ethical Standards

Guarantor The scientific guarantor of this publication is Prof. Dr. J.T.M. Plukker (last author).

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors (Johannes G.M. Burgerhof) has significant statistical expertise.

Informed consent Only if the study is on human subjects: Written informed consent was waived by the Institutional Review Board.

Ethical approval This retrospective study was granted by the Local Institutional Review Board.

Study subjects or cohorts overlap This cohort has not been previously reported.

Methodology
- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. van Hagen P, Hulshof MC, van Lanschot JJ et al (2012) Preoperative chemoradiotherapy for esophageal or junctional cancer. N Engl J Med 366:2074–2084. https://doi.org/10.1056/NEJMoa1112088
2. Kwee RM (2010) Prediction of tumor response to neoadjuvant therapy in patients with esophageal cancer with use of ^{18}F FDG PET: a systematic review. Radiology 254:707–717. https://doi.org/10.1148/radiol.09091324
3. Dexter DL, Leith JT (1986) Tumor heterogeneity and drug resistance. J Clin Oncol 4:244–257. https://doi.org/10.1200/JCO.1986.4.2.244
4. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A (2015) Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res 21:249–257. https://doi.org/10.1158/1078-0432.CCR-14-0990
5. Harris AL (2002) Hypoxia–a key regulatory factor in tumour growth. Nat Rev Cancer 2:38–47. https://doi.org/10.1038/nrc704
6. Yang Z, He B, Zhuang X et al (2019) CT-Based radiomic signatures for prediction of pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant chemoradiotherapy. J Radiat Res 60:538–545. https://doi.org/10.1093/jrr/rrz027
7. Hou Z, Ren W, Li S et al (2017) Radiomic analysis in contrast-enhanced CT: predict treatment response to chemoradiotherapy in esophageal carcinoma. Oncotarget 8:104444–104454. https://doi.org/10.18632/oncotarget.22304
8. Jin X, Zheng X, Chen D et al (2019) Prediction of response after chemoradiation for esophageal cancer using a combination of dosimetry and CT radiomics. Eur Radiol 29:6080–6088. https://doi.org/10.1007/s00330-019-06193-w
9. van Rossum PS, Fried DV, Zhang L et al (2016) The incremental value of subjective and quantitative assessment of ^{18}F-FDG PET for the prediction of pathologic complete response to preoperative chemoradiotherapy in esophageal cancer. J Nucl Med 57:691–700. https://doi.org/10.2967/jnumed.115.163766
10. Beukinga RJ, Hulshoff JB, Mul VEM et al (2018) Prediction of response to neoadjuvant chemotherapy and radiation therapy with

baseline and restaging ^{18}F-FDG PET imaging biomarkers in patients with esophageal cancer. Radiology 287:983–992. https://doi.org/10.1148/radiol.2018172229
11. Tixier F, Le Rest CC, Hatt M et al (2011) Intratumor heterogeneity characterized by textural features on baseline ^{18}F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med 52:369–378. https://doi.org/10.2967/jnumed.110.082404
12. Tan S, Kligerman S, Chen W et al (2013) Spatial-temporal [^{18}F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. Int J Radiat Oncol Biol Phys 85:1375–1382. https://doi.org/10.1016/j.ijrobp.2012.10.017
13. Yip SS, Coroller TP, Sanford NN, Mamon H, Aerts HJ, Berbeco RI (2016) Relationship between the temporal changes in positron-emission-tomography-imaging-based textural features and pathologic response and survival in esophageal cancer patients. Front Oncol 6:72–82. https://doi.org/10.3389/fonc.2016.00072
14. Beukinga RJ, Hulshoff JB, van Dijk LV et al (2017) Predicting response to neoadjuvant chemoradiotherapy in esophageal cancer with textural features derived from pretreatment ^{18}F-FDG PET/CT imaging. J Nucl Med 58:723–729. https://doi.org/10.2967/jnumed.116.180299
15. Nakajo M, Jinguji M, Nakabeppu Y et al (2017) Texture analysis of ^{18}F-FDG PET/CT to predict tumour response and prognosis of patients with esophageal cancer treated by chemoradiotherapy. Eur J Nucl Med Mol Imaging 44:206–214. https://doi.org/10.1007/s00259-016-3506-2
16. Wang D, Plukker JT, Coppes RP (2017) Cancer stem cells with increased metastatic potential as a therapeutic target for esophageal cancer. Semin Cancer Biol. https://doi.org/10.1016/j.semcancer.2017.03.010
17. Bollschweiler E, Hölscher AH, Schmidt M, Warnecke-Eberz U (2015) Neoadjuvant treatment for advanced esophageal cancer: response assessment before surgery and how to predict response to chemoradiation before starting treatment. Chin J Cancer Res 27:221–230. https://doi.org/10.3978/j.issn.1000-9604.2015.04.04
18. Tao CJ, Lin G, Xu YP, Mao WM (2015) Predicting the response of neoadjuvant therapy for patients with esophageal carcinoma: an in-depth literature review. J Cancer 6:1179–1186. https://doi.org/10.7150/jca.12346
19. Vallböhmer D, Brabender J, Grimminger P, Schröder W, Hölscher AH (2011) Predicting response to neoadjuvant therapy in esophageal cancer. Expert Rev Anticancer Ther 11:1449–1455. https://doi.org/10.1586/era.11.126
20. Gowryshankar A, Nagaraja V, Eslick GD (2014) HER2 status in Barrett's esophagus & esophageal cancer: a meta analysis. J Gastrointest Oncol 5:25–35. https://doi.org/10.3978/j.issn.2078-6891.2013.039
21. Miller S, Hung M (1995) Regulation of her2/neu gene-expression (review). Oncol Rep 2:497–503. https://doi.org/10.3892/or.2.4.497
22. Bartley AN, Washington MK, Colasacco C et al (2017) HER2 testing and clinical decision making in gastroesophageal adenocarcinoma: Guideline from the College of American Pathologists, American Society for Clinical Pathology, and the American Society of Clinical Oncology. J Clin Oncol 35:446–464. https://doi.org/10.1200/JCO.2016.69.4836
23. Prins MJ, Ruurda JP, van Diest PJ, van Hillegersberg R, Ten Kate FJ (2013) The significance of the HER-2 status in esophageal adenocarcinoma for survival: an immunohistochemical and an in situ hybridization study. Ann Oncol 24:1290–1297. https://doi.org/10.1093/annonc/mds640
24. Wang D, Nagle PW, Wang HH et al (2019) Hedgehog pathway as a potential intervention target in esophageal cancer. Cancers 11:821. https://doi.org/10.3390/cancers11060821
25. Gros SJ, Kurschat N, Drenckhan A et al (2012) Involvement of CXCR4 chemokine receptor in metastastic HER2-positive esophageal cancer. PLoS One 7:e47287. https://doi.org/10.1371/journal.pone.0047287
26. Bao W, Fu HJ, Xie QS et al (2011) HER2 interacts with CD44 to up-regulate CXCR4 via epigenetic silencing of microRNA-139 in gastric cancer cells. Gastroenterology 141:2076–2087.e6. https://doi.org/10.1053/j.gastro.2011.08.050
27. Rice TW, Blackstone EH, Rusch VW (2010) 7th edition of the AJCC cancer staging manual: esophagus and esophagogastric junction. Ann Surg Oncol 17:1721–1724. https://doi.org/10.1245/s10434-010-1024-1
28. Mandard AM, Dalibard F, Mandard JC et al (1994) Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. clinicopathologic correlations. Cancer 73:2680–2686. https://doi.org/10.1002/1097-0142(19940601)73:11<2680::aid-cncr2820731105>3.0.CO;2-C
29. Boellaard R, Delgado-Bolton R, Oyen WJG et al (2015) FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging 42:328–354. https://doi.org/10.1007/s00259-014-2961-x
30. Zwanenburg A, Vallières M, Abdalah MA et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology:328–338. https://doi.org/10.1148/radiol.2020191145
31. Hofmann M, Stoss O, Shi D et al (2008) Assessment of a HER2 scoring system for gastric cancer: results from a validation study. Histopathology 52:797–805. https://doi.org/10.1111/j.1365-2559.2008.03028.x
32. Tubbs RR, Pettay JD, Roche PC, Stoler MH, Jenkins RB, Grogan TM (2001) Discrepancies in clinical laboratory testing of eligibility for trastuzumab therapy: apparent immunohistochemical false-positives do not get the message. J Clin Oncol 19:2714–2721. https://doi.org/10.1200/JCO.2001.19.10.2714
33. Honing J, Pavlov KV, Mul VE et al (2015) CD44, SHH and SOX2 as novel biomarkers in esophageal cancer patients treated with neoadjuvant chemoradiotherapy. Radiother Oncol 117:152–158. https://doi.org/10.1016/j.radonc.2015.08.031
34. Steyerberg EW, Vickers AJ, Cook NR et al (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21:128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2
35. Vallböhmer D, Kuhn E, Warnecke-Eberz U et al (2008) Failure in downregulation of intratumoral survivin expression following neoadjuvant chemoradiation in esophageal cancer. Pharmacogenomics 9:681–690. https://doi.org/10.2217/14622416.9.6.681
36. Smit JK, Faber H, Niemantsverdriet M et al (2013) Prediction of response to radiotherapy in the treatment of esophageal cancer using stem cell markers. Radiother Oncol 107:434–441. https://doi.org/10.1016/j.radonc.2013.03.027
37. Liang JW, Zhang JJ, Zhang T, Zheng ZC (2014) Clinicopathological and prognostic significance of HER2 overexpression in gastric cancer: a meta-analysis of the literature. Tumour Biol 35:4849–4858. https://doi.org/10.1007/s13277-014-1636-3
38. Fang M, Wu J, Lai X et al (2016) CD44 and CD44v6 are correlated with gastric cancer progression and poor patient prognosis: evidence from 42 studies. Cell Physiol Biochem 40:567–578. https://doi.org/10.1159/000452570
39. Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A (2017) The impact of image reconstruction settings on ^{18}F-FDG PET radiomic features: multi-scanner phantom and patient studies. Eur Radiol 27:4498–4509. https://doi.org/10.1007/s00330-017-4859-z
40. van Velden FH, Kramer GM, Frings V et al (2016) Repeatability of radiomic features in non-small-cell lung cancer [^{18}F]FDG-PET/CT studies: impact of reconstruction and delineation. Mol Imaging Biol 18:788–795. https://doi.org/10.1007/s11307-016-0940-2
41. Pfaehler E, Beukinga RJ, de Jong JR et al (2019) Repeatability of ^{18}F-FDG PET radiomic features: a phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method. Med Phys 46:665–678. https://doi.org/10.1002/mp.13322