

Development and Validation of a Gene Signature for Patients with Head and Neck Carcinomas Treated by Postoperative Radio(chemo)therapy



Stefan Schmidt^{1,2,3}, Annett Linge^{1,4,5,6}, Alex Zwanenburg^{1,3,6,7}, Stefan Leger^{1,4}, Fabian Lohaus^{1,4,5,6}, Constanze Krenn⁴, Steffen Appold^{5,6}, Volker Gudziol^{6,8}, Alexander Nowak^{6,9}, Cläre von Neubeck^{1,4}, Inge Tinhofer^{10,11}, Volker Budach^{10,11}, Ali Sak^{12,13}, Martin Stuschke^{12,13}, Panagiotis Balcermpas^{14,15}, Claus Rödel^{14,15}, Hatice Bunea^{16,17}, Anca-Ligia Grosu^{16,17}, Amir Abdollahi^{18,19,20,21,22}, Jürgen Debus^{18,19,20,21,23}, Ute Ganswindt^{24,25,26}, Claus Belka^{24,25,26}, Steffi Pigorsch^{24,27}, Stephanie E. Combs^{24,27,28}, David Mönnich^{29,30}, Daniel Zips^{29,30}, Gustavo B. Baretton^{1,6,31,32}, Frank Buchholz^{1,33}, Michael Baumann^{1,2,4,5,6,7}, Mechthild Krause^{1,2,4,5,6}, and Steffen Löck^{1,3,5} for the DTKK-ROG

Abstract

Purpose: The aim of this study was to identify and independently validate a novel gene signature predicting locoregional tumor control (LRC) for treatment individualization of patients with locally advanced HPV-negative head and neck squamous cell carcinomas (HNSCC) who are treated with postoperative radio(chemo)therapy (PORT-C).

Experimental Design: Gene expression analyses were performed using NanoString technology on a multicenter training cohort of 130 patients and an independent validation cohort of 121 patients. The analyzed gene set was composed of genes with a previously reported association with radio(chemo)sensitivity or resistance to radio(chemo)therapy. Gene selection and model building were performed comparing several machine-learning algorithms.

Results: We identified a 7-gene signature consisting of the three individual genes *HILPDA*, *CD24*, *TCF3*, and one metagene combining the highly correlated genes *SERPINE1*, *INHBA*, *P4HA2*,

and *ACTN1*. The 7-gene signature was used, in combination with clinical parameters, to fit a multivariable Cox model to the training data (concordance index, $ci = 0.82$), which was successfully validated ($ci = 0.71$). The signature showed improved performance compared with clinical parameters alone ($ci = 0.66$) and with a previously published model including hypoxia-associated genes and cancer stem cell markers ($ci = 0.65$). It was used to stratify patients into groups with low and high risk of recurrence, leading to significant differences in LRC in training and validation ($P < 0.001$).

Conclusions: We have identified and validated the first hypothesis-based gene signature for HPV-negative HNSCC treated by PORT-C including genes related to several radiobiological aspects. A prospective validation is planned in an ongoing prospective clinical trial before potential application in clinical trials for patient stratification. *Clin Cancer Res*; 24(6); 1364–74. ©2018 AACR.

¹German Cancer Consortium (DKTK), Partner Site Dresden, and German Cancer Research Center (DKFZ), Heidelberg, Germany. ²Helmholtz-Zentrum Dresden – Rossendorf, Institute of Radiooncology – OncoRay, Germany. ³OncoRay – National Center for Radiation Research in Oncology, Department of Biostatistics and Modelling in Radiation Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden – Rossendorf, Dresden, Germany. ⁴OncoRay – National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden – Rossendorf, Dresden, Germany. ⁵Department of Radiotherapy and Radiation Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Dresden, Technische Universität Dresden, Dresden, Germany. ⁶National Center for Tumor Diseases (NCT), Partner Site Dresden, Germany; German Cancer Research Center (DKFZ), Heidelberg, Germany; Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany; and Helmholtz Association / Helmholtz-Zentrum Dresden – Rossendorf (HZDR), Dresden, Germany. ⁷German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁸Department of Otorhinolaryngology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. ⁹Department of Oral and Maxillofacial Surgery, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. ¹⁰German Cancer Consortium (DKTK), Partner Site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany.

¹¹Department of Radiooncology and Radiotherapy, Charité University Medicine, Berlin, Germany. ¹²German Cancer Consortium (DKTK), Partner Site Essen, and German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹³Department of Radiotherapy, Medical Faculty, University of Duisburg-Essen, Essen, Germany. ¹⁴German Cancer Consortium (DKTK), Partner Site Frankfurt, and German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁵Department of Radiotherapy and Oncology, Goethe-University Frankfurt, Frankfurt, Germany. ¹⁶German Cancer Consortium (DKTK), Partner Site Freiburg, and German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁷Department of Radiation Oncology, Medical Center, Medical Faculty, University of Freiburg, Freiburg, Germany. ¹⁸German Cancer Consortium (DKTK), Partner Site Heidelberg, and German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁹Heidelberg Institute of Radiation Oncology (HIRO), National Center for Radiation Research in Oncology (NCRO), University of Heidelberg Medical School and German Cancer Research Center (DKFZ), Germany. ²⁰Heidelberg Ion Therapy Center (HIT), Department of Radiation Oncology, University of Heidelberg Medical School, Heidelberg, Germany. ²¹National Center for Tumor Diseases (NCT), University of Heidelberg Medical School and German Cancer Research Center (DKFZ), Heidelberg, Germany. ²²Translational Radiation Oncology, University of Heidelberg Medical School and German Cancer Research Center (DKFZ), Heidelberg, Germany. ²³Clinical Cooperation Unit Radiation Oncology, University of Heidelberg Medical School and German Cancer Research Center (DKFZ), Heidelberg, Germany. ²⁴German Cancer Consortium (DKTK), Partner Site Munich,

Translational Relevance

Patients with HPV-positive, locally advanced head and neck squamous cell carcinomas (HNSCC) show a very good loco-regional tumor control (LRC) after postoperative radiochemotherapy (PORT-C) and are therefore candidates for trials on treatment deescalation to reduce toxicity. For patients with HPV-negative HNSCC, additional biomarkers are urgently needed to identify subgroups of patients, (i) who are unlikely to respond to standard PORT-C and may benefit from treatment escalation, or (ii) who will likely not develop locoregional recurrences. We developed and independently validated a 7-gene signature prognostic for LRC of HPV-negative tumors, which is based on several radiobiological parameters or mechanisms. The prognostic performance of this radiobiology-based signature combined with clinical parameters was higher than that of a model containing hypoxia-associated genes and CSC markers only. After additional prospective validation, the 7-gene signature may be applied in clinical trials for patient stratification.

Introduction

Head and neck cancer is the sixth most frequently occurring tumor entity worldwide (1) with an overall 5-year survival rate of about 50% (2). Patients with resectable, locally advanced head and neck squamous cell carcinomas (HNSCC) who are at high risk for tumor recurrence are being routinely treated with postoperative radiochemotherapy (PORT-C). According to the results of three randomized clinical trials, concurrent chemotherapy leads to improved locoregional tumor control (LRC) and prolonged overall survival (OS) compared with postoperative radiotherapy (PORT) alone (3–5). Within the last years, radiotherapy of locally advanced HNSCC has been further improved through the development of new treatment techniques, such as intensity-modulated radiotherapy. Despite the increase in treatment efficacy, patients show a very heterogeneous treatment response. Therefore, the consideration of the individual tumor biology by appropriate biomarkers in addition to well-established clinical parameters may further improve patient stratification for treatment escalation or deescalation strategies.

Besides the consumption of alcohol and tobacco as well-known risk factors for the development of HNSCC, infection with the human papillomavirus (HPV) has been identified as another independent parameter. Also, the incidence of HPV infection in

HNSCC has been increasing within the last decade (6). Preclinical and clinical studies have shown that HPV-positive HNSCCs are more radiosensitive than HPV-negative tumors (7, 8). To investigate the impact of HPV in patients who receive PORT-C and to identify additional biomarkers for patient selection, a retrospective, multicenter study of the German Cancer Consortium Radiation Oncology Group (DKTK-ROG) was conducted (9–13). For this cohort, we have shown that patients with HPV16 DNA-positive tumors have superior LRC and OS compared with patients with HPV-negative tumors (9). In particular, 98% of the HPV-positive and only 80% of the HPV-negative oropharyngeal tumors were locoregionally controlled. For patients with HPV DNA-negative tumors, additional biomarkers are urgently needed to identify subgroups of patients, who are unlikely to respond to PORT-C and may benefit from treatment escalation, or who are not anticipated to develop locoregional recurrences.

Tumor hypoxia has been shown to be correlated with increased radioresistance (14). For patients with locally advanced HNSCC, pretreatment hypoxia was significantly associated with low tumor control and OS after primary radio(chemo)therapy compared with patients with highly oxygenated tumors (15, 16). Several hypoxia gene classifiers have been developed in the last decade to assess hypoxia or hypoxia-related changes on the transcriptional level using routinely taken pretreatment biopsies (17, 18). We have recently shown their prognostic validity for patients at high risk of loco-regional failure receiving PORT-C (12). The association of hypoxia and LRC after PORT-C is unexpected as the gross tumor has been removed and subsequently remaining tumor cells are very unlikely to differ in hypoxia (12). This suggests that hypoxia impacts LRC not only by a direct biochemical effect on cellular radioresistance but also by other radiobiological mechanisms (12, 19). Recent studies reported that hypoxia as an external factor also favors increased radioresistance of cancer stem cells (CSC) and invasive tumor growth (reviewed in refs. 20, 21), and CSCs are known to play a major role in radioresistance and tumor recurrence (reviewed in ref. 22). The putative CSC markers *CD44*, *SLC3A2*, and *MET* were shown to be prognostic for LRC in patients who received PORT-C (12). The combined application of hypoxia-associated gene panels and CSC markers further improved patient stratification regarding their risk of locoregional treatment failure (23, 24), which was independently validated (25).

A gene signature including additional radiobiological aspects may predict patient outcome with even higher accuracy. In the literature, the number of gene panels for stratification of patients with HNSCC is steadily growing. However, to the best of our knowledge, they have been developed for patients who received primary radiotherapy (26, 27) or have not been linked to a specific

and German Cancer Research Center (DKFZ), Heidelberg, Germany. ²⁵Department of Radiotherapy and Radiation Oncology, University Hospital, Ludwig-Maximilians-Universität, Munich, Germany. ²⁶Clinical Cooperation Group Personalized Radiotherapy in Head and Neck Cancer, Helmholtz Zentrum Munich, Neuherberg, Germany. ²⁷Department of Radiation Oncology, Technische Universität München, Munich, Germany. ²⁸Department of Radiation Sciences (DRS), Institut für Innovative Radiotherapie (IRT), Helmholtz Zentrum Munich, Neuherberg, Germany. ²⁹German Cancer Consortium (DKTK), Partner Site Tübingen, and German Cancer Research Center (DKFZ), Heidelberg, Germany. ³⁰Department of Radiation Oncology, Faculty of Medicine and University Hospital Tübingen, Eberhard Karls Universität Tübingen, Tübingen, Germany. ³¹Institute of Pathology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. ³²Tumour- and Normal Tissue Bank, University Cancer Centre (UCC), University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. ³³University Cancer

Centre (UCC), Medical Systems Biology, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

S. Schmidt and A. Linge contributed equally to this article.

Corresponding Author: Mechthild Krause, OncoRay – National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Fetscherstr. 74, PF 41, Dresden 01307, Germany. Phone: 4935-1458-5441; Fax: 4935-1458-5716; E-mail: mechthild.krause@uniklinikum-dresden.de

doi: 10.1158/1078-0432.CCR-17-2345

©2018 American Association for Cancer Research.

treatment (28–30). Gene signatures prognostic for the response of patients with locally advanced HNSCC to PORT-C covering a broad spectrum of radiobiological aspects are still missing.

Therefore, the major aim of this study was to develop and validate a gene signature and corresponding statistical model for patient stratification beyond HPV infection status to improve the risk assessment for patients with locally advanced HNSCC who receive PORT-C. For the development of this gene signature, a gene set was composed in-house using a hypothesis-driven approach. The gene set incorporated genes that cover many radiobiologically important aspects such as DNA repair, cell cycle, epithelial-mesenchymal transition, CSC markers, hypoxia, proliferation, invasion, metastasis, as well as genes that were reported to be involved in cisplatin resistance. The signature was developed and independently validated on two large patient cohorts for the primary endpoint LRC and the secondary endpoints OS and freedom from distant metastases (DM). To find the optimal results, internal validation methods were applied, and several statistical methods were compared, including advanced machine learning techniques.

Materials and Methods

Patients

Two different cohorts of patients with locally advanced HNSCC were being considered for this study. The training cohort consisted of 221 patients who were treated with PORT-C between 2004 and 2012 within the 9 partner sites of the DTK-ROG. Inclusion criteria, data collection, handling and analyses of biomaterial were previously described (9, 12). Briefly, all patients received curatively intended cisplatin-based PORT-C according to standard protocols with a minimum follow-up of 24 months and presented with a tumor stage pT4 and/or >3 positive lymph nodes and/or positive microscopic resection margins and/or extracapsular spread. The validation cohort consisted of 152 patients who were enrolled by the following criteria: not included in the previous DTK-ROG training cohort, histologically proven HNSCC, treatment between 1999 and 2006 with PORT or PORT-C according to standard radiotherapy protocols with curative intention (25).

Preparation of biomaterials and biomarker analyses

Formalin-fixed paraffin-embedded (FFPE) blocks of the primary tumor specimens (removed by surgery) were first subjected to hematoxylin and eosin staining to histologically confirm the presence of squamous cell carcinoma. Afterwards, they were processed under standardized procedures for biomarker investigations. DNA extraction and PCR-array based analyses of HPV status have been performed as described previously (9). Briefly, genomic DNA was extracted from 5- μ m FFPE sections using the QIAamp DNA FFPE tissue kit (Qiagen). HPV DNA analyses including genotyping were performed using the LCD-Array HPV 3.5 kit (CHIPRON GmbH) according to the manufacturer's instructions.

For both cohorts, gene expression analyses were performed consecutively using nanoString elements technology (nanoString Technologies, Seattle, WA, USA) as described in (12, 25). Briefly, total RNA as well as reporter and capture probes specific to the genes of interest were mixed and incubated at 62°C for 22 hours. Samples were then kept at 4 °C for a maximum of 18 hours and subjected to the nCounter system. Raw counts were logarithmized and then normalized by subtracting the mean of the log-transformed counts of the reference genes *ACTR3*, *B2M*, *GNB2L1*, *NDFIP1*, *POLR2A*, *RPL11*, and *RPL37A*. Because of insufficient

tumor material or too low RNA yield, some of the samples had to be omitted from the analysis. In the training and validation cohort, NanoString and HPV analyses could be performed for 195 and 142 samples, respectively. The expression levels of 178 genes were evaluated by NanoString analyses for both cohorts. The genes were selected by a literature search on a hypothesis-driven basis. Genes were included that have previously been reported to be associated with sensitivity or resistance to radio(chemo)therapy, that is, genes involved in proliferation, invasion, and metastasis; tumor hypoxia-associated genes, genes encoding for putative CSC markers, and DNA repair as well as genes that have been associated with cisplatin resistance (see Supplementary Table S1).

Study design

The aim of this study was to develop and validate a gene signature for patient stratification beyond HPV infection status to further improve the risk assessment for patients with locally advanced HNSCC who receive PORT-C. Therefore, only patients with HPV16 DNA-negative tumors and available NanoString gene expressions were included ($N = 130/221$ training, $N = 121/152$ validation). The study design is presented in Fig. 1. Prognostic models including the following parameters should be compared: the identified gene signature alone, clinical parameters alone, and the combination of clinical parameters and the identified gene signature.

Statistics and clinical endpoints

The primary endpoint was locoregional tumor control (LRC). Secondary endpoints were freedom from distant metastases (DM) and overall survival (OS). All endpoints were calculated from the first day of radiotherapy to the date of event or censoring. Death was considered a competing risk for locoregional recurrence and DM, while locoregional recurrence and DM did not cause censoring. Survival curves were estimated by the Kaplan–Meier method and compared by log-rank tests. Differences between the training and validation cohort were evaluated by Mann–Whitney U tests for continuous variables and by χ^2 tests for categorical variables. Descriptive analyses and the described statistical tests were performed using SPSS 23 (IBM Corporation). A statistical framework was developed to identify gene signatures and corresponding

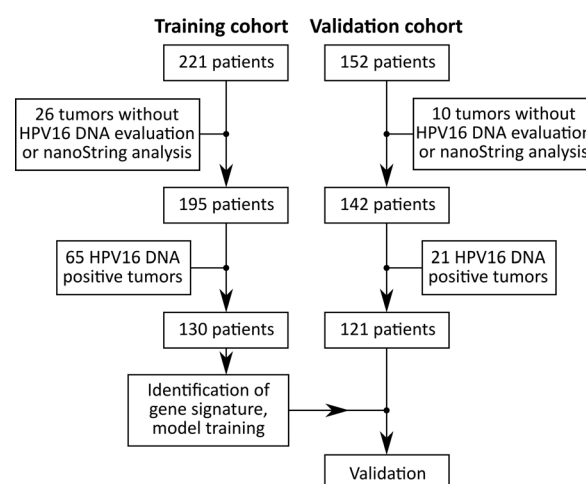
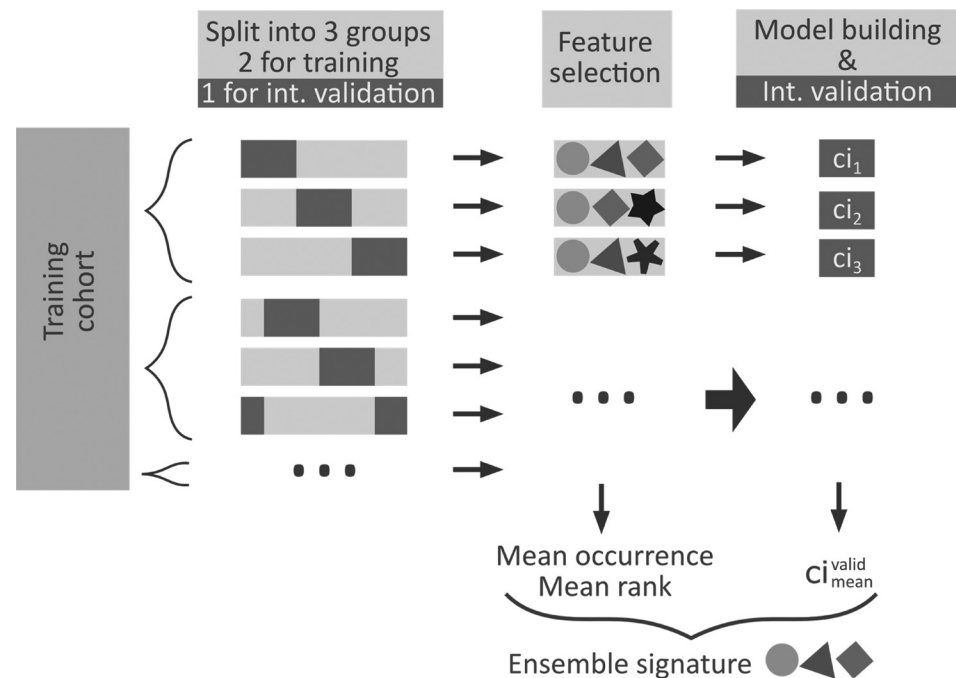


Figure 1.
Study design.

Figure 2.

Cross-validation scheme for identifying the ensemble gene signature. The training cohort was randomly split into 3 equal parts. Each part was used for internal validation and the remaining patients for internal training. This was repeated 333 times. Feature selection was performed on each internal training sample, and a prognostic model was trained using the selected genes. This model was subsequently internally validated. Finally, the occurrence and importance of the genes as well as the validation ci of all cross-validation experiments were used to define the ensemble gene signature.



prognostic models to optimally and robustly predict the primary and secondary endpoints. This framework is described in the Supplementary Materials in detail. To evaluate the prognostic performance of the developed models, the concordance index (ci) was calculated (31). While $ci = 0.5$ is obtained for a noninformative model, $ci = 1.0$ represents a perfectly predicting model. To compare the performance between nested multivariable Cox models, the likelihood ratio test was applied. The framework to determine gene signatures and corresponding prognostic models was implemented in R Statistics version 3.3.2 (R Foundation for Statistical Computing, Vienna; refs. 32 and 33) and Python (Python Software Foundation. Python Language Reference, version 2.7). An overview of the used programs and packages is given in Supplementary Table S2. For all analyses, two-sided tests were performed and P values below 0.05 were considered as statistically significant.

Statistical framework to identify gene signatures and perform model predictions

The statistical framework to identify gene signatures consists of four main steps, which are described in detail in Supplementary sections S1–S3: (i) the gene expression data are preprocessed. Genes are removed from analysis, if their median expression is below twice the median negative control in the training cohort. The expression of each gene is z-transformed to mean 0 and SD 1 on the training cohort, which is favorable for most machine-learning algorithms. The gene expressions of the validation cohort are transformed on the basis of the means and SDs of the training cohort; (ii) a feature selection method is applied to select the most relevant genes using internal 3-fold cross validation on the training cohort (repeated 333 times). The genes are combined to an ensemble signature based on their frequency of occurrence and their importance (Fig. 2). The resulting signature is then used to build prognostic models on 1,000 bootstrap samples of the training cohort to predict the considered outcome. Several feature selection methods, prognostic models, and different signature sizes (1–10) are compared and the best signature is chosen using

the out of the bag data of the bootstrap samples; (iii) to increase the robustness of the signature, genes which are highly correlated to one of the signature genes in the training cohort are combined with this gene to create a new metagene (median expression of the highly correlated genes). The resulting metagene replaces the original gene within the gene signature; (iv) finally, the model is validated using the independent validation cohort. The 95% confidence interval (CI) of the ci is estimated from 1,000 bootstrap samples of the validation cohort. Finally, the validation is declared successful if the 95% CI does not contain 0.5.

Data and materials availability

The final models and the raw genomic data used for creating the models are available upon request.

Results

Patient cohorts

In this retrospective study, a multicenter training cohort of 130 patients and an independent, monocenter validation cohort of 121 patients with HPV16 DNA-negative locally advanced HNSCC were available for the development of a gene signature to predict the clinical endpoints LRC, OS, and DM. Patient data, treatment parameters, and tumor characteristics of both patient cohorts were published previously (9, 34) and are summarized in Table 1. Patients in the validation cohort were treated with PORT ($n = 90$) or PORT-C ($n = 31$), while all patients of the training cohort received PORT-C as the standard treatment. The training cohort included 44.6% patients with oropharyngeal and 37.7% patients with oral cavity carcinomas. In the validation cohort, 21.5% of the patients have been diagnosed with oropharyngeal and 62.0% with oral cavity carcinomas. Patients in the validation cohort showed lower LRC (statistical trend) and OS, while the incidence of DM was not significantly different. Actuarial rates of LRC, freedom from DM, and OS two years after radiotherapy for the training and validation cohort were 83.8%

Table 1. Patient characteristics for the training and validation cohort

Characteristics	Training cohort (2004-2011)		Validation cohort (1999-2006)		P
	Median (range)		Median (range)		
Follow-up (months)	57.4 (11.5-94.5) ^a		62.1 (24.7-153.0) ^a		<0.001 ^b
Age (years)	56.5 (32.0-74.0)		52.3 (36.3-70.6)		0.005
Dose (Gy)	64.0 (56.0-68.0)		64.0 (60.0-66.0)		0.006
	Number of pts	%	Number of pts	%	
Gender					
Male/female	101/29	77.6/22.3	105/16	86.8/13.2	0.061
ECE status					
No/yes/unknown	62/68/0	47.7/52.3	82/39/0	67.8/32.2/0	0.001
Localization					
Oropharynx/oral cavity/Hypopharynx/Larynx	58/49/ 23/0	44.6/37.7/ 17.7/0	26/75/ 13/7	21.5/62.0/ 10.7/5.8	<0.001
Grading					
1/2/3/unknown	4/84/ 42/0	3.1/64.6/ 32.3/0	3/67/ 51/0	2.5/55.4/ 42.1/0	0.27
Chemotherapy					
Yes/no	130/0	100/0	31/90	25.6/74.4	<0.001
Locoregional recurrences	26	20.0	35	28.9	0.096 ^b
Distant metastases	31	23.8	29	24.0	0.72 ^b
Deaths	54	41.5	73	60.3	0.042 ^b

^a95% CI.^bLog-rank test.

versus 75.0% ($P = 0.096$), 79.0% versus 82.8% ($P = 0.72$), and 76.4% versus 64.9% ($P = 0.042$), respectively.

7-Gene signature predicts LRC for HPV16 DNA-negative tumors

To identify a prognostic gene signature for the primary endpoint LRC, the four steps (i)–(iv) of the statistical framework outlined in Materials and Methods were performed.

During the preprocessing step (i), the genes *FGFR2*, *PROM1*, and *TAF7L* were removed from the analysis. The mean validation ci from the 3-fold internal cross validation of step (ii) ranged between 0.57 and 0.68 and was similar between different feature selection methods and statistical models (see Supplementary Fig. S1 for signature size 4). An ensemble gene signature was

determined for each combination of feature selection method and prognostic model, as described in Supplementary Section S4. The performance of these signatures was evaluated using 1,000 bootstrap samples of the whole training cohort (see Fig. 3). The highest mean ci of 0.78 was obtained for a signature, which contained the genes *SERPINE1*, *CD24*, *HILPDA*, and *TCF3*. For the final prediction model, Cox regression was chosen, as it is the most simple of the well performing models. Signature size 4 was chosen based on the mean ci and the signature score of the genes (Supplementary Section S4; Supplementary Fig. S2). The signature score was highest for *SERPINE1*, followed by *CD24*, *HILPDA*, and *TCF3*, which showed a similar score (see Supplementary Fig. S3). To improve the robustness of the identified 4-gene ensemble

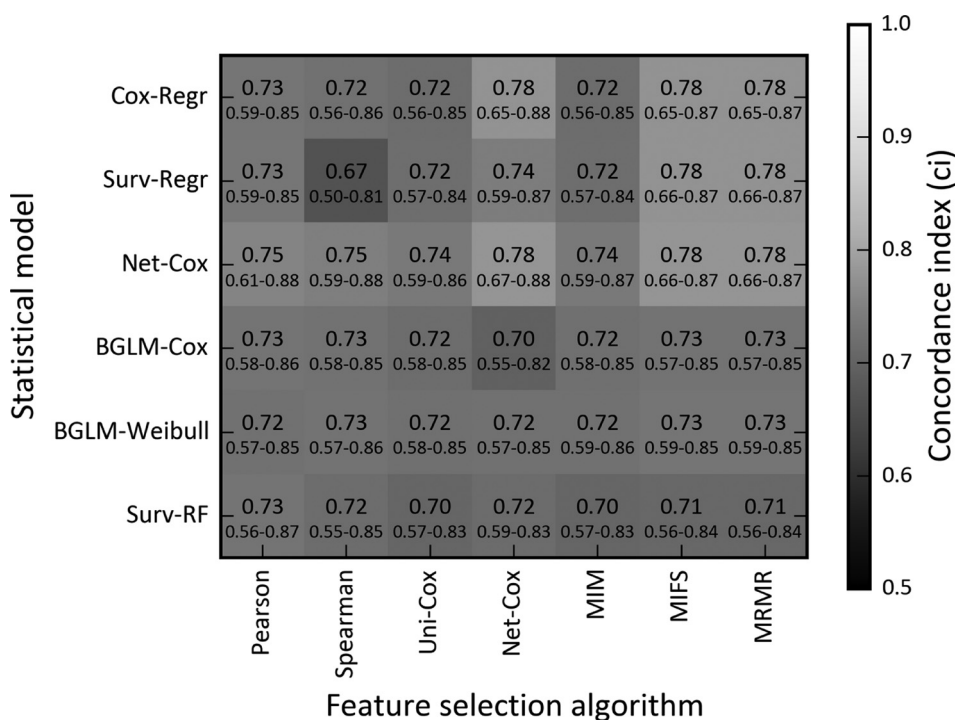


Figure 3. Performance of ensemble gene signatures for locoregional tumor control on the training cohort. For each combination of feature selection algorithm and statistical model, the mean out-of-the-bag (oob) validation ci of the training cohort and its 95% CI are shown. Performance for the endpoint locoregional tumor control was estimated using 1,000 bootstrap samples of the entire training cohort with signature size 4.

Table 2. Multivariable Cox regression of locoregional tumor control

Parameter	HR (95% CI)	P	ci Training (95% CI)	ci Validation (95% CI)	ci Validation, chemotherapy (95% CI)
7-Genes signature					
Metagene from <i>SERPINE1</i> , <i>INHBA</i> , <i>ACTN1</i> , and <i>P4HA2</i>	2.13 (1.18–3.88)	0.012			
<i>HILPDA</i>	1.48 (1.00–2.18)	0.049			
<i>CD24</i>	0.71 (0.48–1.04)	0.072			
<i>TCF3</i>	0.54 (0.32–0.88)	0.017	0.81 (0.75–0.88)	0.69 (0.60–0.77)	0.69 (0.39–0.87)
Clinical parameters					
ECE status	1.26 (0.57–2.82)	0.57			
Localization oral cavity	2.07 (0.95–4.56)	0.069	0.61 (0.53–0.74)	0.66 (0.57–0.74)	0.65 (0.30–0.84)
7-genes signature and clinical parameters					
Metagene from <i>SERPINE1</i> , <i>INHBA</i> , <i>ACTN1</i> , and <i>P4HA2</i>	1.98 (1.09–3.83)	0.026			
<i>HILPDA</i>	1.52 (1.02–2.26)	0.041			
<i>CD24</i>	0.69 (0.46–1.05)	0.083			
<i>TCF3</i>	0.55 (0.32–0.94)	0.031			
ECE status	1.40 (0.62–3.24)	0.43			
Localization oral cavity	1.27 (0.51–3.19)	0.61	0.82 (0.77–0.88)	0.71 (0.62–0.78)	0.72 (0.43–0.90)
Improvement of combined model compared to			dLL	Degrees of freedom	P
7-Genes signature only			1.26	2	0.53
Clinical parameters only			24.19	4	<0.001

NOTE: Three multivariable Cox regression models were built using the training cohort: a model consisting of only the 7-genes signature (top), a model consisting only of the clinical ECE status and tumor localization (center), and a model combining both the 7-genes signature and clinical parameters (bottom). HRs are given with their 95% CIs and the corresponding *P* values. For each model, the concordance index (ci) is given for the training and validation cohort as well as for the patients of the validation cohort who received concurrent chemotherapy. Its 95% CI is determined from 1,000 bootstrap samples of the respective cohort. The improvement of the combined model, including the 7-genes signature and the clinical parameters, compared with the 7-genes signature and clinical parameters alone is shown (bottom) based on the difference in log-likelihood (dLL).

signature, it was extended by other highly correlated genes in step (iii), as described in Supplementary Section S4. *INHBA*, *ACTN1*, and *P4HA2* were found to be highly correlated with *SERPINE1*, while for *CD24*, *TCF3*, and *HILPDA* no additional correlated genes were found (Supplementary Table S3). Thus, our final 7-genes signature for LRC consisted of the genes *SERPINE1*, *INHBA*, *ACTN1*, *P4HA2*, *CD24*, *TCF3*, and *HILPDA*. For evaluation, the median of the z-transformed, reference-gene normalized expression of *SERPINE1*, *INHBA*, *ACTN1*, and *P4HA2* was considered as a new metagene variable (Supplementary Section S4). The whole training cohort was used to fit the final Cox regression model, leading to a training ci of 0.81 (95% CI, 0.75–0.88). The resulting model parameters are shown in Table 2. In the last step (iv), the validation of the final model was performed on the independent validation cohort. A validation ci of 0.69 (0.60–0.77) was obtained, which represents a successful validation of the gene signature for the endpoint LRC.

Patient stratification into groups of low and high risk of recurrence was performed for the final Cox model depending on the risk score, which is given by the linear predictor of the model. The optimal cut-off was chosen based on the training cohort at a risk score of 0.10. This cutoff led to the highest fraction of patient stratifications with a significant difference in LRC, based on 1,000 bootstrap samples of the training cohort (992/1,000). Patients in the high risk group showed significantly lower LRC than patients in the low risk group, both for the training ($P < 0.001$) and the validation cohort ($P = 0.001$; see Supplementary Fig. S4).

Inclusion of clinical parameters to the 7-genes LRC signature

For the training cohort, it was shown that the established clinical parameters tumor localization and ECE status were significantly correlated with LRC or the secondary endpoints (9, 12). Using only these two parameters in a multivariable Cox model resulted in a lower performance [training: ci = 0.61 (0.53–0.74), validation: ci = 0.66 (0.57–0.74)] compared with the 7-genes signature. Finally, a

multivariable Cox model including both, the clinical parameters ECE status and tumor localization (oral cavity vs. others) as well as the 7-genes signature, increased the training ci to 0.82 (0.77–0.89) and the validation ci to 0.71 (0.62–0.78; see Table 2). While in training the clinical Cox model was significantly improved by adding the 7-genes signature ($P < 0.001$), adding the clinical parameters to the 7-genes signature resulted in only small improvements ($P = 0.53$). The difference in validation ci was not statistically significant. An additional validation was performed using only those patients who received concurrent chemotherapy, leading to similar results (validation ci = 0.72).

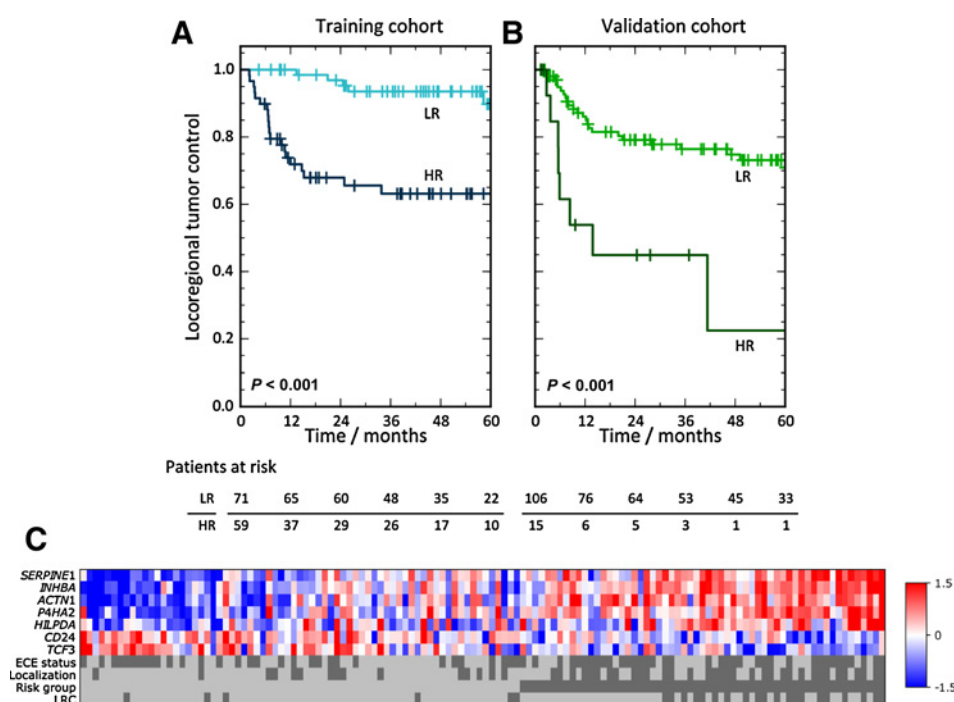
The extended model was used to stratify the patients into two risk groups (cutoff = 0.37), leading to highly significant differences in LRC for the training ($P < 0.001$) and the validation cohort ($P < 0.001$). The corresponding Kaplan–Meier curves are presented in Fig. 4 together with a heatmap of the signature for the training cohort (see also Supplementary Fig. S5).

Comparison with models based on CSC markers and hypoxia classifiers

In a previous study, it was shown that the expression of CSC markers and hypoxia-related genes were prognostic in patients with locally advanced HPV16 DNA-negative HNSCC, who were treated by PORT-C (12). These results were validated in ref. 25. Here, the performance of these models was compared with the 7-genes signature. While in ref. 25, the best performing model, consisting of ECE status, tumor localization, $CD44 > 0.2$, and the 15-genes hypoxia classifier (17), showed a validation ci of 0.65 (0.54–0.74), the identified 7-genes signature combined with the clinical parameters led to a validation ci of 0.71 (0.62–0.78).

7-Genes signature predicts for secondary endpoints

As secondary endpoints, overall survival (OS) and freedom from distant metastases (DM) were considered. The 7-genes signature determined for LRC, combined with the clinical features

**Figure 4.**

Patient stratification by the 7-gene signature and clinical parameters for locoregional tumor control.

Kaplan-Meier estimates of locoregional tumor control (LRC) are shown for the training cohort (**A**) and the validation cohort (**B**). Patients were stratified into a low-risk group (LR) or a high-risk group (HR) by the linear predictor of the multivariable Cox model which included the 7-gene signature and the clinical parameters ECE status and tumor localization. The cut-off risk score (0.37) was determined on the training cohort and applied to the validation cohort.

C, Heatmap of the 7-gene signature as well as ECE status (0, light; 1, dark), localization (oral cavity, dark; others, light), risk group (LR, light; HR, dark), and LRC during follow-up (yes, light; no, dark) for the training cohort.

ECE status and tumor localization, was trained and validated for OS and DM (Supplementary Tables S4 and S5). For OS, training and validation led to a ci of 0.71 (0.65–0.79) and 0.64 (0.57–0.70), respectively. For DM, the ci was 0.69 (0.64–0.80) for training and 0.63 (0.52–0.73) for validation. Cox models including only the clinical features ECE status and tumor localization led to a validation ci of 0.60 (0.54–0.66) for OS and of 0.61 (0.52–0.71) for DM, respectively. Hence, the 7-gene signature could improve the prognostic performance also for the secondary endpoints OS and DM compared with clinical parameters alone. Validation on the subgroup of patients receiving concurrent chemotherapy led to a higher ci for the 7-gene signature [OS: 0.72 (0.56–0.84), DM: 0.74 (0.52–0.90)]. Kaplan-Meier survival analyses are presented in Supplementary Figs. S6 and S7.

Discussion

The overall aim of this study was to identify and validate a gene signature for the stratification of patients with HPV-negative, locally advanced HNSCC who are treated by PORT-C based on the clinical endpoint LRC. We identified a 7-gene signature, which contains genes from an extended in-house gene set compared with previous work, which showed that patients with HPV-negative tumors could be further stratified by the expression of CSC markers and hypoxia-associated genes (12). In addition to the HPV infection status, CSC marker expression levels and tumor hypoxia-associated genes, this gene set included genes related to DNA repair, cell-cycle regulation, epithelial–mesenchymal transition, proliferation, or invasion. A statistical framework was developed with the objective of identifying a gene signature which accurately and robustly predicts the risk of locoregional failure. The framework contains data preprocessing, internal cross validation, signature selection, model building, and independent validation.

The identified 7-gene signature contained the genes *SERPINE1*, *INHBA*, *ACTN1*, and *P4HA2* (which were combined into a single metagene due to high mutual correlation) as well as the genes *CD24*, *TCF3*, and *HILPDA*. *SERPINE1* (also known as *PAI-1*), *HILPDA*, *INHBA*, and *P4HA2* are being induced by the hypoxia-inducible factor *HIF1* leading to extracellular matrix remodeling (35–37). *SERPINE1* plays a role in enhanced migration and cell proliferation as well as decreased cisplatin-induced apoptosis (38, 39). In a prospective clinical study including 190 patients, high expression of *SERPINE1* has been shown to be associated with poor local recurrence-free, progression-free, and cancer-specific survival (38). In a panel of head and neck xenograft tumors, *SERPINE1* expression levels were overexpressed prior to treatment mainly in hypoxic tumors (40). After fractionated irradiation, a correlation between *SERPINE1* expression levels and local tumor control was found *in vivo* (40). In addition, mild hypoxia has been shown to induce *SERPINE1* expression via the hypoxia-inducible factor *HIF1* (37). *SERPINE1* is functionally associated with *INHBA* (41) and *ACTN1* (42). In the hypoxia-associated signatures, *HILPDA* (also known as *HIG2*) and *P4HA2* have also been included (17, 43). *HILPDA* has been shown to promote proliferation and invasion (44). In the literature, conflicting data exist for *CD24*, which is expressed in different tumor entities such as breast cancer and cervical cancer and has shown to be associated with increased tumor growth and progression (45). *CD24* has also been shown to be involved in cisplatin resistance (46) and a shortened progression-free survival was observed for several tumor entities with higher expression (45, 47). In contrast, *CD24* overexpression has been shown to be correlated with better survival in patients with oral carcinoma (48). They further showed that *CD24*^{−/−} mice are able to develop progressive oral cancer. Lack of the surface protein *CD24* resulted in the expansion of a highly immunosuppressive *CD11b*⁺*Gr1*⁺ myeloid cell population leading to oral cancer progression. To date, very little is

known about the transcription factor 3 (*TCF3*) and its potential role in cancer. *TCF3* is a critical cell signaling molecule (49) and has been shown to promote cell migration and wound repair (50). In contrast, *TCF3* was found to be a cell-intrinsic inhibitor of pluripotent self-renewal through limiting the steady-state levels of self-renewal factors such as Oct-4, Sox2, and Nanog in mouse embryonic stem cells (51). Lack of *TCF3* leads to increasing levels of Nanog and other self-renewal genes, minimizing the response to differentiation stimuli (51). According to the factors of the final Cox model, overexpression of *SERPINE1* (as well as the highly correlated genes *INHBA*, *ACTN1*, and *P4HA2*) and *HILPDA* increased the risk for locoregional failure, which is in line with the literature (38, 39). In contrast, a high expression of *CD24* led to decreased risk of recurrence. For oral cavity cancer it has been shown that *CD24* dampens the functional expansion of myeloid-derived suppressor cells and gives rise to a more favorable prognosis as described above (48), which is in line with our findings. The final Cox model also predicted that a high expression of *TCF3* is related to improved LRC, which may be due to its role in the suppression of self-renewal genes (51). However, the functional role of *TCF3* in HNSCC needs to be explored in further mechanistic studies.

The identified 7-gene signature showed a good prognostic ability for the endpoint LRC on the validation cohort ($ci = 0.69$). When combined with the clinical parameters ECE status and tumor localization, its performance could be further improved ($ci = 0.71$). This indicates that the combination of well-established clinical parameters and prognostic biomarkers may lead to a more accurate prognosis than each of them alone. The model including only the clinical parameters showed the lowest validation performance ($ci = 0.66$). In the Cox model combining clinical parameters with the 7-gene signature, most signature genes were significantly associated with LRC. While this may be expected on the training cohort, the impact of the 7-gene signature in validation is less clear, as the relevant improvement in ci by 0.05 was not statistically significant. Evaluating the signature combined with HPV status for all patients increased the validation ci to 0.74, which is similar to or even higher than in other studies (30, 52).

The final Cox model showed a better performance on the training cohort ($ci=0.82$) than on the validation cohort ($ci=0.71$). This difference is expected, since the final Cox model is adjusted to the training cohort and potential overfitting might occur. In addition, the validation of the proposed 7-gene signature might be impeded by the significant differences between both patient cohorts. Patients in the validation cohort were clinically characterized by a higher percentage of prognostically favorable R0 resections of primary tumors and less lymph nodes with ECE. On the other hand, the validation cohort had a higher percentage of prognostically unfavorable oral cavity tumors, much less concurrent chemotherapy (31/121) than the training cohort and was treated with outdated radiation technologies (53). These negative prognostic factors outbalanced the positive ones resulting in worse outcome in terms of LRC ($P = 0.096$) and OS ($P = 0.042$; ref. 25). Lack of concurrent chemotherapy may impede the validation of genes related to cisplatin resistance. For the 7-gene signature, however, only *CD24* has been reported to be strongly involved in resistance to cisplatin (46), but also in other mechanisms (48).

On the basis of the final Cox model, a risk score was calculated for each patient, which allowed stratification into groups of low

and high risk of recurrence. However, mean gene expressions (Supplementary Table S6) as well as clinical parameters were significantly different between the training and validation cohort. These differences caused a shift in the risk score, such that the stratification cutoff, which was based on the training cohort, led to imbalanced patient risk groups for the validation cohort. While in training approximately 45% of the patients were stratified in the low-risk group and 55% in the high-risk group, for the validation cohort only about 12% of the patients were classified as high risk. Such imbalances may be caused by the differing tumor and treatment characteristics between the cohorts. In addition to clinical reasons, differences in gene expression might also be caused by several biomaterial-related factors such as storage time of FFPE material (3–18 years) or batch effects and stability of reagents and consumables (Supplementary Table S7). Renormalizing the validation data to the training data, as described in refs. 12 and 54, gives the same fraction of patients in the low- and high-risk group and similar LRC rates for both cohorts (Supplementary Fig. S5). However, to apply this renormalization method for individual patient prognosis within clinical trials, the inclusion of reference samples may be required, for which the expected gene expression levels are known. This methodology will be applied to the planned prospective validation of the 7-gene signature. In addition, the application of broadly available and cost-effective PCR-based methods may further improve biomarker stability.

In this study, several algorithms for gene selection and risk prediction were compared. Feature selection algorithms based on mutual information, such as MIFS and MRMR, typically led to a higher ci than simple univariable methods such as Pearson or Spearman correlations (Fig. 3). This behavior can be expected, as the more complex algorithms do not only account for the correlation of the gene expressions to outcome but also consider correlations between the selected genes. Therefore, each gene in the signature represents additional information, which increases the performance of the signature. The performance of prediction models, ranging from the well-known Cox model to complex random forests, was similar on the training cohort. Therefore, the performance of the signature was finally assessed by multivariable Cox regression, which allows easy interpretation. Most of the considered models require additional hyperparameters, such as the regularization parameters λ_1 and λ_2 for penalized Cox models or node size and node depth for random forests (see Supplementary Section S3). In an initial experiment, these parameters were chosen based on their default values given in the used software packages and then tuned by a grid search using 2-fold internal cross validation on the training cohort. The resulting parameters were applied in this study and are reported in Supplementary Table S8. While random forests did not outperform simple Cox regression in this study, this may not hold in other situations (55).

The presented 7-gene signature was identified for patients with HPV16 DNA-negative tumors and the primary endpoint LRC. However, it also improved the prognostic value of the clinical parameters for the secondary endpoint OS, while for DM, no significant difference was observed. In particular, for patients receiving concurrent chemotherapy, the validation performance of the 7-gene signature was improved by 10%. This may further enhance the clinical potential of this signature.

A limitation of this study might also be the limited number of genes contained in the initial gene set. Although this has been composed on a hypothesis-driven basis and comprehensive

literature search, it may not include all genes of radiobiological relevance. For example CD44, which has been shown to be a prognosticator for LRC in patients with locally advanced HNSCC who received PORT-C (12), had to be omitted from the NanoString analysis due to incorrect probe design. Since the set-up of our gene set, other genes have been shown to be prognostic for outcome in HNSCC. For example, TCGA analyses (56) suggested several genes, related to HPV status. Of these genes *CCND1*, *NOTCH1*, *YAP1*, and *SOX2* were found to overlap with our gene set. In the TCGA dataset, patients with *CCND1*-overexpressing tumors, who received surgery with or without postoperative radiochemotherapy showed worse prognosis. In our study, *CCND1* had no impact on the primary endpoint LRC ($P = 0.72$). Therefore, it was not selected in the gene signature. However, *CCND1* showed a significant correlation to the secondary endpoints OS and DM using univariable Cox regression for all 195 patients. For the subgroup of patients with HPV-negative tumors, *CCND1* neither correlated with OS nor with DM. This could be explained by the strong correlation of *CCND1* with the HPV status in our cohort. In contrast, *YAP1* was significantly associated with LRC in our study, but was rated only at rank 14 such that it was not included in the 7-gene signature. *NOTCH1* and *SOX2* were not related to LRC. Another example is PD-L1, which was strongly associated with local failure in HPV-negative HNSCC (13, 57), but not included in our gene set. To consider these novel developments and identify further biomarkers, whole transcriptome analyses supplemented by whole methylome analyses might be performed and potentially further improve patient stratification.

Currently, an adaptive clinical biomarker matrix trial is set-up within the DTK-ROG for dose escalation and deescalation in HNSCC. In the first stage, patients with HPV-positive tumors treated by PORT-C will receive a 10% lower radiation dose of the standard concurrent radiochemotherapy schedule. In the second stage, the 7-gene signature is one candidate biomarker for selecting patients with high-risk HPV-negative tumors for dose escalation. To reduce toxicities, especially at higher doses, proton therapy will be considered (58).

In conclusion, this study introduces a novel 7-gene signature predicting LRC for patients with locally advanced HNSCC treated by PORT-C. A prognostic Cox model was trained on a large multicenter patient cohort and independently validated. Although the validation cohort differed in many aspects from the training cohort, a successful validation was achieved, which indicates the robustness of the signature. Prospective validation of the signature is planned within an ongoing prospective clinical

trial of the DTK-ROG before regular application in clinical trials for patient stratification.

Disclosure of Potential Conflicts of Interest

V. Gudziol and V. Budach are consultant/advisory board members for Bristol-Myers Squibb. I. Tinhofer is a consultant/advisory board member for Merck Serono. C. Belka reports receiving commercial research grants from and is a consultant/advisory board member for Merck Darmstadt. S.E. Combs reports receiving speakers bureau honoraria from Accuray and Brainlab, and is a consultant/advisory board member for Bristol-Myers Squibb and Daiichi Sankyo. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: A. Linge, F. Lohaus, A. Abdollahi, M. Baumann, M. Krause, S. Löck

Development of methodology: S. Schmidt, A. Linge, A. Zwanenburg, F. Lohaus, S. Löck

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Linge, F. Lohaus, C. Krenn, S. Appold, V. Gudziol, A. Nowak, C. von Neubeck, I. Tinhofer, V. Budach, A. Sak, M. Stuschke, P. Balermipas, C. Rödel, H. Bunea, A.-L. Grosu, A. Abdollahi, J. Debus, U. Ganswindt, C. Belka, S. Pigorsch, S.E. Combs, D. Mönnich, D. Zips, G.B. Baretton

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S. Schmidt, A. Linge, A. Zwanenburg, S. Leger, C. Belka, M. Baumann, S. Löck

Writing, review, and/or revision of the manuscript: S. Schmidt, A. Linge, A. Zwanenburg, S. Leger, F. Lohaus, C. Krenn, S. Appold, V. Gudziol, A. Nowak, I. Tinhofer, V. Budach, P. Balermipas, C. Rödel, H. Bunea, A.L. Grosu, J. Debus, U. Ganswindt, C. Belka, S. Pigorsch, S.E. Combs, D. Mönnich, D. Zips, G.B. Baretton, M. Baumann, M. Krause, S. Löck

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Linge, A. Zwanenburg, J. Debus, G.B. Baretton, F. Buchholz, S. Löck

Study supervision: J. Debus, S.U. Pigorsch, M. Baumann, M. Krause, S. Löck

Acknowledgments

The study was financed by a Joint Funding Grant within the German Cancer Consortium (DKTK) awarded to the DTK-ROG (principal investigator: M. Baumann). The DTK is funded as one of the National German Health Centers by the Federal German Ministry of Education and Research. The authors gratefully acknowledge the excellent technical assistance by Mrs. Sigrid Balschukat, Mrs. Daniela Friede, and Mrs. Liane Stolz-Kieslich. The authors thank all pathologists, head and neck surgeons, and maxillofacial surgeons at the treatment centers who provided tumor material and data for this study.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 11, 2017; revised November 4, 2017; accepted December 29, 2017; published OnlineFirst January 3, 2018.

References

1. Duray A, Descamps G, Decaestecker C, Rummelink M, Sirtaine N, Lechien J, et al. Human papillomavirus DNA strongly correlates with a poorer prognosis in oral cavity carcinoma. *Laryngoscope* 2012;122:1558–65.
2. Leemans CR, Braakhuis BJM, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer* 2011;11:9–22.
3. Fietkau R, Lautenschläger C, Sauer R, Dunst J, Becker A, Baumann M, et al. Postoperative concurrent radiochemotherapy versus radiotherapy in high-risk SCCA of the head and neck: results of the German phase III trial ARO 96–3. *J Clin Oncol* 2006;24:5507–5507.
4. Cooper JS, Pajak TF, Forastiere AA, Jacobs J, Campbell BH, Saxman SB, et al. Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the Head and Neck. *N Engl J Med* 2004;350:1937–44.
5. Bernier J, Dommange C, Ozsahin M, Matuszewska K, Lefebvre J-L, Greiner RH, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced Head and Neck Cancer. *N Engl J Med* 2004;350:1945–52.
6. Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol* 2010;11:781–9.
7. Kimple RJ, Smith MA, Blitzer GC, Torres AD, Martin JA, Yang RZ, et al. Enhanced radiation sensitivity in HPV-positive head and neck cancer. *Cancer Res* 2013;73:4791–800.
8. Gupta AK, Lee JH, Wilke WW, Quon H, Smith G, Maity A, et al. Radiation response in two HPV-infected head-and-neck cancer cell lines in comparison to a non-HPV-Infected cell line and relationship to signaling through AKT. *Int J Radiat Oncol* 2009;74:928–33.

9. Lohaus F, Linge A, Tinhofer I, Budach V, Gliko E, Stuschke M, et al. HPV16 DNA status is a strong prognosticator of loco-regional control after post-operative radiochemotherapy of locally advanced oropharyngeal carcinoma: results from a multicentre explorative study of the German cancer consortium radiation oncology group. *Radiother Oncol* 2014;113:317–23.
10. Balcermpas P, Rödel F, Rödel C, Krause M, Linge A, Lohaus F, et al. CD8+ tumour-infiltrating lymphocytes in relation to HPV status and clinical outcome in patients with head and neck cancer after postoperative chemoradiotherapy: A multicentre study of the German cancer consortium radiation oncology group (DKTK-ROG). *Int J Cancer* 2016;138:171–81.
11. Tinhofer I, Budach V, Saki M, Korschak R, Niehr F, Jöhrens K, et al. Targeted next-generation sequencing of locally advanced squamous cell carcinomas of the head and neck reveals druggable targets for improving adjuvant chemoradiation. *Eur J Cancer* 2016;57:78–86.
12. Linge A, Löck S, Gudziol V, Nowak A, Lohaus F, von Neubeck C, et al. Low cancer stem cell marker expression and low hypoxia identify good prognosis subgroups in HPV(-) HNSCC after postoperative radiochemotherapy: a multicenter study of the DKTK-ROG. *Clin Cancer Res* 2016;22:2639–49.
13. Balcermpas P, Rödel F, Krause M, Linge A, Lohaus F, Baumann M, et al. The PD-1/PD-L1 axis and human papilloma virus in patients with head and neck cancer after adjuvant chemoradiotherapy: a multicenter study of the German cancer consortium radiation oncology group (DKTK-ROG). *Int J Cancer* 2017;141:594–603.
14. Overgaard J. Hypoxic radiosensitization: adored and ignored. *J Clin Oncol* 2007;25:4066–74.
15. Nordsmark M, Overgaard M, Overgaard J. Pretreatment oxygenation predicts radiation response in advanced squamous cell carcinoma of the head and neck. *Radiother Oncol* 1996;41:31–9.
16. Nordsmark M, Bentzen SM, Rudat V, Brizel D, Lartigau E, Stadler P, et al. Prognostic value of tumor oxygenation in 397 head and neck tumors after primary radiation therapy. An international multi-center study. *Radiother Oncol* 2005;77:18–24.
17. Toustrup K, Sørensen BS, Nordsmark M, Busk M, Wiuf C, Alsner J, et al. Development of a hypoxia gene expression classifier with predictive impact for hypoxic modification of radiotherapy in head and neck cancer. *Cancer Res* 2011;71:5923–31.
18. Buffa FM, Harris AL, West CM, Miller CJ. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia meta-gene. *Br J Cancer* 2010;102:428–35.
19. Yaromina A, Thames H, Zhou X, Hering S, Eicheler W, Dörfler A, et al. Radiobiological hypoxia, histological parameters of tumour microenvironment and local tumour control after fractionated irradiation. *Radiother Oncol* 2010;96:116–22.
20. Bütof R, Dubrovskaya A, Baumann M. Clinical perspectives of cancer stem cell research in radiation oncology. *Radiother Oncol* 2013;108:388–96.
21. Peitzsch C, Perrin R, Hill RP, Dubrovskaya A, Kurth I. Hypoxia as a biomarker for radioresistant cancer stem cells. *Int J Radiat Biol* 2014;90:636–52.
22. Baumann M, Krause M, Hill R. Exploring the role of cancer stem cells in radioresistance. *Nat Rev Cancer* 2008;8:545–54.
23. Yaromina A, Krause M, Baumann M. Individualization of cancer treatment from radiotherapy perspective. *Mol Oncol* 2012;6:211–21.
24. Wu Y, Posner MR, Schumaker LM, Nikitakis N, Goloubeva O, Tan M, et al. Novel biomarker panel predicts prognosis in human papillomavirus-negative oropharyngeal cancer. *Cancer* 2012;118:1811–7.
25. Linge A, Löck S, Krenn C, Appold S, Lohaus F, Nowak A, et al. Independent validation of the prognostic value of cancer stem cell marker expression and hypoxia-induced gene expression for patients with locally advanced HNSCC after postoperative radiotherapy. *Clin Transl Radiat Oncol* 2016;1:19–26.
26. De Cecco L, Nicolau M, Giannoccaro M, Daidone MG, Bossi P, Locati L, et al. Head and neck cancer subtypes with biological and clinical relevance: Meta-analysis of gene-expression data. *Oncotarget* 2015;6:9627–42.
27. van den Broek GB, Wildeman M, Rasch CRN, Armstrong N, Schuurings E, Begg AC, et al. Molecular markers predict outcome in squamous cell carcinoma of the head and neck after concomitant cisplatin-based chemoradiation. *Int J Cancer* 2009;124:2643–50.
28. Agada FO, Patmore H, Alhamarneh O, Stafford ND, Greenman J. Genetic profile of head and neck squamous cell carcinoma: clinical implications. *J Laryngol Otol* 2009;123:266–72.
29. Ginos MA, Page GP, Michalowicz BS, Patel KJ, Volker SE, Pambuccian SE, et al. Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck. *Cancer Res* 2004;64:55–63.
30. De Cecco L, Bossi P, Locati L, Canevari S, Licita L. Comprehensive gene expression meta-analysis of head and neck squamous cell carcinoma microarray data defines a robust survival predictor. *Ann Oncol* 2014;25:1628–35.
31. Harrell FE. Regression modeling strategies. 1st edition. New York, NY: Springer; 2001.
32. R Development Core Team. R: a language and environment for statistical computing [Internet]. Vienna, Austria; 2008. Available from: <http://www.r-project.org>.
33. RStudio Team. RStudio: integrated development environment for R [Internet]. Boston, MA; 2015. Available from: <http://www.rstudio.com/>.
34. Linge A, Lohaus F, Löck S, Nowak A, Gudziol V, Valentini C, et al. HPV status, cancer stem cell marker expression, hypoxia gene signatures and tumour volume identify good prognosis subgroups in patients with HNSCC after primary radiochemotherapy: a multicenter retrospective study of the DKTK-ROG. *Radiother Oncol* 2016;121:364–73.
35. Gilkes DM, Bajpai S, Chaturvedi P, Wirtz D, Semenza GL. Hypoxia-inducible Factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts. *J Biol Chem* 2013;288:10819–29.
36. Depoix CL, de Selliers I, Hubinont C, Debieve F. HIF1A and EPAS1 potentiate hypoxia-induced upregulation of inhibin alpha chain expression in human term cytotrophoblasts in vitro. *Mol Hum Reprod* 2017;23:199–209.
37. Kietzmann T, Roth U, Jungermann K. Induction of the plasminogen activator inhibitor-1 gene expression by mild hypoxia via a hypoxia response element binding the hypoxia-inducible factor-1 in rat hepatocytes. *Am Soc Hematol* 1999;94:4177–85.
38. Pavón MA, Arroyo-Solera I, Téllez-Gabriel M, León X, Virós D, López M, et al. Enhanced cell migration and apoptosis resistance may underlie the association between high SERPINE1 expression and poor outcome in head and neck carcinoma patients. *Oncotarget* 2015;6:29016–33.
39. Pavón MA, Arroyo-Solera I, Céspedes MV, Casanova I, León X, Mangués R. uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy. *Oncotarget* 2016;7:57351–66.
40. Bayer C, Schilling D, Hoetzel J, Egermann HP, Zips D, Yaromina A, et al. PAI-1 levels predict response to fractionated irradiation in 10 human squamous cell carcinoma lines of the head and neck. *Radiother Oncol* 2008;86:361–8.
41. Wang Y-H, Tian X, Liu O-S, Fang X-D, Quan H-Z, Xie S, et al. Gene profiling analysis for patients with oral verrucous carcinoma and oral squamous cell carcinoma. *Int J Clin Exp Med* 2014;7:1845–52.
42. Menicanin D, Bartold PM, Zannettino ACW, Gronthos S. Genomic profiling of mesenchymal stem cells. *Stem Cell Rev Reports* 2009;5:36–50.
43. Eustace A, Mani N, Span PN, Irlam JJ, Taylor J, Betts GNJ, et al. A 26-gene hypoxia signature predicts benefit from hypoxia-modifying therapy in laryngeal cancer but not bladder cancer. *Clin Cancer Res* 2013;19:4879–88.
44. Mao X, Wang C, Liu D, Zhang X, Wang L, Yan M, et al. Hypoxia upregulates HIG2 expression and contributes to bevacizumab resistance in glioblastoma. *Oncotarget* 2016;7:47808–20.
45. Lee J-H, Kim S-H, Lee E-S, Kim Y-S. CD24 overexpression in cancer development and progression: a meta-analysis. *Oncol Rep* 2009;22:1149–56.
46. Modur V, Joshi P, Nie D, Robbins KT, Khan AU, Rao K. CD24 expression may play a role as a predictive indicator and a modulator of cisplatin treatment response in head and neck squamous cellular carcinoma. *PLoS One* 2016;11:e0156651.
47. Smith SC. The metastasis-associated gene CD24 is regulated by Ral GTPase and is a mediator of cell proliferation and survival in human cancer. *Cancer Res* 2006;66:1917–22.
48. Fugle CW, Zhang Y, Hong F, Sun S, Westwater C, Rachidi S, et al. CD24 blunts oral squamous cancer development and dampens the functional expansion of myeloid-derived suppressor cells. *Oncoimmunology* 2016;5:e1226719.
49. Zhang Z, Filho MSS, Nör JE. The biology of head and neck cancer stem cells. *Oral Oncol* 2012;48:1–9.

50. Miao Q, Ku AT, Nishino Y, Howard JM, Rao AS, Shaver TM, et al. Tcf3 promotes cell migration and wound repair through regulation of lipocalin 2. *Nat Commun* 2014;5:4088.
51. Yi F, Pereira L, Merrill BJ. Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells* 2008;26:1951–60.
52. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.
53. Baumann M, Krause M, Overgaard J, Debus J, Bentzen SM, Daartz J, et al. Radiation oncology in the era of precision medicine. *Nat Rev Cancer* 2016;16:234–49.
54. Toustrup K, Sørensen BS, Metwally MAH, Tramm T, Mortensen LS, Overgaard J, et al. Validation of a 15-gene hypoxia classifier in head and neck cancer for prospective use in clinical trials. *Acta Oncol* 2016;55:1091–8.
55. Zheng G. Identification of candidate colon cancer biomarkers by applying a random forest approach on microarray data. *Oncol Rep* 2012; 28:1036–42.
56. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 2015;517: 576–82.
57. Skinner HD, Giri U, Yang LP, Kumar M, Liu Y, Story MD, et al. Integrative analysis identifies a novel AXL-PI3 kinase-PD-L1 signaling axis associated with radiation resistance in head and neck cancer. *Clin Cancer Res* 2017;23:2713–22.
58. Jakobi A, Bandurska-Luque A, Stützer K, Haase R, Löck S, Wack L-J, et al. Identification of patient benefit from proton therapy for advanced head and neck cancer patients based on individual and subgroup normal tissue complication probability analysis. *Int J Radiat Oncol Biol Phys* 2015; 92:1165–74.