

Differential Proteomics Identifies Protein Biomarkers That Predict Local Relapse of Head and Neck Squamous Cell Carcinomas

Tieneke B.M. Schaaij-Visser,1,2 A. Peggy Graveland,3 Sharon Gauci,1,2 Bouwewijn J.M. Braakhuis,3 Marijke Buijze,3 Albert J.R. Heck,1,2 Dirk J. Kuik,4 Elisabeth Bloemena,5 C. René Leemans,3 Monique Slijper,1,2 and Ruud H. Brakenhoff3

Abstract

Purpose: The 5-year survival rates of head and neck squamous cell carcinomas (HNSCC) remain disappointing. HNSCCs develop in precursor fields of genetically altered cells that are often not completely resected when the tumor is excised, causing local relapse. These precursor fields are mostly recognized as dysplasia, but histologic grading cannot reliably predict malignant transformation. Our aim was to discover and validate protein biomarkers that can detect precursor fields and predict local relapse in HNSCC using immunostaining of surgical margins.

Experimental Design: We compared paired and genetically characterized normal, precursor, and tumor tissues of eight patients by proteome analysis to identify differentially expressed proteins. The prognostic value of candidate protein biomarkers was evaluated by immunohistochemical analysis of 222 surgical margins of 46 HNSCC patients who developed local relapse or remained disease free. Significant associations were determined by Kaplan-Meier survival analysis and Cox-proportional hazards models.

Results: Forty proteins showed significant differential expression (false discovery rate–corrected \( P < 0.05 \)). Most discriminative markers suited for immunostaining were keratin 4 and cornulin. Low expression in the surgical margins of keratin 4 (hazard ratio, 3.8; \( P = 0.002 \)), cornulin (hazard ratio, 2.7; \( P = 0.025 \)), and their combination (hazard ratio, 8.8; \( P = 0.0005 \)) showed a highly significant association with the development of local relapse. Dysplasia grading had no prognostic relevance.

Conclusions: Immunohistochemical assessment of keratin 4 and cornulin expression in surgical margins of HNSCC patients outperforms histopathologic grading in predicting the risk for local relapse. These markers can be used to initiate more frequent and lifelong surveillance of patients at high risk of local relapse, and enable selection for adjuvant treatment or tertiary prevention trials. (Clin Cancer Res 2009;15(24):7666–75)

Head and neck squamous cell carcinoma (HNSCC) develops in the mucosal linings of the upper aerodigestive tract and is the sixth most common cancer worldwide (1). The 5-year survival rates of HNSCC are approximately 60% (1) and have only moderately improved the last decades (2) mainly because 20% to 40% of all patients develop a local relapse in the same or adjacent anatomic region even when the surgical margins are histologically tumor free (3, 4). Clinically these relapses at the primary and adjacent anatomic sites are assigned as local recurrences when they develop within three years and at <2 cm distance of the primary tumor. Relapses not fulfilling these criteria are clinically classified as second primary tumors. Despite the difference in clinical assignment, many local relapses have in fact the same pathobiological origin (3–8).

It has been well established that HNSCC is the result of a multistep process characterized by the accumulation of genetic and epigenetic alterations (9). Genetic analysis of surgical margins has shown that HNSCC frequently develops in a field of...
translated in DNA isolated from the surgical margins, and these as loss of heterozygosity and mutated not as a single criterion. Alternatively, genetic markers such as dysplasia is the gold standard for identification and risk assessment of precursor fields in surgical margins, but it is subjective and has limited value to predict local relapse in objective and has limited value to predict local relapse in HNSCC patients based on immunohistochemical analysis of the surgical margins of the index tumor, whereas dysplasia grading failed.

Materials and Methods

Experimental design and workflow

This study was conducted in three consecutive phases: first a proteomics-based discovery screen was done, second an initial evaluation of candidate biomarkers, and third a clinical validation of the most promising protein biomarkers. For the discovery screen, a comparative proteome analysis was carried out to identify differentially expressed proteins between paired and genetically characterized samples of normal, precursor, and tumor tissues taken from the excised HNSCC specimens of eight patients. For the initial evaluation of the candidates with the most significant differential level of expression an independent set of formalin-fixed paraffin-embedded normal, preneoplastic, and tumor tissues were selected for immunostaining, using commercially available antibodies. The two most promising protein biomarkers were validated in a retrospective case-control study by immunostaining of the surgical margins of surgically treated head and neck cancer patients who either developed a local relapse or remained local disease free.

The study was approved by the Institutional Review Board and in accordance with the Dutch guidelines on the use of human specimens for research. Patients included for both the biomarker discovery screen and the retrospective clinical validation study were diagnosed with a primary HNSCC and scheduled for surgery. The tumors analyzed in the studies are representative for the patient population. All histologic evaluations and dysplasia grading were done by an experienced pathologist (E.B.) according to the WHO standard criteria (17).

Patients and tissue samples

Tissue selection for the biomarker discovery screen. For the biomarker discovery screen we analyzed the proteomes of frozen normal, precursor, and tumor tissues for changes in protein expression. Patients who were scheduled for surgical treatment of a squamous cell carcinoma in the oral cavity or oropharynx and consented to enrolment were included in the study. Biopsies were taken directly from the surgical specimen and frozen immediately in liquid nitrogen. One biopsy was taken from the tumor and four biopsies from the surrounding mucosal epithelium with a clinically normal appearance. Allelic loss analysis was done on these biopsies to assess the genetic changes to distinguish normal mucosa from precursor field (see below). To increase the chance of detecting focal precursor fields, the biopsies were immunostained for p53 (often overexpressed when mutated) and reviewed for morphologic abnormalities (8, 18). Guided by p53 immunostaining and morphology, tissues were microdissected for allelic loss analysis and preparation of protein extracts.

Tissue selection for initial evaluation of promising protein biomarker candidates. For the initial evaluation of the most promising protein biomarkers, a panel of formalin-fixed paraffin-embedded specimens was selected containing five normal mucosa samples, five severely dysplastic mucosa samples (preneoplastic tissue with a high risk for malignant progression), and five tumor samples.

Tissue selection for the retrospective clinical validation study. After tumor excision the surgical margins are routinely investigated by histologic examination for the presence of residual cancer as well as unresected precursor fields that are usually recognized as dysplastic mucosal changes. Histopathologic grading of dysplasia is the gold standard for identification and risk assessment of precursor fields in surgical margins, but it is subjective and has limited value to predict local relapse in individual cases. Only severe dysplasia might be considered an indication for adjuvant postoperative radiotherapy, but not as a single criterion. Alternatively, genetic markers such as loss of heterozygosity and mutated TP53 can be determined in DNA isolated from the surgical margins, and these have been reported to predict local relapse (6, 14, 16). However, the prognostic value of these approaches is still limited, and the required methods are laborious and not easy to implement in current clinical management.

Here we identified a panel of novel protein biomarkers using differential proteome analysis of paired and genetically characterized normal, precursor, and tumor tissues. With two of these candidate protein biomarkers we could reliably predict the development of local relapse in HNSCC patients based on immunohistochemical analysis of the surgical margins of the index tumor, whereas dysplasia grading failed.
were included who either developed a local relapse during follow-up in the same or adjacent anatomical region, or remained disease free. Further criteria for inclusion were: (a) primary tumor in the oral cavity or oropharynx, (b) surgical treatment with >5 mm histologically tumor-free surgical margins, and (c) N-stage ≤ N2b. Cases and controls were matched for primary treatment (surgery with/without postoperative radiotherapy) and T-stage (T1-2 or T3-4). For comparison with the current standard, every surgical margin was graded for dysplastic changes as either 0, 1, 2, or 3 for no, mild, moderate, or severe dysplasia, respectively. As final readout, the highest grade of dysplasia scored in any of the surgical margins of a patient was taken. Clinical and histopathologic details are listed in Supplementary Table S2.

Genetic characterization of tissues for the biomarker discovery screen

**P53 immunostaining.** All antibodies were obtained from DAKO. Frozen sections of tumor tissue and epithelial margins were fixed for 10 min in 4% paraformaldehyde in PBS and preincubated for 15 min with 2% normal rabbit serum. Primary antibodies against p53 (DO-7) were incubated for 1 h in the dilution 1:100. The second step was done with a biotinylated rabbit anti-mouse antibody (1:500) and in the final step horse radish peroxidase–labeled streptavidin–biotin complex was applied. The staining was developed with diaminobenzidine and H2O2 as chromogen. The sections were counterstained with haematoxylin and coverslipped with Kaiser’s glycerin. Mouse monoclonal antibody U36 directed against CD44v6 always stains squamous cells independent of genetic changes and was utilized as positive control.

**Allelic loss analysis.** Allelic loss analysis decisively distinguished normal tissue from precursor field and was also applied to determine the clonal relationship between precursor field and tumor. Allelic loss analysis was done as described previously (21). Briefly, from all selected specimens, 10 to 20 frozen 10-μm sections were cut, mounted on glass slides, and stained with methylene blue/toluidin blue. Areas of interest were manually microdissected under a stereomicroscope. Stroma tissue sections were cut, mounted on glass slides, and stained with haematoxylin and eosin, and immunostained with DO7 for mutated p53 as described above to guide microdissection. Microdissected tissue samples were incubated with primary antibodies against p53 (DO7) and labeled streptavidin–biotin complex was applied. The staining was developed with diaminobenzidine and H2O2 as chromogen. The sections were counterstained with haematoxylin and coverslipped with Kaiser’s glycerin. Mouse monoclonal antibody U36 directed against CD44v6 always stains squamous cells independent of genetic changes and was utilized as positive control.

Twelve microsatellite markers were used to assess allelic loss on chromosomes 3p, 9p, 11q, and 17p: D3S1766, D3S1029, D3S1293, D9S171, D9S162, D9S157, D11S2002, D11S1369, D11S1883, CHRNA1, TP53, and D17S1866. Ten nanograms of DNA were used to PCR-amplify the microsatellite markers and label the products with fluorescent dye. The ABI 3130 Genetic Analyzer (Applied Biosystems) was used to separate the PCR products and measure the signal intensities. Allelic loss was assigned if the signal intensity of one of the two alleles decreased with ≥50% compared with normal control DNA isolated from the stroma.

**Biomarker discovery by two-dimensional differential gel electrophoresis**

The proteomes of seven normal, eight precursor, and eight tumor tissues (see Table 1) were compared in a pairwise manner with two-dimensional differential gel electrophoresis (2D-DIGE; ref. 22, 23). For one patient all margins showed genetic changes and a normal sample could not be selected. All equipment and software for 2D-DIGE were obtained from GE Healthcare, unless stated otherwise.

**Protein extraction.** For each tissue sample, 30 frozen 10-μm sections were cut, mounted on glass slides, and stained with haematoxylin. Relevant epithelial (tumor) parts of the tissues were manually microdissected under a stereomicroscope, and transferred into lysis buffer [30 mmol/L TRIS (pH 8.5), 15 mmol/L NaCl, 1.5 mmol/L MgCl2, 7 mol/L urea, 2 mol/L thiourea, one tablet EDTA-free complete mini-protease inhibitor cocktail/100 mL; Roche], snap-frozen, and stored at -80°C until further use. Protein lysates were prepared by tip sonication (LABSONIC M homogenizer, Sartorius Filtratie BV) followed by centrifugation and collection of the soluble fraction. Protein lysates were cleaned from other cellular molecules with the Plus One 2D Clean-up kit (GE Healthcare), dissolved in labeling buffer [30 mmol/L TRIS (pH 8.8), 7 mol/L urea, 2 mol/L thiourea, 4% CHAPS, one tablet EDTA-free complete mini-protease inhibitor cocktail/100 mL], and the protein content was determined with the Plus One 2D Quant kit (GE Healthcare).

**Protein labeling and 2-DIGE.** Protein lysates were labeled with CyDye DIGE Fluor minimal dyes following the manufacturer’s protocol. However, due to the minute amounts of tissue samples, the labeling procedure was scaled down. In brief, 10 μg of protein sample were mixed with 80 pmol of either Cy3 or Cy5, and after a 30-min incubation procedure was scaled down. In brief, 10 μg of protein sample were mixed with 80 pmol of either Cy3 or Cy5, and after a 30-min incubation on ice the labeling reaction was stopped by adding 10 nmol of lysine. Equal amounts of all samples were mixed and 10 μg of this mix were labeled with Cy2 for each internal standard. Samples to be run on the same gel were combined, snap-frozen, and stored at -80°C.

Proteins were separated in two dimensions as described before (24). In the first dimension, proteins were resolved by iso-electric focusing using both pH 3-7 (24 cm, nonlinear) and pH 6-11 (18 cm, linear) pH-gradients. After equilibration of the pH-strips in buffer [50 mmol/L TRIS (pH 8.8), 6 mol/L urea, 2% SDS, 30% glycerol] containing 1% DTT (reduction, 10 min) and 2.5% iodoacetamide (alkylation, 10 min), the strips were placed on top of a 11% polyacrylamide gel for the second dimension. The gels were scanned on a Typhoon 9400 Imager, and the gel images were cropped with ImageQuant TL software (version 2005).

**DeCyder analysis of 2D gel images.** The spot codetection, quantification by normalization, and ratio calculation for the gel images from one single gel was done in the Differential In-Gel Analysis module of DeCyder software.

**Table 1. Patients and tissue selected for 2D-DIGE proteome analysis**

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Age (y)</th>
<th>Site</th>
<th>Subsite</th>
<th>T-stage</th>
<th>Precursor tissue</th>
<th>Normal tissue</th>
<th>P53 staining</th>
<th>Allelic loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>46</td>
<td>Oral cavity</td>
<td>Mobile tongue</td>
<td>T2 N2 BM0</td>
<td>No</td>
<td>9p</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>48</td>
<td>Oral cavity</td>
<td>Mobile tongue</td>
<td>T2 N2 BM0</td>
<td>Yes</td>
<td>3p, 9p</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>67</td>
<td>Oral cavity</td>
<td>Mobile tongue</td>
<td>T2 N2 BM0</td>
<td>No</td>
<td>9p</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>43</td>
<td>Hypopharynx</td>
<td>Piniform sinus</td>
<td>T2 N2 BM0</td>
<td>Yes</td>
<td>3p, 9p, 11q</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>78</td>
<td>Oral cavity</td>
<td>Mucosa lower jaw</td>
<td>T2 N2 BM0</td>
<td>No</td>
<td>3p</td>
<td></td>
<td>Yes*</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>64</td>
<td>Oral cavity</td>
<td>Retromolar trigone</td>
<td>T2 N2 BM0</td>
<td>Yes</td>
<td>9p, 11q</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>55</td>
<td>Larynx</td>
<td>Glottic larynx</td>
<td>T2 N2 BM0</td>
<td>Yes</td>
<td>11q, 17p</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>77</td>
<td>Oropharynx</td>
<td>Retromolar trigone</td>
<td>T2 N2 BM0</td>
<td>Yes</td>
<td>3p, 9p, 11q, 17p</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

*Mucosal tissues were classified as normal when p53 immunostaining was negative and no allelic loss was detected.
the DeCyder 2D software (version 6.5). For spot detection the V6 spot detection algorithm was used with the estimated number of spots set to 10,000 and the exclude filter set to >30,000 for volume. In the Biological Variation Analysis module of DeCyder, all the gel images were processed and matched. To select protein spots of which the intensities were significantly different among the normal, precancerous, and tumor tissues, a paired one-way ANOVA was done on the whole sample group. Paired Student’s r-tests were applied to detect differential protein spot intensities between either normal and precancerous tissue, normal and tumor tissue, or precancerous and tumor tissue, and to calculate average ratios between each of these two tissue types. P values were corrected for the false discovery rate (FDR). Proteins were considered differentially expressed when the FDR-corrected P values of the ANOVA analysis were <0.05. Hierarchical clustering of the samples was done in the Extended Data Analysis module of DeCyder.

Protein identification by liquid chromatography-tandem mass spectrometry. Spots from differentially expressed proteins were picked from the analytical gels with the Ettan Spot Picker (GE Healthcare) according to the manufacturer’s protocol. The gel plugs were washed in milliQ water followed by tryptic in-gel digestion as described (25). Peptides were extracted with 5% formic acid to a total volume of 20 μL.

Ten microliters of peptide sample was analyzed on either a nanoLC-LTQ-Orbitrap-MS or a nanoLC-LTQ-FT(ICR)-MS (both Thermo Electron) as described before (26, 27). All tandem mass spectrometry files were merged into a single file that was used as input for the MASCOT search engine (version 2.2.04, Matrix Science). Spectra were searched against the IPI_Human_3.36 database that contained 69,012 Homo sapiens protein sequences. Further search settings were: trypsin digestion with maximal one missed cleavage, carbamidomethyl (C) and oxidation (M) as fixed and variable modification, respectively, peptide tolerance of 5 ppm with 1+, 2+, and 3+ peptide charges and tandem mass spectrometry tolerance of 0.9 Da. A Mascot cutoff score of 30 was set for peptides and a minimum of two peptides were required for identification of a protein. All protein spots were excised and analyzed by mass spectrometry at least three times.

The second step was done with a biotinylated rabbit anti-mouse or rabbit anti-goat antibody (1:500) and in the final step horseradish peroxidase–labeled streptavidin-biotin complex was applied. The staining was developed with diaminobenzidine and H2O2 as chromogen. The sections were counterstained with haematoxylin and coverslipped with Kaiser’s glycerin. Positive (normal mucosa) and negative (tumor) controls were added with each immunohistochemical assay to ensure reproducibility of the antibodies. Mouse monoclonal antibody U36 and normal mouse IgG were included as positive and negative controls, respectively.

**Clinical validation of keratin 4 and cornulin in a retrospective case-control study**

**Semi-quantitative assessment of protein expression.** In total 222 surgical margins of 46 patients with known follow-up were immunohistochemically stained for keratin 4 and cornulin as described above. The level of protein expression in the immunostained samples was recorded by two independent observers who were blinded for the clinical outcome data, and defined as the percentage of positive cells. After independent scoring, the results were compared and a final consensus score was appointed for each surgical margin. The overall score that was used for statistical analysis in the prognostic study was the averaged mean of all the surgical margins of that particular patient. To enable statistical analysis for ordinal data, percentages were categorized as either high (above median) or low (below median). To explore the prognostic value of the combination of decreased keratin 4 and cornulin expression, the categorical values for both proteins were combined to make a division between normal (high keratin 4 and high cornulin) and aberrant (low keratin 4 and/or low cornulin) protein expression.

**Statistical analysis.** Local disease free survival time was calculated from the date of histologic diagnosis of the primary tumor to the date of local relapse or the date of last follow-up. Kaplan-Meier curves were computed to assess local disease free survival (28), and the log-rank test was done to determine the difference between the curves (29). Variables that showed significant relationships with local disease free survival in the univariate models were included in a multivariate Cox proportional hazards model to determine associations (30). Dysplasia in the surgical margins was included in the statistical analysis, both divided into four (no, mild, moderate, and severe) and into two (no/mild and moderate/severe) categories. Receiver operating characteristic curves were calculated with the original averaged staining percentages to explore the sensitivity and specificity of the immunostaining for the respective biomarkers. The SPSS 15.0.1 statistical package was used for all analyses, and P < 0.05 was considered statistically significant.
Table 2. Clinicopathologic characteristics of the patients in the prognostic validation study

<table>
<thead>
<tr>
<th></th>
<th>Disease free</th>
<th>Local recurrence</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number</td>
<td>23</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>14</td>
<td>15</td>
<td>1.00</td>
</tr>
<tr>
<td>Female</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Average age</td>
<td>57</td>
<td>57</td>
<td>0.85</td>
</tr>
<tr>
<td>Tumor site†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral cavity</td>
<td>19</td>
<td>17</td>
<td>0.72</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>N status†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>12</td>
<td>11</td>
<td>1.00</td>
</tr>
<tr>
<td>N+</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Average no. margins‡</td>
<td>4.6</td>
<td>5.1</td>
<td>0.17</td>
</tr>
<tr>
<td>Dysplasia margins‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>9</td>
<td>1.00</td>
</tr>
<tr>
<td>Mild</td>
<td>7</td>
<td>6</td>
<td>1.00</td>
</tr>
<tr>
<td>Moderate</td>
<td>4</td>
<td>4</td>
<td>1.00</td>
</tr>
<tr>
<td>Severe</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Any grade</td>
<td>15</td>
<td>14</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Fisher’s Exact for categorical variables and †-test for continuous variables.
†Categorical variables.
‡Continuous variables.

Results

Biomarker discovery screen

Figure 1 presents an overview of the tissue characterization and the selection of the samples for the discovery screen by comparative proteomics. Table 1 lists the characteristics of the patients and tissues analyzed. Genetic characterization by allelic loss analysis ensured the selection of normal and precursor epithelium, and confirmed the clonal relation between precursor and tumor tissue. With 2D-DIGE, approximately 1,000 common protein spots could be compared among the tissue samples. In total 40 unique proteins were identified that showed significantly different expression levels among the three tissue types (FDR-corrected P < 0.05, Supplementary Table S1). The identified proteins are involved in various cellular processes such as epidermis and mucosa development, epithelial differentiation, protein synthesis, protein folding, and stress response.

A prominent functional group of proteins with a high expression in the normal mucosa and relatively low expression level in tumors is related to epithelial differentiation. Examples are keratins 4 and 13, cornulin, and small proline-rich protein 3. Figure 2 provides an example of a 2D gel image with the four indicated proteins marked, including the level of expression in each sample. Figure 2 also shows the hierarchical clustering of all tissue samples on basis of the 40 identified proteins that showed significantly different levels of expression. Normal and precursor tissues fall into one cluster, whereas all tumors are classified in a separate cluster.

Initial evaluation of candidate biomarkers

We obtained commercial antibodies against 12 proteins with the largest tumor to normal ratios and most significant FDR-corrected P values. To achieve optimal staining without background signals on formalin-fixed paraffin-embedded tissue sections, dilutions of all primary antibodies were evaluated using varying incubation times and three antigen retrieval protocols. Detailed information on the antibodies and retrieval protocols tested is available on request. Four antibodies directed against keratin 4, keratin 13, cornulin, and small proline-rich protein 3 showed specific staining patterns on sections of formalin-fixed paraffin-embedded specimens of normal mucosa and tumors. The different levels of expression of these proteins were first confirmed on the original patient samples used for the proteomics biomarker discovery screen (data not shown). In addition, we assessed the expression levels in a panel of independent normal tissues, tumors, and severe dysplasias, precursor fields that show multiple genetic changes (16) and have a relatively high risk for malignant progression. As shown in Supplementary Fig. S1, keratin 4 and cornulin discriminated most optimally among the various tissue types and with a large range in expression levels of 0% to 80% and 5% to 70%, respectively. In addition, both proteins were almost undetectable in tumors and therefore selected for subsequent clinical validation in a retrospective case-cohort study.

Clinical validation of keratin 4 and cornulin in a retrospective case-control study

Study cohort. The clinicopathologic characteristics of the case and the control groups are summarized in Table 2. Detailed information on all individual patients is listed in Supplementary Table S2. Our study cohort consisted of 29 men and 17 women with an average age of 57.0 years (range, 38 to 86 years) at the time of diagnosis. In total 23 patients had tumor-free lymph nodes, whereas 23 patients showed lymph node metastases. Twenty-two patients were treated by surgery only and the remaining received adjuvant radiotherapy. The median local disease-free survival time was 5.4 years (range, 0.5 to 13.9 years), during which the 23 case patients developed a local recurrence (n = 19) or second primary tumor (n = 4). For every patient, the surgical margins were divided over 2 to 7 separate tissue blocks, which added up to a total of 222 different surgical margins.

Semiquantitative scoring. All 222 margins were immunostained and independently evaluated by two observers. Supplementary Fig. S2 illustrates the high interobserver accuracy of the quantitative evaluation of keratin 4 and cornulin expression with correlation coefficients of 0.92 and 0.87, respectively. Examples of the staining patterns are shown in Fig. 3, which represents the images of one patient who remained disease free and one patient who developed a local relapse. Overall scores for all individual patients are provided in Supplementary Table S2.

Statistics. The median of the averaged scores of all patients was used to divide the samples into two categories, one with
high expression and one with low expression. For keratin 4 the median was 43.0% and for cornulin 49.5%. Kaplan-Meier curves and the log-rank test revealed that the development of local relapse was significantly associated with a low expression of keratin 4 and cornulin in the surgical margins of the index tumor as depicted in Fig. 4. Hazard ratios calculated by univariate Cox regression analysis, were 3.8 (95% confidence interval, 1.6-9.5; \( P = 0.004 \)) for keratin 4 and 2.7 (95% confidence interval, 1.1-6.5; \( P = 0.031 \)) for cornulin. Receiver operating characteristic curves, based on the continuous averaged percentage data, were computed to explore sensitivity and specificity of the individual markers, displaying an area under the curve of 0.77 for keratin 4 and 0.71 for cornulin (Supplementary Fig. S3).

The keratin 4 and cornulin immunostaining data were combined to form one group of patients with high keratin 4 and high cornulin expression in the surgical margins, reflecting a normal staining pattern, and another group containing all other patients (low keratin 4 and/or low cornulin), reflecting an aberrant staining pattern. This classification showed a very strong association with local disease free survival (\( P < 0.0005 \)) and has a superior prognostic power with a hazard ratio of 8.8 (95% confidence interval, 2.0-37.6; \( P = 0.004 \)).

Grade of dysplasia in the surgical margins did not show a significant association with local disease free survival (Fig. 4), neither when categorized in four classes (no, mild, moderate, or severe) nor in two classes (no/mild versus moderate/severe).

**Discussion**

A simple straightforward immunostaining procedure on the surgical margins of surgically treated HNSCC patients would be an ideal diagnostic tool to predict local relapse if protein biomarkers were available. For this study we used a proteomics platform for biomarker discovery, and we clinically validated the selected biomarkers in a retrospective case-control study.

In most recent proteomics studies done on HNSCC cell lines, tissues, saliva, and secretomes aiming at biomarker discovery for head and neck cancer, normal samples and precancerous tissue (when used) are usually defined by abnormal histology (31–39). To our knowledge, our study is the first to use detailed genetic characterizations to ascertain the selection of strictly normal and precancerous tissues. We selected the 2D-DIGE technology as a comparative proteomics approach as it has been proven to be a valuable platform for the large-scale proteome analysis of multiple paired clinical samples (22, 23, 40), due to the use of a common internal standard and the 4-log dynamic range of protein detection. With only 10 \( \mu \)g of protein per sample, high-quality 2D gel images were acquired and 24 different patient tissues could be compared, which would be unfeasible with any other current quantitative proteomics platform. Notwithstanding, the use of 2D gel electrophoresis restricts the proteome analysis to the approximately 1,000 most abundant proteins, unless cellular subfractions are investigated (40).

We noted that only small differences were found between normal tissue and precancerous tissue. The most logical explanation for this is that we used genetic markers (allelic losses of chromosome arms 3p, 9p, 11q, and 17p) to ensure selection of normal, precancerous, and tumor tissues that are known to detect early changes. Therefore, the precancerous tissue samples were likely at early stages of progression, which explains why they seemed to resemble normal mucosa.

Subsequent validation of our panel of candidate protein biomarkers was restricted by the availability of antibodies that allow immunostaining on formalin-fixed paraffin-embedded samples. Notwithstanding, we could show in the retrospective case-control study that low expression of keratin 4 and cornulin in the resection margins of surgically treated HNSCC patients accurately predicts local relapse and outperforms histopathologic grading. An explanation for this better performance may be that the immunostaining is scored semiquantitatively, which

---

**Fig. 3.** Immunostaining examples of surgical margins in the prognostic validation study. Representative images of immunostained surgical margins of two patients, one who remained disease-free (df; pt 5) and one who developed a local relapse (loc rec; pt 14). Consensus score for every margin of both keratin 4 and cornulin is indicated between brackets. Staining patterns can vary slightly throughout the margin. Averaged staining percentages per patient are given between brackets at the right-hand side of the images.
is more objective and reproducible. The subjectivity and the intra-observer and interobserver variability form a major obstacle in dysplasia grading. The semiquantitative scoring of immunostaining might also tackle the heterogeneity in morphology of the mucosa from different locations within the head and neck region, which also might hinder accurate dysplasia grading. Still, the accuracy is also protein biomarker specific. Keratin 4 showed a much better prognostic performance than keratin 13 (analyzed...
on a subset of samples; data not shown), and seemed better when compared with cornulin. This could be explained by its broad range of expression, which is illustrated in Supplementary Figs. S1 and S2, allowing a larger differential power, a more reliable assessment of the immunostained sections, and a higher correlation between observers.

The candidate protein biomarkers we identified, cornulin and small proline-rich protein 3, as well as keratins 4 and 13, are all proteins involved in differentiation processes. Although it is known that the expression of these proteins, like many others, changes during the carcinogenic process (33, 41–47), no studies have been published that specifically elaborate on the prognostic value of these proteins to predict local relapses when analyzed on the resection margins of surgically treated HNSCC patients. Cornulin and small proline-rich protein 3 belong to the proteins forming the cornified envelope, an important protective barrier of the mucosa and skin. The gene encoding cornulin has recently been identified as one of the “fused genes” in the epidermal differentiation complex locus at chromosome 1q21 (41). The genes in this locus are classified into three groups. One group contains proteins with short tandem repeated peptide motifs that are cross-linked in the formation of the cornified envelope. Small proline-rich proteins belong to this same group. The second group of proteins encoded in the epidermal differentiation complex locus contains proteins with EF domains that bind calcium. The third group, to which cornulin belongs, contains the fused genes that both have EF domains and short tandem repeats.

Besides proteins belonging to the epidermal differentiation complex, both keratins 4 and 13 were identified in our screen. Cytokeratins form the intermediate filaments of the cytoskeleton in epithelial cells. There are multiple so-called acidic and basic cytokeratins, and these form specific dimers that aggregate into filaments (48, 49). Interestingly, keratin 4 and 13 are such a dimer, supporting the validity of our findings.

Due to our stringent inclusion criteria, a relative small number of patients was considered in the retrospective clinical validation study with keratin 4 and cornulin. Nevertheless, the statistical outcomes were highly significant, particularly when the markers were combined in a “normal” (both high) and “aberrant” (one or both low) staining pattern. For this study, we used the (unbiased) median expression to define high and low expression; for clinical applications a cutoff could be set based on the desired sensitivity and specificity (see also Supplementary Fig. S3).

Keratin 4 and cornulin immunostaining of the surgical margins might be used in the decision-making process to determine the surveillance policy for treated HNSCC patients during follow-up. Depending on the medical center, the follow-up surveillance usually encompasses a visit every 6 to 8 weeks during the first year, gradually decreasing to once per year and then stopping at five years. Patients with apparent unresected precancerous fields, indicated by low keratin 4 and cornulin expression in the surgical margins, should remain under long, and perhaps lifelong, frequent surveillance, whereas for patients without these fields, surveillance could remain the same as applied at present. Further, loss of keratin 4 and cornulin expression can be a valuable enrolment criterion for tertiary prevention trials in treated HNSCC patients. Several experimental approaches are being studied and showed promising initial results (50–52).

Our approach can be easily implemented for patients scheduled for tumor resection as surgical margins are routinely obtained from the excised specimen. It remains to be determined whether it could also be exploited for analysis of multiple biopsies taken around the tumor when patients are scheduled for chemoradiation. As the number of biopsies that can be taken is limited, sampling error might easily cause false-negative results. The first recent successes in visualizing these precursor fields by autofluorescence might turn out to be a great aid to select the mucosal regions at risk in these patients (12, 13). Hence, these visualization tools may be a suitable approach when used in combination with biopsy and keratin 4/cornulin immunostaining for the identification of patients with fields who are scheduled for treatment regimens without surgery. Taken together, our proteome analysis has revealed a number of potential biomarkers. Of these, keratin 4 and cornulin, evaluated by a simple immunostaining on routine formalin-fixed paraffin-embedded surgical margin specimens, seem to have large prognostic value to predict local relapse. Large-scale, prospective multicenter trials will have to be carried out to further elucidate their value in the clinic and to study relations with other variables such as the neck status.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References


Downloaded from clincancerres.aacrjournals.org on July 15, 2017. © 2009 American Association for Cancer Research.
Differential Proteomics Identifies Protein Biomarkers That Predict Local Relapse of Head and Neck Squamous Cell Carcinomas
