Morphoproteomic Characterization of Lung Squamous Cell Carcinoma Fragmentation, a Histological Marker of Increased Tumor Invasiveness

Ruben Casanova1, Daniel Xia2,3, Undine Rulle1, Paolo Nanni4, Jonas Grossmann4, Bart Vruyt1, Reto Wettstein2, Rafael Ballester3, Alberto Astolfo6, Walter Weder7, Holger Moch1, Marco Stampalone6,8, Andrew H. Beck9, and Alex Soltermann1

Abstract

Accurate stratification of tumors is imperative for adequate cancer management. In addition to staging, morphologic subtyping allows stratification of patients into additional prognostic groups. In this study, we used an image-based computational method on pan-cytokeratin IHC stainings to quantify tumor fragmentation (TF), a measure of tumor invasiveness of lung squamous cell carcinoma (LSCC). In two independent clinical cohorts from tissue microarrays (TMA: n = 208 patients) and whole sections (WS: n = 99 patients), TF was associated with poor prognosis and increased risk of blood vessel infiltration. A third cohort from The Cancer Genome Atlas (TCGA: n = 335 patients) confirmed the poor prognostic value of TF using a similar human-based score on hematoxylin-eosin staining. Integration of RNA-seq data from TCGA and LC-MS/MS proteomics from WS revealed an upregulation of extracellular matrix remodeling genes such as overall architecture or cell and nuclear pleomorphism, which could serve as a metric for tumor aggressiveness, which could serve to stratify LSCC into prognostic groups. However, there is still no well-established grading system for LSCC (4).

Tumor invasion is supported by de novo formation of desmoplastic stroma, which provides not only physical support to cancer cells but also favors tumor expansion and invasion as a net effect of tumor-stroma inter-talk (5). LSCC invasion is histologically characterized by tumor clusters of variable sizes surrounded by such specialized stroma. This is reflected by an apparent fragmentation of both central and peripheral portions of the tumor mass. The quantification of such tumor fragments on whole histologic sections may be used as a metric for tumor aggressiveness, which could serve to stratify LSCC into prognostic groups.

In this study, we performed an image-based computational analysis for unbiased tumor fragmentation (TF) quantification and extended the method with a human-based scoring system relevant for pathologists. In addition, we integrated tumor histology with molecular data to evaluate the biological processes associated to this morphologic feature. A retrospective study on three independent clinical cohorts was performed to assess the potential prognostic relevance of TF as novel grading parameter for LSCC.

Patients and Methods

Patient cohorts

In this retrospective study, consecutive patients with surgical resection of their primary lung SCC at the University Hospital...
Cancer Research 2586
chronic or metachronic second primary tumor, in particular p40, p63, CK7, and synaptophysin antibodies. Cases with mixed or unclear histology were excluded. Patients with synchronous or metachronic second primary tumor, in particular head and neck SCC, with neoadjuvant chemo-/radiotherapy or with overall survival (OS) <1-month after surgery were also excluded. The tissue microarray (TMA) cohort consisted of 233 lung SCC patients between January 1993 and December 2002, as described previously (6). Patients were further excluded whether the quality of the pan-cytokeratin IHC was insufficient for automatic image analysis on both cores, for example, due to presence of noncancerous tissue, more than 80% stroma or major cutting artefacts. In total 208 patients met inclusion criteria. Adjuvant treatment was administered to 75 (36%) patients, from which 62 (82%) received radiotherapy, 8 (11%) chemotherapy, and 5 (7%) both. The WS cohort consisted of nonredundant 167 patients between January 2003 and February 2010 having complete clinical data. After revision for exclusion criteria, 99 patients were included. Adjuvant therapy was administered to 30 (30%) of the patients, from which 6 (20%) received radiotherapy, 20 (67%) chemotherapy, and 4 (13%) both. The Cancer Genome Atlas (TCGA) clinical data were downloaded from cBioPortal for Cancer Genomics, Memorial Sloan Kettering Cancer Center (7) for 504 lung SCC patients, from which 62 (82%) received radiotherapy, 8 (11%) chemotherapy, and 5 (7%) both. The WS cohort consisted of nonredundant 167 patients between January 2003 and February 2010 having complete clinical data. After revision for exclusion criteria, 99 patients were included. Adjuvant therapy was administered to 30 (30%) of the patients, from which 6 (20%) received radiotherapy, 20 (67%) chemotherapy, and 4 (13%) both. The Cancer Genome Atlas (TCGA) clinical data were downloaded from cBioPortal for Cancer Genomics, Memorial Sloan Kettering Cancer Center (7) for 504 lung SCC patients. After exclusion due to OS<1 month, neoadjuvant chemotherapy or insufficient image quality, 335 patients remained. Thirty-one (9%) patients had a documented adjuvant therapy whereby 3 (10%) received radiotherapy, 21 (68%) targeted therapy, and 7 (22%) both. Our study was approved by the Ethical Commission of the Canton of Zurich under reference number KEK ZH-Nr. 29-2009/14.

Tissue microarray construction and IHC
For the TMA cohort, two representative paraffin blocks with tumor were selected. One tissue core of 600-µm diameter was taken from each tissue block and transferred into a recipient block. For the whole sections cohort, two blocks representing best the overall tumor morphology were selected and 2-µm-thick sections were cut from each block. IHC was performed on an automated platform (Ventana Medical Systems) using mouse monoclonal anti-human cytokeratin AE1/AE3 (M3515, DAKO, dilution 1:50), rabbit polyclonal Periostin (Abcam, 1:1,000) and mouse monoclonal Versican (2B1, Seikagaku, 1:500). Detection was completed with a respective secondary antibody and the OptiView DAB Kit (Ventana Medical Systems).

Image acquisition
Immunohistochemically stained sections were scanned with a high resolution whole-slide scanner (Nanozoomer Digital Pathology) using a ×40 objective with spatial resolution of 0.23 µm/pixel. TMA and whole sections images were analyzed with a spatial resolution of 1.84 µm/pixel and 9.2 µm/pixel, respectively. Whole sections were further annotated by a surgical pathologist (A. Soltermann) to select tumor tissue. The surrounding nontumor lung tissue was excluded from the analysis. A formalin-fixed cylindrical tissue sample of 1-cm diameter from a pT3 lung SCC was imaged by X-ray micromotography at the beamline for tomographic microscopy and coherent radiology experiments (TOMCAT), Paul Scherrer Institute (PSI, Würenlingen, Switzerland). Sample was scanned using a microscope equipped with a PICO-Edge camera-mounted with a ×10 objective for a spatial resolution of 0.63 µm/pixel using propagation-based X-ray phase-contrast CT, as described previously (8).

Image processing
Automatic morphometric analysis was performed using Fiji (9). Color-based segmentation allowed separating the tumor tissue (brown) from its surrounding stroma (blue/gray). Color threshold values were previously validated by the pathologist. A similar approach for tumor segmentation by cytokeratin stained sections has been described for colon carcinoma (10). The segmentation algorithm included mean shift and median filtering, color thresholding using the $L \times a \times b$ color space and particle size filtering with a minimum size threshold of 100 µm$^2$, considered as noise. The resulting segmentation masks were used for automated scoring of tumor segmentation.

Tumor fragmentation scoring
Tumor fragments refer to tumor clusters completely separated from each other on a two-dimensional plane by intervening desmoplastic stroma. Tumor fragmentation was automatically scored on the pan-CK stainings of both TMA and WS by counting the total number of disconnected particles larger than 800 µm$^2$ (circa >5 cells) for each segmentation mask. TF scores were summed up over two images when available or duplicated for cases with single images. In addition, 88/99 WS patients were also represented on a secondary TMA, from which 77 had sufficient image quality for automated TF scoring. TF scores from this subset were used to address scoring heterogeneity between TMA cores and whole sections for the same tumor. Human-based TF scoring was done by one observer (obs1) on H&E-stained tissues. On TMA cores, TF was scored using all available magnifications and on the area of highest fragmentation for the TCGA and WS cohorts, under a magnification of ×50. This corresponded to 1,920 × 1,036 pixels (3.5 × 1.9 mm) for TCGA image frames and to 4 mm FOV for WS with light microscopy. Two additional observers (obs2-3) evaluated 20% of the TCGA images ($n = 67$) to address interobserver variability.

Sample preparation for mass spectrometry
A total of 48 samples from the whole sections cohort were selected for mass spectrometry based on the availability of at least 1 tumor paraffin block containing >80% carcinoma epithelium per total tissue surface. Two 20-µm-thick microtome cuts were deparaffinized in xylene and washed with 96% ethanol. Samples were suspended in 120 µL of SDS buffer (4% SDS, 100 mM/L Tris/HCl pH 8.2, 0.1 mol/L DTT–dithiothreitol) and boiled at 95°C for 20 minutes followed by 2 hours at 80°C and processed with high intensity focused ultrasound for 10 minutes, setting the ultrasonic amplitude to 65%. Protein concentration was determined using the Qubit Protein Assay Kit (Life Technologies). For each sample, 20 µg protein were taken for on-filter digestion using an adaptation of the filter-aided sample preparation protocol (11). Briefly, proteins were diluted in 200 µL of U1 buffer (Urea 8 mol/L in 100 mM/L Tris/HCl pH 8.2), loaded on Ultrapel 30,000 MWCO
centrifugal unit (Amicon Ultra, Merck) and centrifuged at 14,000 x g. SDS buffer was exchanged by one centrifugation round of 200 µL UT buffer. Alkylation of reduced proteins was carried out by 5 minutes incubation with 100 µL iodoacetamide 0.05 mol/L in UT buffer, followed by three 100 µL washing steps with UT and three 100 µL washing steps with NaCl 0.5 mol/L. Finally, proteins were on-filter digested using 120 µL of 0.05 triethylammonium bicarbonate buffer (pH 8) containing trypsin (Promega) in ratio 1:50 (w/w). Digestion was performed overnight in a wet chamber at room temperature. After elution, the solution containing peptides was acidified to a final 0.1% TFA, 3% acetonitrile concentration. Peptides were desalted using self-packed C18 Stage-Tips, dried, and resolubilized in 15 µL of 3% acetonitrile, 0.1% formic acid for MS analysis.

LC-MS/MS analysis
Mass spectrometry analysis was performed on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) coupled to an Eksigent-Nano-HPLC system (Sciex). Solvent composition at the two channels was 0.1% formic acid for channel A and 0.1% formic acid, 99.9% acetonitrile for channel B. For each sample 2 µL of peptides were loaded on a self-made column (75 µm x 150 mm) packed with reverse-phase C18 material (ReproSil-Pur 120 C18-AQ, 1.9 µm, Dr. Maisch GmbH, Ammerbuch, Germany) and eluted at a flow rate of 300 nL/minute by a gradient from 3 to 25% B in 65 minutes, 35% B in 5 minutes and 97% B in 5 minutes. Samples were acquired in a randomized order. The mass spectrometer (Tune page v1.1) was configured to fragment peptide ions (default filter with 1.6 m/z isolation window and fragmented by higher-energy collisional dissociation (HCD) using a normalized collision energy of 30. Fragments were calibrated on m/z 371.1010 and 445.1200. At 10 ppm. The samples were acquired using internal lock mass calibration on m/z 371.1010 and 445.1200. Three mass spectra were exported into a Mascot generic filter with 1.6 m/z isolation filter digested using trypsin (Promega) and one missed cleavage was allowed. Carbamidomethylation of cysteine was set as fixed modification, while oxidation (M) and deamidation (N, Q) were set as variable. The resulting .dta file was loaded into Scaffold v4.1.1 (Proteome Software) and filtered at peptide and protein FDR were set to 1% and 9%, respectively. Finally, the Scaffold Spectrum Report was imported back into Progenesis. For quantification, all proteins identified with at least two peptide ions were assessed, resulting in an estimated protein FDR of 0.5%. Proteins were grouped with Progenesis and the relative quantification using Hi-N (N = 3) peptides was used. For protein quantification, the average of the normalized peptide intensities in peak picking was used to calculate the protein quantification.

Protein identification and label-free protein quantification
Protein label-free quantification was performed using the software Progenesis QI for proteomics (v.4.0.4265.42984) software (Nonlinear Dynamics), using as reference for alignment the raw-file of the pool sample. Normalization was kept with default settings. From each Progenesis peptide ion (default sensitivity in peak picking) a maximum of the top five tandem mass spectra were exported into a Mascot generic file using charge deconvolution and deisotoping option and a maximum number of 200 peaks per MS/MS. This Mascot generic file (mgf) was searched with Mascot 2.4.3.3 (Matrix Science Ltd.) against the forward Uniprot database for Homo sapiens, concatenated to a reversed decoyed FASTA database and 260 common mass spectrometry protein contaminants. The parameters for precursor tolerance and fragment ion tolerance were set to ± 10 ppm and ± 0.6 Da. Enzyme was set to trypsin and one missed cleavage was allowed. Carbamidomethylation of cysteine was set as fixed modification, while oxidation (M) and deamidation (N, Q) were set as variable. The resulting .dta file was loaded into Scaffold v4.1.1 (Proteome Software) and filtered at peptide and protein FDR were set to 1% and 9%, respectively. Finally, the Scaffold Spectrum Report was imported back into Progenesis. For quantification, all proteins identified with at least two peptide ions were assessed, resulting in an estimated protein FDR of 0.5%. Proteins were grouped with Progenesis and the relative quantification using Hi-N (N = 3) peptides was used. For protein quantification, the average of the normalized peptide intensities in peak picking was used to calculate the protein quantification.

Table 1. Patient cohorts

<table>
<thead>
<tr>
<th>Patient cohorts</th>
<th>TMA</th>
<th>WS</th>
<th>TCGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>66</td>
<td>65</td>
<td>68</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>42</td>
<td>19</td>
<td>85</td>
</tr>
<tr>
<td>Male</td>
<td>166</td>
<td>80</td>
<td>252</td>
</tr>
<tr>
<td>pT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>43</td>
<td>24</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>107</td>
<td>28</td>
<td>215</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>pN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>110</td>
<td>47</td>
<td>209</td>
</tr>
<tr>
<td>1</td>
<td>73</td>
<td>31</td>
<td>88</td>
</tr>
<tr>
<td>2-3</td>
<td>25</td>
<td>21</td>
<td>35</td>
</tr>
<tr>
<td>ha/b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>196</td>
<td>95</td>
<td>284</td>
</tr>
<tr>
<td>la/b</td>
<td>12</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>71</td>
<td>24</td>
<td>165</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>35</td>
<td>158</td>
</tr>
<tr>
<td>III</td>
<td>50</td>
<td>36</td>
<td>57</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>121</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>87</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Size (cm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>3.9</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Vessel infiltration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>86</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjuvant therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>133</td>
<td>69</td>
<td>304</td>
</tr>
<tr>
<td>Yes</td>
<td>75</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>TF score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer (HIC)</td>
<td>7</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>Human (H&amp;E)</td>
<td>6</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>42</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>OS (months)</td>
<td>169</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>Median (CI)</td>
<td>59</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Follow-up max</td>
<td>169</td>
<td>137</td>
<td>154</td>
</tr>
<tr>
<td>Total patients (n = 642)</td>
<td>208</td>
<td>99</td>
<td>335</td>
</tr>
</tbody>
</table>

NOTE: Summary of clinicopathologic parameters, including OS and RFS for TMA, WS and TCGA datasets. Abbreviations: CI, 95% confidence interval; TF computer/human, tumor fragmentation scores respectively computed or assessed by eye.
abundance from the most intense N peptide ions of each protein group was calculated individually for each sample. This generated the normalized quantitative protein abundance. Protein levels were further log₂ transformed for statistical testing. Differentially expressed proteins were identified by Significance Analysis of Microarrays as described (12). TF scores were log₂ transformed and addressed as quantitative response using the standard regression method on median-centered protein levels. Proteins significantly correlated with TF scores (FDR < 0.05) were considered for further analysis. Gene ontology enrichment analysis was performed using WEB-based Gene Set Analysis Toolkit (13) using default settings. All identified proteins (n = 2,614) were used as background (Supplementary Table S1). Proteomics quantification data has been submitted to The PRIDE Proteomics Identifications (PRIDE) database (project accession: PXD006132; ref. 14).

Gene ontology enrichment networks and pathway analysis for TCGA mRNA-Seq data
Analysis-ready standardized TCGA mRNA-Seq data were downloaded from the Broad GDAC Firehose stddata__2015_11_01 run. Two groups showing clearly distinguishable TF profiles on TCGA H&E stainings, consisting of patients with the 25% lowest (TF < 6) and 25% highest (TF > 18) scores were selected for mRNA differential expression analysis. For statistical testing, the R-package EdgeR was used to compare them. In total mRNA-Seq data for 20531 genes were available. Genes with a minimum of 1 count-per-million, in at least half of the samples, were selected (n = 14151). Default parameter settings were applied. Candidates having a FDR < 0.05 were included in the gene ontology enrichment analysis. Gene ontology enrichment map was generated with the software Cytoscape (15) using BINGO (16) and Enrichment Map (17) plugins. Genes upregulated in

Figure 1.
Computer-based morphometric analysis of tumor fragmentation. A, Left, histologic whole section showing a squamous cell carcinoma of the lung stained with pan-cytokeratin IHC (brown) and its surrounding stroma (blue-gray counterstain). Right, color-based segmentation showing tumor fragments of different sizes labeled on a 32-color scale. Approximate sizes are: single-cell, 200 μm²; buds < 800 μm²; nests < 3,000 μm²; larger fragments > 3,000 μm². Number of epithelial elements = 32, size mean = 44,000 μm², size median = 14,000 μm²; scale bar, 500 μm. B, Left, X-ray microtomography 2D reconstructed slice of a lung SCC example. Arrows, epithelial tumor fragments. t*, tumor; n*, necrosis; s*, stroma. Right, 3D rendering of the tumor microarchitecture imaged by X-ray microtomography.
high and low TF tumors, respectively, were uploaded separately in BINGO. Enrichment analysis was performed with the default hypergeometric test using a significance level <0.05 after Bonferroni correction for multiple testing. Enrichment map was generated using BINGO outputs for each tumor morphotype with the following parameter: $P$ value cutoff <0.001, FDR <0.05, Jaccard coefficient = 0.25. The generated network is represented by nodes (gene sets), edges (mutual overlap) and colors (tumor morphology groups: blue, low TF; red, high TF). Pathway analysis was done using the KEGG database (18) through WEB-based GEne SeT AnaLysis Toolkit.

Data interpretation and statistical analysis
All statistical analyses were performed on SPSS version 22 software (SPSS Inc.). Relapse-free survival was measured from the date of surgery to the date of documented relapse or death as described (19). RFS was assessed only for patients who presented no evidence of remaining tumor (incomplete resection and/or metastases) after surgery. Kaplan–Meier survival curves were evaluated using log-rank tests using TF scores dichotomized at the median. Hazard ratios were assessed by Cox regression. Clinical correlations were addressed using the nonparametric Spearman rank correlation test. The association of clinical parameters with survival was computed by univariate Cox regression. In addition, clinically relevant parameters were introduced into multivariate Cox regressions. $P$ values <0.05 were considered significant.

Results
LSCC shows a variable degree of tumor fragmentation
TF scores were automatically computed as the total number of tumor fragments in pan-cytokeratin stained TMAs of 208 LSCC patients and whole sections (WS) of 99 patients (Table 1). The size of tumor fragments within individual tumors varied considerably (Fig. 1A). By using X-ray microtomography, we saw that tumor fragments represent projections of interconnected epithelial branches from a three-dimensional perspective (Fig. 1B). The distribution of the computed TF scores was comparable in the TMA and WS cohorts (Supplementary Fig. S1A–S1B, left). The assessment of a subset of 77 patients from WS with matching TMA cores from a secondary TMA, showed a positive correlation of TF scores between whole sections and tumor cores from the same tumor (correlation coeff. = 0.484; $P < 0.001$). However, due to intratumor heterogeneity, the WS analysis allows a more reliable assessment of TF scores as demonstrated by the correlation analysis of computed TF scores from two tumor regions (Supplementary Fig. S1A–S1B, right). We further tested the applicability of TF scoring by human eye using H&E-stained images of the TMA and WS cohorts, as well as in an external TCGA LSCC cohort with similar clinical characteristics. The distribution of TF scores was comparable between the computed and human-based evaluations (Supplementary Fig. S1C–S1E). Histologic examples of tumors with high and low fragmentation from TMA, WS, and TCGA cohorts are shown in Fig. 2.
Figure 3.
Overall survival curves based on tumor fragmentation. Kaplan-Meier curves stratified by high/low TF, dichotomized at the median. CI, 95% confidence interval. Computed TF scores were quantified on pan-cytokeratin-stained tissue. Human-based scores were performed on H&E-stained tissue.

Tumor fragmentation is associated with increased invasiveness and worse outcome

Clinicopathologic correlations showed that high TF was consistently associated with vessel infiltration on both TMA and WS (Supplementary Table S2). Perineural and mediastinal invasion were further evaluated on WS and positively correlated with TF as well ($P = 0.001, P = 0.004$, respectively). Survival analysis showed that high TF is a poor prognostic factor for OS and was confirmed using the external TCGA cohort with human-based scores (Fig. 3). Increasing TF scores had a significantly poor impact on OS as showed by both univariate and multivariate analysis (Table 2). Analysis for relapse-free survival showed comparable results for the three patient cohorts (Supplementary Table S3). Human-based scores showed similar results for WS (Supplementary Table S4). By evaluating the subset of 77 patients from WS with matching TMA cores, we observed a decrease of prognostic power using TMA cores (OS/RFS: $P = 0.060/0.152$), in comparison with whole sections (OS/RFS: $P < 0.001/<0.001$), possibly due to the higher impact of intratumor heterogeneity on the TMA. TF scores using different minimal fragment size thresholds were also evaluated on WS. The prognostic relevance of TF was similar for a wide range of size cutoffs (5–100 cells) for both OS and RFS (Supplementary Fig. S2).

Tumor fragmentation is associated with changes in extracellular matrix

We selected TCGA patients with the upper and lower quartiles of TF scores to identify molecular characteristics associated with TF morphotypes and identified 910 genes associated with the high TF. Enriched biological processes involved tissue development, extracellular matrix (ECM) organization, and cell adhesion processes (Supplementary Fig. S3). KEGG analysis notably showed an upregulation of pathways involved in ECM–receptor interaction, focal adhesion and protein digestion. In parallel, mass spectrometry–based label-free quantification identified 154 proteins significantly associated with increased TF (Supplementary Table S5), including extracellular (20%) and cytoskeletal (15%) components. In accordance with mRNA-Seq analysis, we observed an enrichment of genes involved in developmental processes.

In contrast, low TF was associated with 554 genes involved in diverse metabolic pathways, including xenobiotics metabolism.

<table>
<thead>
<tr>
<th>Table 2. Association of TF scores with overall survival</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OS</strong> (n = 208)</td>
</tr>
<tr>
<td><strong>TMA</strong></td>
</tr>
<tr>
<td>TF (low/high) computer IHC</td>
</tr>
<tr>
<td>TF (continuous) computer IHC</td>
</tr>
<tr>
<td>Stage (I, II, III, IV)</td>
</tr>
<tr>
<td>Grade (1,2,3)</td>
</tr>
<tr>
<td>Vessel infiltration</td>
</tr>
<tr>
<td><strong>WS</strong> (n = 99)</td>
</tr>
<tr>
<td>TF (low/high) computer IHC</td>
</tr>
<tr>
<td>TF (continuous) computer IHC</td>
</tr>
<tr>
<td>Stage (I, II, III, IV)</td>
</tr>
<tr>
<td>Grade (1,2,3)</td>
</tr>
<tr>
<td>Vessel infiltration</td>
</tr>
<tr>
<td><strong>TCGA</strong> (n = 335)</td>
</tr>
<tr>
<td>TF (low/high) human H&amp;E</td>
</tr>
<tr>
<td>TF (continuous) human H&amp;E</td>
</tr>
<tr>
<td>Stage (I, II, III, IV)</td>
</tr>
<tr>
<td>Grade (1,2,3)</td>
</tr>
<tr>
<td>Vessel infiltration</td>
</tr>
</tbody>
</table>

NOTE: Survival analysis by Cox univariate and multivariate regressions.
Abbreviations: CI, 95% confidence interval; computer, automatic scoring on pan-cytokeratin IHC-stained tissue; human, fragmentation scored by eye; TF (low/high), scores dichotomized at the median (categorical); TF (continuous), score as continuous variable.
(Supplementary Fig. S3), and 205 proteins representing in majority intracellular (70%) and nuclear proteins (46%). Biological nodes associated with decreasing TF were mostly related to hormone regulation processes.

We validated the expression of periostin and versican, two ECM proteins associated with TF in both RNA and protein analyses, by IHC. The expression of both proteins in the tumor and stromal compartments was significantly associated with TF (Fig. 4).

**Discussion**

LSCC is histologically characterized by a variable degree of keratinization and/or the formation of intercellular bridges between tumor epithelial cells.Collective cell migration is predominant in LSCC as other invasion modes such as EMT (epithelial–mesenchymal transition, fibroblastoid) or amoeboid are less observed probably due to the tight desmosomal adhesions (20, 21). This invasion pattern results in an apparent network of epithelial branches in 3D, which is histologically depicted by a high variability of epithelia sizes in 2D. In this study, we used an image-based computational method to quantify such tumor fragments, hypothesizing that high fragmentation is a trait of increased tumor invasiveness. In support of this hypothesis, we showed that TF was associated with increased blood vessel, mediastinal and perineural invasion and worse patient outcome. Consistently, molecular analysis showed an upregulation of processes involved in ECM remodeling and focal adhesion, major characteristics of increased cellular motility (21–24). The association of TF with two identified ECM proteins, periostin and versican, was validated by IHC. Periostin is a secreted ECM protein observed in the desmoplastic stroma of a variety of cancers and known to promote cell invasion (25–28). It is also involved in cardiac remodeling after myocardial infarction (29). Versican is a major proteoglycan of the extracellular matrix upregulated in several tumor types, including lung cancer (30) and has been notably associated with poor prognosis in NSCLC (31). It is also shown to favor tumor metastasis in Lewis lung carcinoma cell lines (32). In contrast, tumors with low TF showed an enrichment of diverse metabolic and xenobiotic processes notably contributing to the regulation of a number of chemotherapeutic drugs (33–35).

TNM staging is the only established system for predicting LSCC prognosis, whereas the value of tumor grading along the keratinization qualifier remains unclear. In this study, we evaluated the clinical relevance of LSCC stratification based on TF. Survival analysis showed that high TF is a prognostic marker associated with poor survival, independent from stage. This could be a useful additional grading parameter, suggesting tumor invasiveness. Alternative histologic parameters such as tumor budding have been proposed as various measures of tumor invasion in cancers notably in colorectal carcinoma (36–38). In NSCLC, single cells, tumor buds (<5 cells) or nests (≤15 cells) together with stroma thickness have also shown an impact on patient’s survival (2, 3, 39, 40), although small invasive tumor clusters are less frequent in LSCC. As shown in Fig. 1, the typical median and average LSCC epithelia clusters size in 2D is at least one order of magnitude higher than single cells and tumor buds. Furthermore, studies of the 3D tumor microarchitecture suggest that single cells may be part of tumor buds, which in turn may belong to larger tumor nests or branches (41).
Alternatively, current progresses in digital pathology have led to automated identification of prognostic features on histologic sections, notably for NSCLC (42). However, such methods are mostly based on the quantification of cellular (mostly nuclear and cytoplasmic) features. Our computational method is focused on a higher order feature of the tumor epithelia, related to tumor invasion patterns that can be favorably translated into a human-based scoring system on H&E tissue staining. The proposed computational image-based analysis allowed unbiased scoring of tumor fragments, but has nevertheless intrinsic limitations. The main drawback is that normal epithelial lung structures as well as necrotic residues are also immune-reactive with pan-cytokeratin. To minimize false positive tumor fragments, adjacent normal tissue was covered and a minimal size threshold for tumor fragments (>800 μm²) was set. Finally, the clinical relevance of TF in the context of adjuvant therapy could not be addressed in this study, because of the high heterogeneity of treatment modalities across cohorts.

In conclusion, we have shown using an image-based computational approach, that high fragmentation of LSCC is a histologic trait associated with increased aggressiveness. In addition, the integration of molecular data showed an upregulation of proteins favoring extracellular matrix remodeling and focal adhesion, supporting the increased invasive potential of tumors with such high fragmentation. The proposed histologic parameter is an independent unfavorable prognostic marker that could be envisaged as new grading parameter for LSCC.

Disclosure of Potential Conflicts of Interest
A.H. Beck is the founder of PathAI Inc. No potential conflicts of interest were disclosed by the other authors.

References

Authors’ Contributions
Conception and design: R. Casanova, H. Moch, A. Soltermann
Development of methodology: R. Casanova, D. Xia, U. Rulle, P. Nanni, M. Stampanoni, A. Soltermann
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. Casanova, D. Xia, P. Nanni, B. Vrugt, A. Astolf, A. Soltermann
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Casanova, P. Nanni, J. Grossmann, R. Wettstein, R. Ballester, A.H. Beck, A. Soltermann
Writing, review, and/or revision of the manuscript: R. Casanova, D. Xia, U. Rulle, J. Grossmann, B. Vrugt, A. Astolf, W. Wieder, H. Moch, A.H. Beck, A. Soltermann
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): U. Rulle, H. Moch, A. Soltermann
Study supervision: H. Moch, A. Soltermann

Acknowledgments
We would like to acknowledge Dr. Karina Silina (Institute of Experimental Immunology), Prof. Bernd Bodenmüller (Institute of Molecular Life Sciences), Prof. Renato Pajarola (Department of Informatics), Dr. Dimitri Korol (Cancer Registry), Dr. Peter Schraml (Institute of Pathology and Molecular Pathology) for their critical input.

Grant Support
This work was supported by grants from the Swiss Cancer League (reference number F-87701-31-01) and the Swiss National Science Foundation SystemsX (reference number M-87704-01-02 to A. Soltermann). 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 31, 2016, revised January 13, 2017; accepted March 8, 2017, published OnlineFirst March 31, 2017.
in colon cancer and proposed definitions for future trials. J Natl Cancer Inst
21. Friedl P, Alexander S. Cancer invasion and the microenvironment plas-
24. Friedl P, Wolf K. Tube travel: the role of proteases in individual and
25. Gillan L, Matei D, Fishman DA, Gerbin CS, Karlan BY, Chang DD. Periostin
secreted by epithelial ovarian carcinoma is a ligand for alpha(V)beta(3)
and alpha(V)beta(5) integrins and promotes cell motility. Cancer Res
26. Yan W, Shao R. Transduction of a mesenchyme-specific gene periostin into
293T cells induces cell invasive activity through epithelial–mesenchymal
27. Butcher JT, Norris RA, Hoffman S, Mjaatvedt CH, Markwald RR. Periostin
promotes atrioventricular mesenchyme matrix invasion and remodeling
mediated by integrin signaling through Rho/PI 3-kinase. Dev Biol
Suter T, Aebersold R, et al. N-glycoprotein profiling of lung ade-
carcinoma pleural effusions by shotgun proteomics. Cancer Res
2008;68:1124–33.
Periostin is essential for cardiac healing after acute myocardial infarction. J
2B1 antigen characteristically expressed on extracellular matrices of human
malignant tumors is a large chondroitin sulfate proteoglycan, PG-M/
31. Piirinen R, Leinonen T, Bohm J, Johansson R, Ropponen K, Kumpulai-
nen E, et al. Versican in non-small cell lung cancer: relation to hyalur-
onan, clinicopathologic factors, and prognosis. Hum Pathol 2005;
36:44–50.
Carcinoma produced factors activate myeloid cells through TLR2 to stim-
Periostin is essential for cardiac healing after acute myocardial infarction. J
2B1 antigen characteristically expressed on extracellular matrices of human
malignant tumors is a large chondroitin sulfate proteoglycan, PG-M/
Morphoproteomic Characterization of Lung Squamous Cell Carcinoma Fragmentation, a Histological Marker of Increased Tumor Invasiveness

Ruben Casanova, Daniel Xia, Undine Rulle, et al.

Cancer Res 2017;77:2585-2593. Published OnlineFirst March 31, 2017.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-16-2363

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2017/03/30/0008-5472.CAN-16-2363.DC1

Cited articles
This article cites 41 articles, 16 of which you can access for free at:
http://cancerres.aacrjournals.org/content/77/10/2585.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.