


Quantitative proteomics identifies biomarkers to distinguish pulmonary from head and neck squamous cell carcinomas by immunohistochemistry

Annika Richter^{1†*}, Alexander Fichtner^{1†} , Jasmin Joost^{1†}, Philipp Brockmeyer², Philipp Kauffmann², Henning Schliephake², Alexander Hammerstein-Equord³, Stefan Kueffer¹, Henning Urlaub^{4,5}, Thomas Oellerich^{6,7}, Philipp Ströbel¹, Hanibal Bohnenberger^{1†} and Felix Bremmer¹

¹Institute of Pathology, University Medical Centre Göttingen, Göttingen, Germany

²Department of Oral and Maxillofacial Surgery, University Medical Centre Göttingen, Göttingen, Germany

³Department of Thoracic and Cardiovascular Surgery, University Medical Centre Göttingen, Göttingen, Germany

⁴Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

⁵Bioanalytics, Institute for Clinical Chemistry, University Medical Centre Göttingen, Göttingen, Germany

⁶Department of Medicine II, Haematology/Oncology, Goethe University, Frankfurt, Germany

⁷German Cancer Research Centre and German Cancer Consortium, Heidelberg, Germany

*Correspondence to: Annika Richter, Institute of Pathology, University Medical Centre, Robert-Koch-Str. 40, 37075 Göttingen, Germany.
E-mail: annika.richter1@med.uni-goettingen.de

†These authors contributed equally to this study.

Abstract

The differentiation between a pulmonary metastasis and a newly developed squamous cell carcinoma of the lung in patients with prior head and neck squamous cell carcinoma (HNSCC) is difficult due to a lack of biomarkers but is crucially important for the prognosis and therapy of the affected patient. By using high-resolution mass spectrometry in combination with stable isotope labelling by amino acids in cell culture, we identified 379 proteins that are differentially expressed in squamous cell carcinomas of the lung and the head and neck. Of those, CAV1, CAV2, LGALS1, LGALS7, CK19, and UGDH were tested by immunohistochemistry on 194 tissue samples (98 lung and 96 HNSCCs). The combination of CAV1 and LGALS7 was able to distinguish the origin of the squamous cell carcinoma with high accuracy (area under the curve 0.876). This biomarker panel was tested on a cohort of 12 clinically classified lung tumours of unknown origin after HNSCC. Nine of those tumours were immunohistochemically classifiable.

Keywords: SILAC-mass spectrometry; squamous cell carcinoma; head and neck; lung; metastasis; cell culture

Received 23 April 2021; Revised 23 July 2021; Accepted 1 September 2021

Introduction

Head and neck squamous cell carcinomas (HNSCCs) comprise a group of neoplasms developing in the nose, oral cavity, pharynx, and larynx [1] and make up more than 90% of neoplasms of the head and neck [2]. The long-term survival of patients with HNSCC treated with curative intent has not improved significantly over recent decades [2] due to lymph node metastasis, local tumour recurrence, and mainly because of distant metastasis (met-HNSCC), for example to the lung [3].

The differentiation between a pulmonary metastasis of a primary known HNSCC and a secondary primary tumour (SPT) of the lung after a prior HNSCC is difficult. More than two-thirds of SPTs are squamous cell carcinomas [4,5] and both tumours have similar patterns of genetic mutations and pathological molecular alterations [5–8]. No biomarker has yet been established for their differential diagnosis. Therefore, the differentiation is mainly based on radiological and clinical criteria [5,9,10].

However, the distinction between met-HNSCC and SPT is crucially important for further prognosis and the decision between a curative and palliative treatment

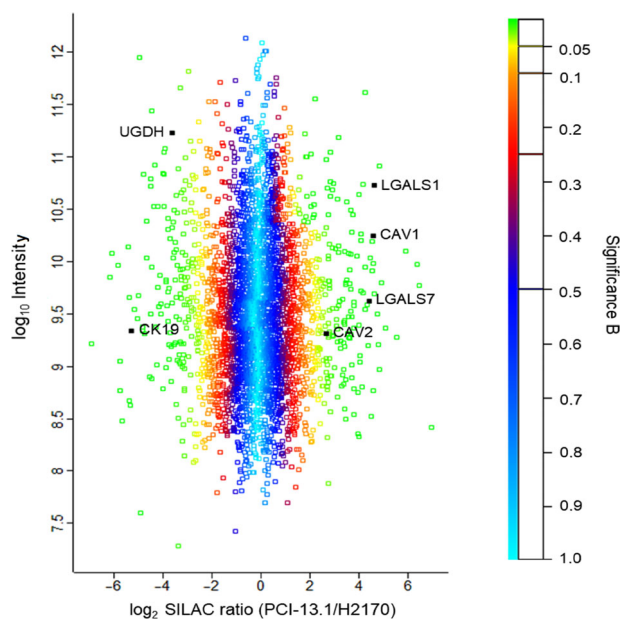


Figure 1. Analysis of the significance of the SILAC-based MS. Distribution of the \log_2 SILAC ratios (PCI-13.1/H2170) of all quantified proteins based on their relative expression in H2170 (SQCLC) and PCI-13.1 (HNSCC) depending on the respective \log_{10} intensity values. Significantly differentially expressed proteins are shown in green according to their significance. Proteins selected for immunohistochemistry are labelled.

regimen. We therefore aimed to identify, by high-resolution mass spectrometry (MS)-based proteomics, differentially expressed proteins that can be used as immunohistochemical markers to distinguish between squamous cell lung carcinoma (SQCLC) and HNSCC.

Methods

Proteomic analysis of squamous cell carcinoma cell lines

The human tumour cell line NCI-H2170 (CRL-5928; American Type Culture Collection, Manassas, VA,

USA) originates from a squamous cell carcinoma of the lung. The PCI-13.1 cell line (Department of Pathology, UPMC, PA, USA) derives from HPV-negative squamous cell carcinoma of the oropharynx. Labelling with stable isotopes in cell culture and proteomic comparison were conducted as described previously [8,11,12]. In brief, NCI-H2170 and PCI3.1 cells were cultured in RPMI 1640 medium supplemented with 10% dialysed foetal calf serum (Invitrogen, Waltham, MA, USA), 4 mM glutamine, antibiotics, 0.115 mM L-arginine- $^{13}\text{C}_6$ and 0.275 mM L-lysine-4,4,5,5- D_4 (Eurisotop, St-Aubin, France), or equimolar levels of the corresponding non-labelled (light) amino acids (Merck Millipore, Burlington, MA, USA) for at least 10 cell cycles. Labelled cells were lysed in 0.5% Nonidet P-40 buffer containing 50 mM Tris/HCl, pH 7.8, 150 mM NaCl, 1 mM Na_3VO_4 , 1 mM NaF, 0.2% lauryl maltoside, and protease inhibitors (Complete, Roche, Basel, Switzerland). Equal amounts of proteins of light-labelled NCI-H2170 were mixed with heavy-labelled PCI-13.1 and vice versa to obtain two biological replicates with different stable isotope labelling by amino acids in cell culture (SILAC). Afterwards, they were separated by 1D-PAGE (4–12% NuPAGE Bis-Tris Gel; Invitrogen) and stained with Coomassie brilliant blue. Next, the stained gel was separated in 23 slices and each one was reduced with 10 mM DTT for 55 min at 56 °C, alkylated with 55 mM iodoacetamide (IAA) for 20 min at 26 °C, and digested with modified trypsin (Promega, Madison, WI, USA) overnight at 37 °C. Resulting peptides were separated by a C18 precolumn (2.5 cm, 360 μm o.d., 100 μm i.d., Reprosil-Pur 120 Å, 5 μm , C18-AQ; Dr. Maisch GmbH, Ammerbuch, Germany) at a flow rate of 10 $\mu\text{l}/\text{min}$ and a C18 capillary column (20 cm, 360 μm o.d., 75 μm i.d., Reprosil-Pur 120 Å, 3 μm , C18-AQ; Dr. Maisch GmbH) at a flow rate of 300 nl/min, with a gradient of acetonitrile ranging from 5 to 35% in 0.1% formic acid for 90 min using an Proxeon nano LC coupled to an Q Exactive mass spectrometer (Thermo

Table 1. Proteins selected for immunohistochemistry with their \log_2 SILAC ratios (PCI-13.1/H2170), \log_{10} -transformed intensity values, and calculated *P* values. The original data are in supplementary material, Table S3.

| Protein | Gene | \log_2 SILAC ratio (PCI-13.1/H2170) | \log_{10} -intensity | <i>P</i> value |
|-----------------------------|--------|---------------------------------------|------------------------|----------------|
| Caveolin-1 | CAV1 | 2.30 | 10.247 | 4.50E-05 |
| Caveolin-2 | CAV2 | 1.32 | 9.311 | 3.19E-02 |
| Galectin-1 | LGALS1 | 2.32 | 10.726 | 4.14E-11 |
| Galectin-7 | LGALS7 | 2.20 | 9.624 | 6.00E-04 |
| Cytokeratin-19 | KRT19 | -2.64 | 9.33638 | 2.12E-05 |
| UDP-glucose 6-dehydrogenase | UGDH | -1.80 | 11.2308 | 6.63E-03 |

Table 2. Results of the marker candidates in the HPA and cohort 1.

| Marker | SILAC-MS (significantly stronger in) | IHC (HPA) (n) positive | | IHC (cohort 1) (n) positive | |
|--------|--------------------------------------|------------------------|-------|-----------------------------|-------|
| CAV1 | PCI-13.1 (HNSCC) | 3/3 | HNSCC | 6/6 | HNSCC |
| | | 1/5 | SQCLC | 0/6 | SQCLC |
| CAV2 | PCI-13.1 (HNSCC) | 3/3 | HNSCC | 6/6 | HNSCC |
| | | 1/4 | SQCLC | 1/6 | SQCLC |
| LGALS1 | PCI-13.1 (HNSCC) | 3/4 | HNSCC | 5/6 | HNSCC |
| | | 1/5 | SQCLC | 1/6 | SQCLC |
| LGALS7 | PCI-13.1 (HNSCC) | 3/3 | HNSCC | 6/6 | HNSCC |
| | | 1/4 | SQCLC | 1/6 | SQCLC |
| CK19 | H2170 (SQCLC) | 1/4 | HNSCC | 2/6 | HNSCC |
| | | 3/4 | SQCLC | 6/6 | SQCLC |
| UGDH | H2170 (SQCLC) | 1/3 | HNSCC | 2/6 | HNSCC |
| | | 5/6 | SQCLC | 5/6 | SQCLC |

The table shows the results of the SILAC-MS as well as those of the IHC in the HPA and cohort 1. The absolute number of positive cases and the total number of cases are given for each case. The proteins that were upregulated in HNSCC are highlighted in blue and the proteins that were downregulated are highlighted in grey.

HPA, Human Protein Atlas; IHC, immunohistochemistry; SILAC-MS, SILAC-based mass spectrometry.

Electron, Waltham, MA, USA). MS conditions were as follows: spray voltage, 1.8 kV; heated capillary temperature, 270 °C; and normalised collision energy,

28. The mass spectrometer automatically switched between MS and MS/MS acquisitions (data-dependent mode). Survey MS spectra were acquired

Table 3. Clinical and pathological data.

| | | SQCLC (n = 98) | HNSCC (n = 96) |
|-----------------------------|---------------|----------------|----------------|
| Age | Mean ± SD | 65.5 ± 8.4 | 62.1 ± 10.4 |
| | Age range | 43–81 | 24–83 |
| Sex | Male | 86 (87.8%) | 71 (74.0%) |
| | Female | 12 (12.2%) | 25 (26.0%) |
| Localisation | Oral cavity | – | 48 (50.0%) |
| | Pharynx | – | 25 (26.0%) |
| | Larynx | – | 23 (24.0%) |
| | Lung | 98 (100%) | – |
| pT stage | pT1 | 19 (19.4%) | 22 (22.9%) |
| | pT2 | 63 (64.3%) | 41 (42.7%) |
| | pT3 | 12 (12.2%) | 19 (19.8%) |
| | pT4 | 4 (4.1%) | 14 (14.6%) |
| pN stage | pN0 | 59 (60.2%) | 52 (54.2%) |
| | pN1 | 22 (22.4%) | 23 (24.0%) |
| | pN2 | 17 (17.3%) | 21 (21.9%) |
| pM stage | pM0 | 98 (100%) | 95 (99.0%) |
| | Uncertain pM1 | 0 (0%) | 1 (1.0%) |
| UICC stage (eighth edition) | I | 33 (33.7%) | 16 (16.7%) |
| | II | 43 (43.9%) | 26 (27.1%) |
| | III | 22 (22.4%) | 25 (26.0%) |
| | IV | 0 (0.0%) | 29 (30.2%) |
| Grade | G1 | 0 (0.0%) | 0 (0%) |
| | G2 | 70 (70.7%) | 87 (90.6%) |
| | G3 | 28 (29.3%) | 9 (9.4%) |
| p16 positive | | – | 12 (12.5%) |

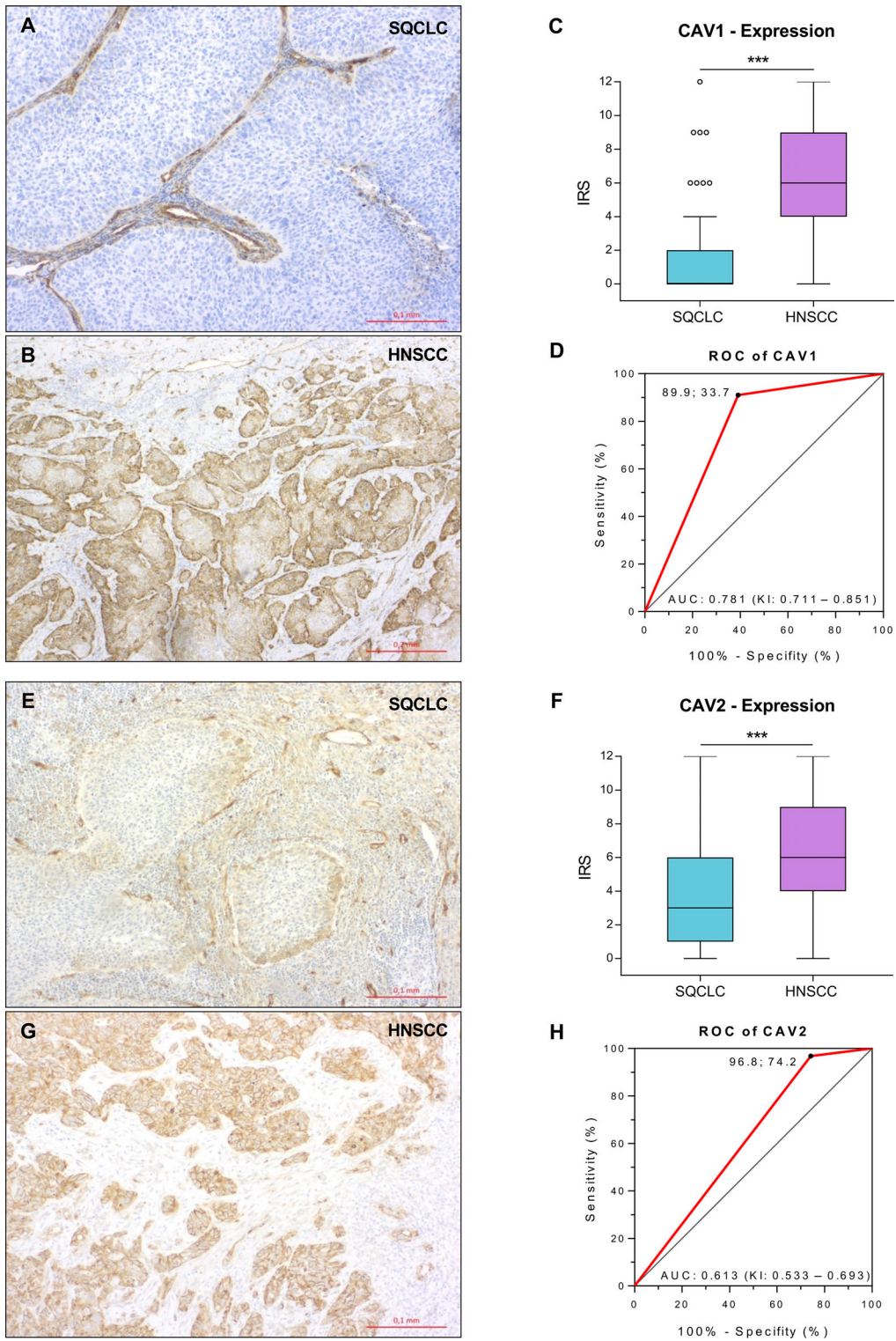


Figure 2. Legend on next page.

in the Orbitrap (m/z 350–1,600) with the resolution set to 70,000 at m/z 200. The 15 most intense ions were sequentially isolated for higher-energy collisional dissociation (HCD) MS/MS fragmentation and detection. Raw data were analysed with MaxQuant (version 1.3.0.5) using Uniprot human as a sequence database. Up to two missed cleavages of trypsin were allowed. Oxidised methionine was searched as variable modification and cysteine carbamidomethylation as fixed modification. The modifications corresponding to arginine and lysine labelled with heavy stable isotopes were handled as fixed modifications. The false positive rate was set to 1% at the peptide level, the false discovery rate was set to 1% at the protein level, and the minimum required peptide length was set to six amino acids.

Resulting data from MaxQuant analysis of the raw data were further analysed by Perseus (version 1.5.2.6; Max Planck Institute for Biochemistry, Martinsried, Germany). Ratios of intensity of heavy- and light-labelled proteins corresponding to either of the two cell lines were \log_2 -transformed and the medians of the two biological replicates were calculated. Intensity values were \log_{10} -transformed. Next, an outlier significance score for SILAC ratios depending on intensity values (significance B in Perseus, see citation for more details [13]) for every protein was calculated and fold discovery rate was corrected by Benjamini–Hochberg correction. The complete results of the statistical analysis with Perseus including the number of unique peptides can be found in supplementary material, Table S1.

Tissue samples

The patient samples were collected at the University Medical Centre in Göttingen, Germany. In total, 98 SQCLC, 96 HNSCC, and 12 lung tumours with squamous cell carcinoma histology of unknown origin after primary HNSCC were included in this study. The samples of the main cohort (98 SQCLC, 96 HNSCC) derived from oncological resections. Approval for using the human patient material in this study was

obtained from the Ethics Committee of the University Medical Centre Göttingen (vote no. 07/06/09, updated in April 2018). All procedures were conducted in accordance with the Declaration of Helsinki and institutional, state, and federal guidelines.

Immunohistochemistry

Immunohistochemical reactions were performed on 2- μm formalin-fixed and paraffin-embedded tissue sections, as described previously [14]. Antigen retrieval was carried out at 97 °C in citrate buffer (pH 6) or EDTA buffer (pH 9). The antibodies and dilutions used are listed in supplementary material, Table S2. The sections were incubated with a ready-to-use horseradish peroxidase-labelled secondary antibody at room temperature for 25 min (anti-rabbit/mouse, produced in goat; REAL EnVision Detection System; Dako, Agilent Technologies, Waldbronn, Germany). The substrate DAB + Chromogen system produces a brown end product, and is applied to visualise the target antigen (REAL DAB + Chromogen; Dako, Agilent Technologies). Tissue samples were counterstained with Mayer's haematoxylin (Dako, Agilent Technologies) for 8 min and analysed using light microscopy.

Two independent investigators evaluated all stained tissue sections by using an immunoreactivity staining score (IRS) as described previously [12]. The percentage of positively stained cells was first classified using a 0–4 scoring system: score 0 = 0% positive cells, score 1 = less than 10% positive cells, score 2 = 10–50% positive cells, score 3 = 51–80% positive cells, and score 4 = >80% positive cells. The intensity of staining was evaluated on a four-tiered scale (0 = negative, 1 = weak, 2 = intermediate, and 3 = strong). Afterwards, the scores of intensity and staining were multiplied and the mean value per patient was calculated, where 0–1 point was interpreted as negative, 2–3 as weakly positive, 4–6 as moderately positive, and 8–12 points as strongly positive. Therefore, 'immunohistochemically positive' tumours

Figure 2. Staining of CAV1 and CAV2 in SQCLC and HNSCC. (A) The SQCLC tumour cells show no or only a slight expression of CAV1 (total magnification $\times 100$). (B) The tumour cells of HNSCC show cytoplasmic and strong membranous expression of CAV1 (total magnification $\times 50$). (C) Box plot for CAV1 expression in SQCLC and HNSCC; the horizontal lines within the boxes represent the median IRS values, *** $p < 0.001$. (D) CAV1 had an AUC value of 0.781, a sensitivity of 89.9%, and a specificity of 66.3%. (E) The SQCLC tumour cells show no or only slight expression of CAV2 (total magnification $\times 100$). (F) Box plot for CAV2 expression in SQCLC and HNSCC; the horizontal lines within the boxes represent the median IRS values, *** $p < 0.001$. (G) The tumour cells of HNSCC show cytoplasmic and strong membranous expression of CAV2 (total magnification $\times 50$). (H) CAV2 had an AUC value of 0.613, a sensitivity of 96.8%, and a specificity of 25.8%.

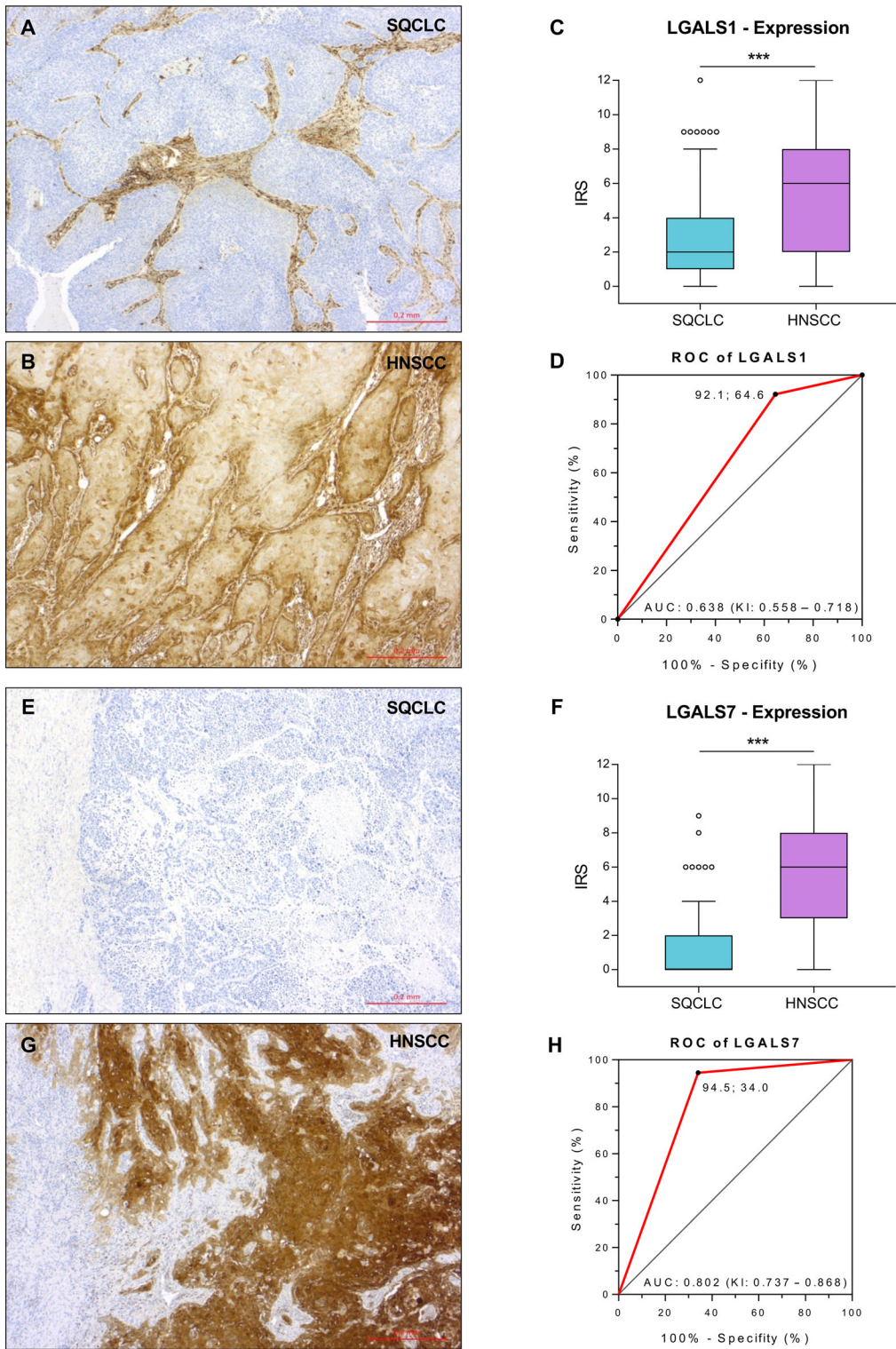


Figure 3. Legend on next page.

have an IRS ≥ 2 . Examples of the various IRSs are shown in supplementary material, Figure S1.

Statistical analysis

Differences of the obtained IRS between the different subtypes of squamous cell carcinomas were statistically evaluated using the Mann–Whitney *U*-test (GraphPad Software, San Diego, CA, USA). A *P* value of <0.05 was considered significant. Using receiver operating characteristic (ROC) curve analysis, we were able to determine the area under the curve (AUC) value and sensitivity and specificity of the antibodies. The cut-off of the ROC analyses was set at IRS ≥ 2 as this value was considered immunohistochemically positive.

Results

Quantitative proteomic comparison of pulmonary and HNSCC cell lines

The differential diagnosis of metachronous primary and metastatic squamous cell carcinoma in patients with HNSCC is crucial for clinical treatment decisions. However, diagnostic biomarkers are still not established in routine diagnostic pathology, partially due to overlapping genetic and morphological features. We therefore aimed to identify new proteomic biomarkers that can be translated to immunohistochemical support for the pathological diagnosis. We therefore compared the protein expression profile of the two human cell lines NCI-H2170 and PCI-13.1. NCI-H2170 originates from a squamous cell carcinoma of the lung and PCI-13.1 from an HPV-negative squamous cell carcinoma from the oropharynx. By using high-resolution MS in combination with SILAC, we were able to quantify 4,606 proteins and 379 of those were significantly differently expressed (Figure 1 and supplementary material, Tables S1 and S3).

We compared those proteins with the expression data of the Human Protein Atlas and chose six markers (two upregulated in the lung cancer cell line H2170

and four upregulated in the HNSCC cell line PCI-13.1) for immunohistochemical validation (Table 1). Staining for these six markers was examined in a small training cohort of HNSCC ($n = 6$) and SQCLC ($n = 6$) (Table 2).

Patient collection for immunohistochemical validation

The clinical and pathological data of 98 patients with SQCLC and 96 patients with HNSCC were evaluated. Of the 96 HNSCC examined, 12 tumours were p16 positive (12.5%); of these, six tumours were located in the oropharynx and six in the pharynx or larynx. In contrast, 84 tumours were p16 negative (87.5%). The SQCLCs were UICC stage I–III and the HNSCCs ranged from stage I to IV (Table 3). All patients were treated by surgery and none of the patients had received neoadjuvant therapy or primary chemo- or radiotherapy. Squamous cell histology of all samples was confirmed by expert pathological review (FB, PS, and HB).

Distinction of SQCLC and HNSCC by immunohistochemical biomarkers

We immunohistochemically stained the described collection of squamous cell carcinomas with antibodies specific for CAV1, CAV2, LGALS1, and LGALS7 (upregulated in HNSCC) as well as CK19 and UDP-glucose-6-dehydrogenase (UGDH) (upregulated in SQCLC) in order to test their ability to distinguish SQCLC and HNSCC.

Analysis of proteins upregulated in HNSCC

Staining of CAV1 in SQCLC and HNSCC

Tumour cells assessed as positive showed cytoplasmic and often membranous staining. For SQCLC, 33.7% (31/92) of tumours were positive (Figure 2A). HNSCC presented significantly higher expression of CAV1 ($p < 0.001$), with 89.9% (80/89) of the tumours assessed as immunohistochemically positive (Figure 2B). Median IRSs of 0 for SQCLC and 6 for HNSCC were observed

Figure 3. Staining of LGALS1 and LGALS7 in SQCLC and HNSCC. (A) The SQCLC tumour cells show no or weakly positive expression of LGALS1 (total magnification $\times 50$). (B) Diffuse expression of LGALS1 in HNSCC (total magnification $\times 50$). (C) Box plot for LGALS1 expression in SQCLC and HNSCC; the horizontal lines within the boxes represent the median IRS values, $***p < 0.001$. (D) LGALS1 had an AUC value of 0.638, a sensitivity of 92.1%, and a specificity of 35.4%. (E) The SQCLC tumour cells show no or only a weak immunohistochemical reaction for LGALS7 (total magnification $\times 50$). (F) Box plot for LGALS7 expression in SQCLC and HNSCC; the horizontal lines within the boxes represent the median IRS values, $***p < 0.001$. (G) The HNSCC tumour cells show strong cytoplasmic expression of LGALS7 (total magnification $\times 50$). (H) LGALS7 had an AUC value of 0.802, a sensitivity of 94.5%, and a specificity of 66.0%.

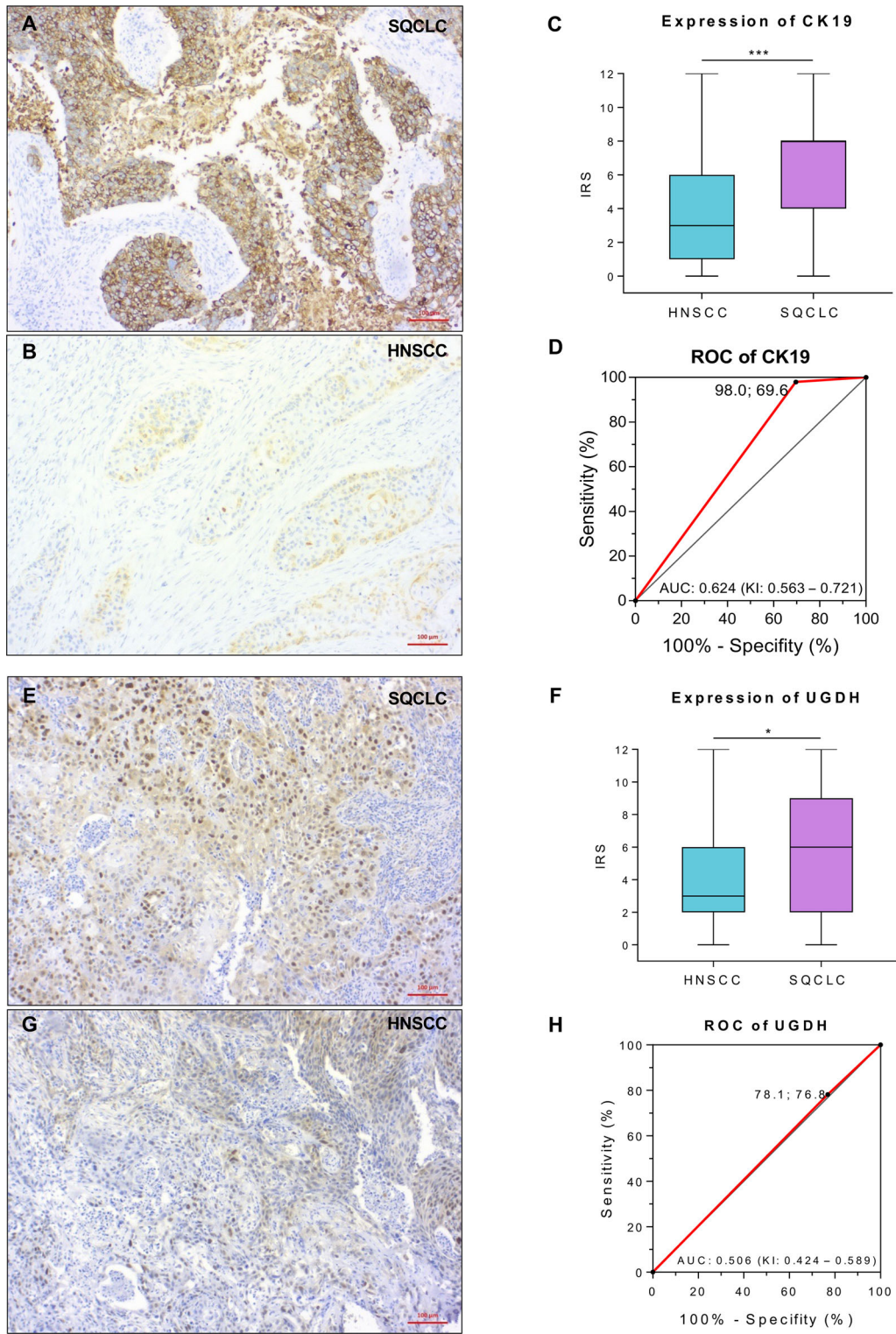


Figure 4. Legend on next page.

(Figure 2C). In our analysis, CAV1 showed an AUC value of 0.781, a sensitivity of 89.9%, a specificity of 66.3%, a positive predictive value of 72.1%, and a negative predictive value of 87.1% (Figure 2D). The immunohistochemical data of CAV1 depending on the subsites of HNSCC are listed in supplementary material, File S1.

Staining of CAV2 in SQCLC and HNSCC

Tumour cells assessed as positive showed cytoplasmic and often membranous staining. For SQCLC, 74.2% (72/97) of tumours were positive (Figure 2E). HNSCC presented significantly higher expression of CAV2 ($p < 0.001$), with 96.8% (90/93) of the tumours assessed as immunohistochemically positive (Figure 2F). Median IRSs of 3 for SQCLC and 6 for HNSCC were observed (Figure 2G). CAV2 showed an AUC value of 0.613, a sensitivity of 96.8%, a specificity of 25.8%, a positive predictive value of 55.6%, and a negative predictive value of 89.3% (Figure 2H). The immunohistochemical data of CAV2 depending on the subsites of HNSCC are listed in supplementary material, File S1.

Expression of LGALS1 in SQCLC and HNSCC

Tumour cells assessed as positive showed cytoplasmic and often nuclear staining. For SQCLC, 64.6% (62/96) of tumours were positive (Figure 3A). HNSCC presented significantly higher expression of LGALS1 ($p < 0.001$), with 92.1% (82/89) of the tumours assessed as immunohistochemically positive (Figure 3B). Median IRSs of 2 for SQCLC and 6 for HNSCC were observed (Figure 3C). LGALS1 showed an AUC value of 0.638, a sensitivity of 92.1%, a specificity of 35.4%, a positive predictive value of 56.9%, and a negative predictive value of 81.6% (Figure 3D). The immunohistochemical data of LGALS1 depending on the subsites of HNSCC are listed in supplementary material, File S1.

Expression of LGALS7 in SQCLC and HNSCC

Tumour cells assessed as positive showed mainly cytoplasmic and focally nuclear staining. For SQCLC, 34% (33/97) of tumours were positive (Figure 3E). HNSCC

presented significantly higher expression of LGALS7 ($p < 0.001$), with 94.5% (86/91) of the tumours assessed as immunohistochemically positive (Figure 3F). Median IRSs of 0 for SQCLC and 6 for HNSCC were observed (Figure 3G). LGALS7 showed an AUC value of 0.802, a sensitivity of 94.5%, a specificity of 66.0%, a positive predictive value of 72.3%, and a negative predictive value of 92.3% (Figure 3H). The immunohistochemical data of LGALS7 depending on the subsites of HNSCC are listed in supplementary material, File S1.

Analysis of proteins upregulated in SQCLC

Staining of CK19 in SQCLC and HNSCC

Tumour cells assessed as positive showed mainly cytoplasmic and membranous staining. For SQCLC, 97.9% (96/98) of tumours were positive (Figure 4A). HNSCC presented significantly lower expression of CK19 ($p < 0.001$), with 69.5% (64/92) of the tumours assessed as immunohistochemically positive (Figure 4B). Median IRSs of 8 for SQCLC and 3 for HNSCC were observed (Figure 4C). CK19 showed an AUC value of 0.624, a sensitivity of 98.0%, a specificity of 30.4%, a positive predictive value of 60.0%, and a negative predictive value of 93.3% (Figure 4D).

Staining of UGDH in SQCLC and HNSCC

Tumour cells assessed as positive showed mainly cytoplasmic and nuclear staining. For SQCLC, 78.1% (75/96) of tumours were positive (Figure 4E). HNSCC presented a significant difference in expression of UGDH ($p = 0.0131$), while showing a nearly similar immunohistochemical positivity of 76.8% (73/95) (Figure 4F). This can be explained by a median IRS of 6 for SQCLC and 3 for HNSCC (Figure 4G). UGDH showed an AUC value of 0.506, a sensitivity of 78.1%, a specificity of 23.2%, a positive predictive value of 50.7%, and a negative predictive value of 51.2% (Figure 4H).

Expression of combination of CAV1 and LGALS7 in SQCLC and HNSCC

With an AUC value between 0.613 and 0.802 (Figures 2D,H and 3D,H), we observed that single

Figure 4. Staining of CK19 and UGDH in SQCLC and HNSCC. (A) The SQCLC tumour cells show diffuse positive expression of CK19 (total magnification $\times 100$). (B) There is weak staining for CK19 in HNSCC ($\times 100$). (C) Box plot for CK19 staining in SQCLC and HNSCC; the horizontal lines within the boxes represent the median IRS values, $***p < 0.001$. (D) CK19 had an AUC value of 0.624, a sensitivity of 98.0%, and a specificity of 30.4%. (E) The SQCLC tumour cells show moderate or only weak immunohistochemical reaction for UGDH (total magnification $\times 100$). (F) Box plot for UGDH staining in SQCLC and HNSCC; the horizontal lines within the boxes represent the median IRS values, $*p < 0.0131$. (G) The HNSCC tumour cells show only weak cytoplasmic staining for UGDH (total magnification $\times 100$). (H) UGDH had an AUC value of 0.506, a sensitivity of 78.1%, and a specificity of 23.2%.

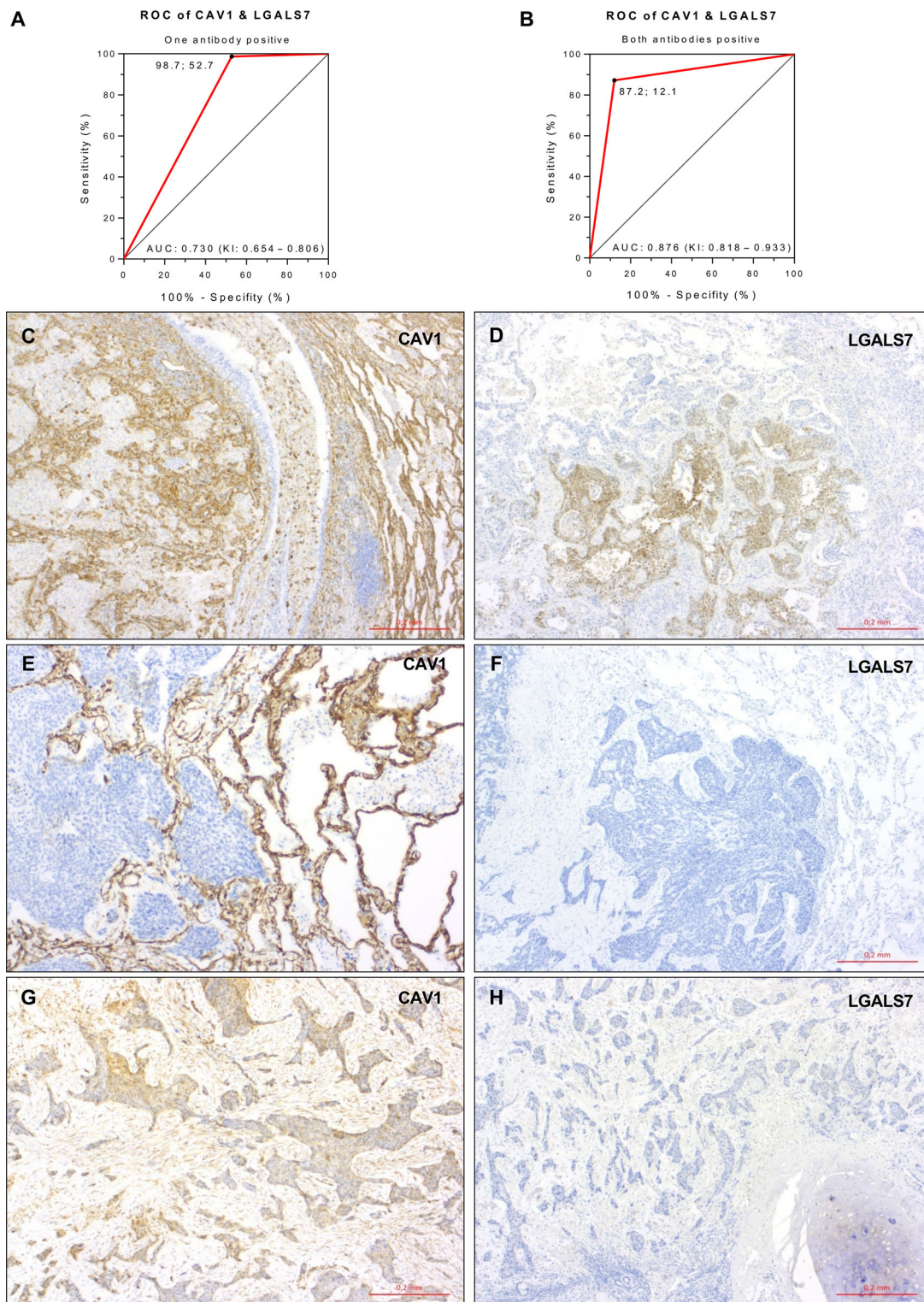


Figure 5. Legend on next page.

immunohistochemical marker was able to distinguish between HNSCC and SQCLC, but the specificity in particular was too low for routine diagnostic use. Therefore, we tested whether a combination of the two most promising markers, CAV1 and LGALS7, was able to further increase the diagnostic accuracy. If only one antibody had to be positive to diagnose an HNSCC, the sensitivity increased to 98.7%. The specificity decreased to 47.3%. The positive predictive value was 61.0% and the negative predictive value 97.7%. The new AUC value was 0.730 (Figure 5A). If both antibodies showed a positive reaction, the sensitivity was 87.2%, specificity improved to 87.9%, the positive predictive value was 86.1%, the negative predictive value 88.9%, and the AUC value was 0.876 (Figure 5B).

Immunohistochemical examination of lung tumours of unknown origin

Based on our findings mentioned above, we used the antibody panel consisting of CAV1 and LGALS7 to examine 12 lung tumours from patients diagnosed previously with an HNSCC (Table 4). Both markers had to be assessed positive (IRS \geq 2) in order to diagnose an HNSCC. Of the 12 carcinomas, 7 showed expression of both markers and were therefore classified as met-HNSCC (Figure 5C,D). Two tumours were negative and were classified as SQCLC (Figure 5E,F). Three cases (Tu-2, Tu-5, and Tu-10) were positive for CAV1 and negative for LGALS7 (Figure 5G,H). These were assessed as tumours of uncertain origin. Further clinical information can be found in Supplementary Table S4.

Discussion

Patients with HNSCC can develop both pulmonary metastases and metachronous primary squamous cell carcinoma of the lung in the course of their disease due to the similar aetiology and risk factors. The distinction between these two entities is a major

diagnostic challenge, because the morphology and the genetic alterations are very similar and there are no reliable biomarkers known yet. However, the correct diagnosis is important for the prognosis and therapy of the patient. Until now, the distinction of those cases was mainly based on clinical and radiological criteria [6,7,10,15]. Therefore, it would be a great diagnostic gain to discover biomarkers that can differentiate between both entities with a high sensitivity and specificity.

To detect new markers, we analysed cell cultures of HNSCC and SQCLC with SILAC-based MS and tested the chosen proteins immunohistochemically on a cohort of HNSCC and SQCLC. These tests were performed using primary tumours, as the differentiation between met-HNSCC and SQCLC is still a problem and we have no gold standard or ground truth. A limitation of this method is tumour evolution between primary and metastasis and therefore a risk of antigen loss. This has to be further analysed by staining more lung tumours in future studies.

Caveolin-1 (CAV1) belongs to a group of caveolins and is an integral membrane protein. As part of the caveolin scaffolding domain, which are cytoskeletal associated proteins, it links cell adhesion molecules and signalling molecules. Therefore, it participates in multiple processes of malignant tumour cells including signal transduction, cell transformation, cell migration, and metastasis [16–19]. Different studies were able to show high expression of CAV1 in HNSCC, squamous cell carcinoma of the oesophagus and the uterine cervix as well as adenocarcinoma of the prostate. In contrast, carcinomas of the lung, pancreas, ovaries, and breast showed low expression [20–25]. With SILAC-based MS as well as immunohistochemistry, significantly higher expression of CAV1 was identified in HNSCC than in SQCLC. The study of Vachani *et al*, who compared the gene expression of 18 HNSCC and 10 SQCLC, showed similar results. The dominant immunohistochemical expression pattern in our series showed strong staining in the periphery and absent or weak staining in the middle of the tumour. This

Figure 5. Expression of CAV1 and LGALS7 in tumours of the lung after HNSCC. (A) If only one antibody had to be positive to diagnose HNSCC, the sensitivity increased to 98.7% and the specificity decreased to 47.3%. The new AUC value was 0.730. (B) If both antibodies had to show a positive reaction, the sensitivity was 87.2%, specificity improved to 87.9%, and the AUC value was 0.876. (C, D) Expression of CAV1 and LGALS7 in tumour cells of a lung tumour in a patient with HNSCC, which was immunohistochemically classified as met-HNSCC (total magnification \times 50). (E, F) No expression of CAV1 and LGALS7 in a lung tumour in a patient with HNSCC, which was immunohistochemically classified as SQCLC (E, total magnification \times 100; F, total magnification \times 50). (G, H) Weak expression of CAV1 and no expression of LGALS7 in a lung tumour in a patient with HNSCC, which was immunohistochemically classified as uncertain (total magnification \times 50).

Table 4. Immunohistochemistry of lung tumours of unknown origin in patients with prior HNSCC.

| Case | Age | Gender | Localisation of HNSCC | Recurrence | Pulmonary foci (n) | Time interval (years) | pT | pN | Grade | CAV1 | LGALS7 | Clinical classification | IHC classification |
|-------|-----|--------|-----------------------|------------|--------------------|-----------------------|----|----|-------|----------|----------|-------------------------|--------------------|
| Tu-01 | 61 | M | Larynx | Yes | 2 | <1 | 4 | 2b | G2 | Positive | Positive | met-HNSCC | met-HNSCC |
| Tu-02 | 76 | M | Larynx | Yes | 1 | ≥3 | 1 | 0 | G2 | Positive | Negative | met-HNSCC | Uncertain |
| Tu-03 | 62 | M | Oral cavity | Yes | 1 | >3 | 3 | 1 | G2 | Positive | Positive | met-HNSCC | met-HNSCC |
| Tu-04 | 73 | M | Oropharynx | No | 2 | <1 | 2 | 0 | G2 | Negative | Negative | met-HNSCC | SQCLC |
| Tu-05 | 71 | M | Hypopharynx | No | 1 | <3 | 3 | 1 | G2 | Positive | Negative | SQCLC | Uncertain |
| Tu-06 | 66 | M | Larynx | No | 1 | <1 | 2 | 2b | G2 | Negative | Negative | met-HNSCC | SQCLC |
| Tu-07 | 79 | M | Hypopharynx | No | 1 | <3 | 4 | 2b | G2 | Positive | Positive | SQCLC | met-HNSCC |
| Tu-08 | 57 | M | Larynx | No | 1 | <3 | 4 | 2c | G2 | Positive | Positive | SQCLC | met-HNSCC |
| Tu-09 | 49 | M | Larynx | No | 1 | <3 | 2 | 0 | G2 | Positive | Positive | SQCLC | met-HNSCC |
| Tu-10 | 23 | F | Oral cavity | No | 1 | <1 | 2 | 2c | G2 | Positive | Negative | met-HNSCC | Uncertain |
| Tu-11 | 39 | M | Oropharynx | No | 1 | <1 | 2 | 1 | G2 | Positive | Positive | met-HNSCC | met-HNSCC |
| Tu-12 | 62 | M | Oropharynx | No | 1 | <3 | 3 | 2c | G3 | Positive | Positive | SQCLC | met-HNSCC |

Data for the 12 lung tumours (Tu-01–Tu-12) of unknown origin, those of the previously known HNSCC (localisation, relapse, pT, pN, grade) as well as the results of the IHC staining of the lung tumours are presented. The negative IHC results and the classification as SQCLC are marked in bold and the positive results and the classification as met-HNSCC in italics. M, male; F, female; IHC, immunohistochemical.

biphasic pattern was also described in other studies and indicated high activity of CAV1 in the area of tumour invasion [24,26–28].

Caveolin-2 (CAV2) is, similar to CAV1, an integral membrane protein. Under physiological conditions, it is expressed in the same cells as CAV1 and is needed for intracellular transport of coated vesicles to the cell membrane [29–32]. In other studies, CAV2 showed strong expression in urothelial and renal cell carcinomas. Expression is low in carcinomas of the breast and lung, and in follicular carcinoma of the thyroid gland [33–39]. Consistent with our findings, Vachani *et al* showed higher expression of CAV2 in HNSCC than in SQCLC in their gene expression study [15].

Galectin-1 (LGALS1) is involved in many cell biological processes such as cell growth, cell adhesion, cell migration, angiogenesis, and apoptosis. Thus, it plays a role in tumour development and progression [40,41]. Increased expression has already been shown in adenocarcinoma of the colon, pancreas, and uterus; urothelial carcinoma; prostate carcinoma; glioblastoma; and HNSCC [42–45]. A possible explanation for the high expression levels of LGALS1 in HNSCC was provided by Le *et al*. They showed a positive correlation to the hypoxia marker carbonic anhydrase IX (CA IX) and a negative correlation to CD3 in 101 cases of HNSCC. They established the hypothesis that the strong expression of CA IX is caused by the many hypoxic areas in HNSCC. This leads to increased production of LGALS1, which promotes the apoptosis of T-cells and blocks the activation of T-cells [46].

Galectin-7 (LGALS7) has a high specificity for squamous epithelium and is physiologically expressed in all squamous epithelia and myoepithelium of the breast [47]. It takes part in differentiation and development of epithelia as well as tissue repair, cell–cell interaction, cell–matrix interaction, and apoptosis [40,48]. LGALS7 has positive and negative regulating functions depending on the tumour entity [49]. Decreased expression was shown in squamous cell carcinoma of the cervix uteri, adenocarcinoma of the stomach, and urothelial carcinoma. In contrast, increased expression was shown in squamous cell carcinoma of the head and neck and oesophagus [50–55]. We were able to show significantly higher expression of LGALS7 in HNSCC than in SQCLC with MS as well as immunohistochemistry. These findings are in accordance with the results of a study by Bohnenberger *et al* who performed MS-based proteomics and immunohistochemical analysis in a large group of SQCLC and HNSCC [8].

The two proteins CK19 and UGDH, which were upregulated in SQCLC, are also of interest. UGDH was not suitable to distinguish SQCLC from HNSCC. In

contrast, CK19 is positive in 97.9% of the SQCLC (AUC 0.624, sensitivity 98.0%, and specificity 30.4%). This is in line with the results of Ichinose *et al* who showed CK19 in combination with MMP3, ZNF830, and PI3 as a protein to differentiate between SQCLC and HNSCC [56].

As a combination of antibodies in a biomarker panel can increase the sensitivity and specificity regarding the differential diagnosis, we combined the two most promising antibodies anti-CAV1 and anti-LGALS7. The quality criteria were improved when both markers had to be positive to diagnose an HNSCC.

We tested our biomarker panel on a cohort of 12 lung tumours of unknown origin after HNSCC. The tumours were clinically classified using a classification score according to Ichinose *et al* [56]. This score includes clinical criteria such as local relapse of the HNSCC, the number of pulmonary tumours, and the time between the HNSCC and the lung tumour. Bohnenberger *et al* showed in their study that the clinical classification (here especially by Ichinose *et al*) did not show a difference in survival [8]. They presented differentiation into met-HNSCC and SQCLC based on proteomic analyses. The newly established groups showed a significant difference in survival. Thereby, they were able to demonstrate that the clinical classification is unreliable and therefore not ground truth. In our cohort, 3 of the 12 tumours (25%) showed concordance and 6 discordance (50%) between the immunohistochemical and clinical classification. Tu-04 and Tu-06 were immunohistochemically classified as SQCLC in contrast to the clinical classification. The pathological parameters (solitary pulmonary nodule, no relapse, pN0) of Tu-04 supported the immunohistochemical diagnosis of SQCLC. The new classification of Tu-06 is supported by the clinical information that there was no local relapse. The pN2b and death due to tumour are in favour of the diagnosis of met-HNSCC. This is a good example of difficulties trying to differentiate SQCLC and met-HNSCC clinically. Tu-07, Tu-08, Tu-09, and Tu-12 were immunohistochemically categorised as met-HNSCC in contrast to the clinical classification. The clinical and pathological data of Tu-07, Tu-08, and Tu-12 (pN2b/pN2c, death) supported the immunohistochemical classification, whereas the data for Tu-09 (pN0, no local recurrence, alive) are in favour of SQCLC.

In summary, we were able to establish an immunohistochemical marker panel that can be used to differentiate between squamous cell carcinomas of the lung (SQCLC) and the head and neck (HNSCC). We conclude that CAV1 and LGALS7, especially in combination as an antibody panel, represent promising marker candidates to differentiate between primary SQCLC and pulmonary metastases (met-HNSCC) of previously known HNSCC.

Author contributions statement

FB, HB and AR conceived the study concepts and together with JJ and PS the study design. JJ, AR, AF, PB, PK, HS and AH-E acquired data and HB, HU and TO analysed and interpreted them. All authors were involved in writing the paper and had final approval of the submitted and published versions.

References

1. Vigneswaran N, Williams MD. Epidemiologic trends in head and neck cancer and aids in diagnosis. *Oral Maxillofac Surg Clin North Am* 2014; **26**: 123–141.
2. Marur S, Forastiere AA. Head and neck squamous cell carcinoma: update on epidemiology, diagnosis, and treatment. *Mayo Clin Proc* 2016; **91**: 386–396.
3. Ferlito A, Shaha AR, Silver CE, *et al*. Incidence and sites of distant metastases from head and neck cancer. *ORL J Otorhinolaryngol Relat Spec* 2001; **63**: 202–207.
4. Atabek U, Mohit-Tabatabai MA, Raina S, *et al*. Lung cancer in patients with head and neck cancer. Incidence and long-term survival. *Am J Surg* 1987; **154**: 434–438.
5. Leong PP, Rezai B, Koch WM, *et al*. Distinguishing second primary tumors from lung metastases in patients with head and neck squamous cell carcinoma. *J Natl Cancer Inst* 1998; **90**: 972–977.
6. Geurts TW, Nederlof PM, van den Brekel MW, *et al*. Pulmonary squamous cell carcinoma following head and neck squamous cell carcinoma: metastasis or second primary? *Clin Cancer Res* 2005; **11**: 6608–6614.
7. Talbot SG, Estilo C, Maghami E, *et al*. Gene expression profiling allows distinction between primary and metastatic squamous cell carcinomas in the lung. *Cancer Res* 2005; **65**: 3063–3071.
8. Bohnenberger H, Kaderali L, Ströbel P, *et al*. Comparative proteomics reveals a diagnostic signature for pulmonary head-and-neck cancer metastasis. *EMBO Mol Med* 2018; **10**: e8428.
9. Finley RK 3rd, Verazin GT, Driscoll DL, *et al*. Results of surgical resection of pulmonary metastases of squamous cell carcinoma of the head and neck. *Am J Surg* 1992; **164**: 594–598.
10. Jones AS, Morar P, Phillips DE, *et al*. Second primary tumors in patients with head and neck squamous cell carcinoma. *Cancer* 1995; **75**: 1343–1353.
11. Bohnenberger H, Oellerich T, Engelke M, *et al*. Complex phosphorylation dynamics control the composition of the Syk interactome in B cells. *Eur J Immunol* 2011; **41**: 1550–1562.
12. Bremmer F, Bohnenberger H, Küffer S, *et al*. Proteomic comparison of malignant human germ cell tumor cell lines. *Dis Markers* 2019; **2019**: 8298524.
13. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008; **26**: 1367–1372.
14. Fichtner A, Richter A, Filmar S, *et al*. The detection of isochromosome i(12p) in malignant germ cell tumours and tumours with somatic malignant transformation by the use of quantitative

- real-time polymerase chain reaction. *Histopathology* 2021; **78**: 593–606.
15. Vachani A, Nebozhyn M, Singhal S, et al. A 10-gene classifier for distinguishing head and neck squamous cell carcinoma and lung squamous cell carcinoma. *Clin Cancer Res* 2007; **13**: 2905–2915.
 16. Liu P, Rudick M, Anderson RG. Multiple functions of caveolin-1. *J Biol Chem* 2002; **277**: 41295–41298.
 17. Razani B, Woodman SE, Lisanti MP. Caveolae: from cell biology to animal physiology. *Pharmacol Rev* 2002; **54**: 431–467.
 18. Williams TM, Lisanti MP. Caveolin-1 in oncogenic transformation, cancer, and metastasis. *Am J Physiol Cell Physiol* 2005; **288**: C494–C506.
 19. Goetz JG, Lajoie P, Wiseman SM, et al. Caveolin-1 in tumor progression: the good, the bad and the ugly. *Cancer Metastasis Rev* 2008; **27**: 715–735.
 20. Wiechen K, Diatchenko L, Agoulnik A, et al. Caveolin-1 is down-regulated in human ovarian carcinoma and acts as a candidate tumor suppressor gene. *Am J Pathol* 2001; **159**: 1635–1643.
 21. Kato K, Hida Y, Miyamoto M, et al. Overexpression of caveolin-1 in esophageal squamous cell carcinoma correlates with lymph node metastasis and pathologic stage. *Cancer* 2002; **94**: 929–933.
 22. Hung KF, Lin SC, Liu CJ, et al. The biphasic differential expression of the cellular membrane protein, caveolin-1, in oral carcinogenesis. *J Oral Pathol Med* 2003; **32**: 461–467.
 23. Yoo SH, Park YS, Kim HR, et al. Expression of caveolin-1 is associated with poor prognosis of patients with squamous cell carcinoma of the lung. *Lung Cancer* 2003; **42**: 195–202.
 24. Zhao X, Yu G, Yu X, et al. Caveolin-1 is overexpressed in hypopharyngeal squamous cell carcinoma and correlates with clinical parameters. *Oncol Lett* 2016; **12**: 2371–2374.
 25. Fu P, Chen F, Pan Q, et al. The different functions and clinical significances of caveolin-1 in human adenocarcinoma and squamous cell carcinoma. *Onco Targets Ther* 2017; **10**: 819–835.
 26. Zhang H, Su L, Müller S, et al. Restoration of caveolin-1 expression suppresses growth and metastasis of head and neck squamous cell carcinoma. *Br J Cancer* 2008; **99**: 1684–1694.
 27. Xue J, Chen H, Diao L, et al. Expression of caveolin-1 in tongue squamous cell carcinoma by quantum dots. *Eur J Histochem* 2010; **54**: e20.
 28. Masuelli L, Budillon A, Marzocchella L, et al. Caveolin-1 overexpression is associated with simultaneous abnormal expression of the E-cadherin/alpha-beta catenins complex and multiple ErbB receptors and with lymph nodes metastasis in head and neck squamous cell carcinomas. *J Cell Physiol* 2012; **227**: 3344–3353.
 29. Scherer PE, Lewis RY, Volonte D, et al. Cell-type and tissue-specific expression of caveolin-2. Caveolins 1 and 2 co-localize and form a stable hetero-oligomeric complex in vivo. *J Biol Chem* 1997; **272**: 29337–29346.
 30. Engelman JA, Zhang XL, Galbiati F, et al. Chromosomal localization, genomic organization, and developmental expression of the murine caveolin gene family (Cav-1, -2, and -3). Cav-1 and Cav-2 genes map to a known tumor suppressor locus (6-A2/7q31). *FEBS Lett* 1998; **429**: 330–336.
 31. Parolini I, Sargiacomo M, Galbiati F, et al. Expression of caveolin-1 is required for the transport of caveolin-2 to the plasma membrane. Retention of caveolin-2 at the level of the Golgi complex. *J Biol Chem* 1999; **274**: 25718–25725.
 32. Williams TM, Lisanti MP. The caveolin proteins. *Genome Biol* 2004; **5**: 214.
 33. Aldred MA, Ginn-Pease ME, Morrison CD, et al. Caveolin-1 and caveolin-2, together with three bone morphogenetic protein-related genes, may encode novel tumor suppressors down-regulated in sporadic follicular thyroid carcinogenesis. *Cancer Res* 2003; **63**: 2864–2871.
 34. Fong A, Garcia E, Gwynn L, et al. Expression of caveolin-1 and caveolin-2 in urothelial carcinoma of the urinary bladder correlates with tumor grade and squamous differentiation. *Am J Clin Pathol* 2003; **120**: 93–100.
 35. Sagara Y, Mimori K, Yoshinaga K, et al. Clinical significance of caveolin-1, caveolin-2 and HER2/neu mRNA expression in human breast cancer. *Br J Cancer* 2004; **91**: 959–965.
 36. Wikman H, Seppänen JK, Sarhadi VK, et al. Caveolins as tumour markers in lung cancer detected by combined use of cDNA and tissue microarrays. *J Pathol* 2004; **203**: 584–593.
 37. Elsheikh SE, Green AR, Rakha EA, et al. Caveolin 1 and caveolin 2 are associated with breast cancer basal-like and triple-negative immunophenotype. *Br J Cancer* 2008; **99**: 327–334.
 38. Sugie S, Mukai S, Yamasaki K, et al. Significant association of caveolin-1 and caveolin-2 with prostate cancer progression. *Cancer Genomics Proteomics* 2015; **12**: 391–396.
 39. Liu F, Shangli Z, Hu Z. CAV2 promotes the growth of renal cell carcinoma through the EGFR/PI3K/Akt pathway. *Onco Targets Ther* 2018; **11**: 6209–6216.
 40. Barondes SH, Cooper DN, Gitt MA, et al. Galectins. Structure and function of a large family of animal lectins. *J Biol Chem* 1994; **269**: 20807–20810.
 41. Ebrahim AH, Alalawi Z, Mirandola L, et al. Galectins in cancer: carcinogenesis, diagnosis and therapy. *Ann Transl Med* 2014; **2**: 88.
 42. Gillenwater A, Xu XC, el-Naggar AK, et al. Expression of galectins in head and neck squamous cell carcinoma. *Head Neck* 1996; **18**: 422–432.
 43. Sanjuan X, Fernández PL, Castells A, et al. Differential expression of galectin 3 and galectin 1 in colorectal cancer progression. *Gastroenterology* 1997; **113**: 1906–1915.
 44. Shen J, Person MD, Zhu J, et al. Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry. *Cancer Res* 2004; **64**: 9018–9026.
 45. Camby I, Le Mercier M, Lefranc F, et al. Galectin-1: a small protein with major functions. *Glycobiology* 2006; **16**: 137R–157R.
 46. Le QT, Shi G, Cao H, et al. Galectin-1: a link between tumor hypoxia and tumor immune privilege. *J Clin Oncol* 2005; **23**: 8932–8941.
 47. Magnaldo T, Fowles D, Darmon M. Galectin-7, a marker of all types of stratified epithelia. *Differentiation* 1998; **63**: 159–168.
 48. Liu FT, Rabinovich GA. Galectins as modulators of tumour progression. *Nat Rev Cancer* 2005; **5**: 29–41.
 49. Kaur M, Kaur T, Kamboj SS, et al. Roles of galectin-7 in cancer. *Asian Pac J Cancer Prev* 2016; **17**: 455–461.

50. Matsui Y, Ueda S, Watanabe J, *et al.* Sensitizing effect of galectin-7 in urothelial cancer to cisplatin through the accumulation of intracellular reactive oxygen species. *Cancer Res* 2007; **67**: 1212–1220.
51. Saussez S, Decaestecker C, Lorfèvre F, *et al.* Increased expression and altered intracellular distribution of adhesion/growth-regulatory lectins galectins-1 and -7 during tumour progression in hypopharyngeal and laryngeal squamous cell carcinomas. *Histopathology* 2008; **52**: 483–493.
52. Saussez S, Camby I, Toubeau G, *et al.* Galectins as modulators of tumor progression in head and neck squamous cell carcinomas. *Head Neck* 2007; **29**: 874–884.
53. Zhu X, Ding M, Yu ML, *et al.* Identification of galectin-7 as a potential biomarker for esophageal squamous cell carcinoma by proteomic analysis. *BMC Cancer* 2010; **10**: 290.
54. Zhu H, Wu TC, Chen WQ, *et al.* Roles of galectin-7 and S100A9 in cervical squamous carcinoma: clinicopathological and in vitro evidence. *Int J Cancer* 2013; **132**: 1051–1059.
55. Kim SJ, Hwang JA, Ro JY, *et al.* Galectin-7 is epigenetically-regulated tumor suppressor in gastric cancer. *Oncotarget* 2013; **4**: 1461–1471.
56. Ichinose J, Shinozaki-Ushiku A, Takai D, *et al.* Differential diagnosis between primary lung squamous cell carcinoma and pulmonary metastasis of head and neck squamous cell carcinoma. *Expert Rev Anticancer Ther* 2016; **16**: 403–410.

SUPPLEMENTARY MATERIAL ONLINE

File S1. Expression of immunohistochemical markers according to tumour subsite

Figure S1. Examples of IRS

Table S1. Complete proteomic data

Table S2. Antibodies used in this study

Table S3. Differential protein expression analysis

Table S4. Immunohistochemistry and complete clinical data of lung tumours of unknown origin in patients with prior HNSCC