Quantitative proteomics identify Tenascin-C as a promoter of lung cancer progression and contributor to a signature prognostic of patient survival

The MIT Faculty has made this article openly available. *Please share* how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Gocheva, Vasilena et al. “Quantitative Proteomics Identify Tenascin-C as a Promoter of Lung Cancer Progression and Contributor to a Signature Prognostic of Patient Survival.” Proceedings of the National Academy of Sciences 114, 28 (June 2017): E5625–E5634 © 2017 The Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/PNAS.1707054114">http://dx.doi.org/10.1073/PNAS.1707054114</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences (U.S.)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Tue Feb 05 21:12:47 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/114911">http://hdl.handle.net/1721.1/114911</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Quantitative proteomics identify Tenascin-C as a promoter of lung cancer progression and contributor to a signature prognostic of patient survival

Vasilena Goecheva, Alexandra Naba, Arjun Bhutkar, Talia Guardia, Kathryn M. Miller, Carman Man-Chung Li, Talya L. Dayton, Francisco J. Sanchez-Rivera, Caroline Kim-Kiselak, Noor Jailkhani, Monte M. Winslow, Amanda Del Rosario, Richard O. Hynes, and Tyler Jacks

The extracellular microenvironment is an integral component of normal and diseased tissues that is poorly understood owing to its complexity. To investigate the contribution of the microenvironment to lung fibrosis and adenocarcinoma progression, two pathologies characterized by excessive stromal expansion, we used mouse models to characterize the extracellular matrix (ECM) composition of normal lung, fibrotic lung, lung tumors, and metastases. Using quantitative proteomics, we identified and assayed the abundance of 113 ECM proteins, which revealed robust ECM protein signatures unique to fibrosis, primary tumors, or metastases. These analyses indicated significantly increased abundance of several S100 proteins, including Fibronectin and Tenascin-C (Tnc), in primary lung tumors and associated lymph node metastases compared with normal tissue. We further showed that Tnc expression is repressed by the transcription factor Nkx2-1, a well-established suppressor of metastatic progression. We found that increasing the levels of Tnc, via CRISPR-mediated transcriptional activation of the endogenous gene, enhanced the metastatic dissemination of lung adenocarcinoma cells. Interrogation of human cancer gene expression data revealed that high TNC expression correlates with worse prognosis for lung adenocarcinoma, and that a three-gene expression signature comprising TNC, S100A10, and S100A11 is a robust predictor of patient survival independent of age, sex, smoking history, and mutational load. Our findings suggest that the poorly understood ECM composition of the fibrotic and tumor microenvironment is an underexplored source of diagnostic markers and potential therapeutic targets for cancer patients.

extracellular matrix | Tenascin-C | lung cancer | quantitative proteomics | tumor microenvironment

Tumor progression is a function of the combined effects of genetic and epigenetic changes in cancer cells, as well as the influence of the tumor microenvironment. Both cellular and noncellular stromal components in the tumor microenvironment have been shown to affect growth of the primary tumor, progression to metastasis, and response to various anticancer agents (1). Characterization of the tumor-associated stroma in many different cancer types has revealed an extensive array of cell types and structural components that directly affect tumorigenesis (1). In particular, the noncellular component of the tumor microenvironment, the extracellular matrix (ECM), has emerged as an important regulator of cancer development (2). The ECM is a complex network of secreted macromolecules that surrounds most cells within tissues and contributes to the establishment and maintenance of tissue architecture (3). Along with providing the structural foundation for tissue function and mechanical integrity, the ECM has various other roles that make it relevant to the pathology of cancer. The ECM serves as a substrate for cell attachment and guides the migration of cells along its fibers. Moreover, the ECM affects proliferation by acting as a reservoir for growth factors and chemokines, and regulates the presentation of these molecules to their corresponding receptors (3). Thus, it is not surprising that the ECM plays a critical role during tumor cell invasion and metastasis (4, 5). In fact, different adhesive characteristics of the ECM have been suggested to control key steps of the metastatic process, including tumor cell intravasation, extravasation, and metastatic colonization (2).

Although the ECM is a major component of the tumor microenvironment, our understanding of the changes in ECM composition during cancer progression remains incomplete, largely due to technical challenges in comprehensively assaying the ECM in a sensitive and unbiased manner. In this study, we focused on analyzing the ECM changes associated with lung cancer, the most prevalent and deadly cancer type worldwide. Non–small-cell lung cancers account for 83% of all lung cancers, with lung adenocarcinomas the most common subtype. Lung cancer accounts for more deaths than any other cancer in both men and women (6). Several clinical studies have noted that increased expression of certain ECM components, such as versican and hyaluronic acid, can predict survival in patients with lung adenocarcinoma. Three-gene expression signature comprising TNC, S100A10, and S100A11 can predict survival in patients with lung adenocarcinoma. These factors could serve as disease markers that could be exploited for better diagnosis of lung cancer, and their future study could be used to inform the design of more potent treatments for patients.

**Significance**

Quantitative mass spectrometric profiling of the extracellular matrix composition of normal lung, fibrotic lung, primary lung tumors, and lung metastases to the lymph nodes uncovered specific signatures distinguishing these tissues. CRISPR/Cas9-mediated gene activation of one of the identified factors, Tenascin-C (Tnc), showed that this protein plays a role in mediating lung adenocarcinoma metastasis. Tnc expression is repressed, directly or indirectly, by the transcription factor Nkx2-1. Bioinformatic analysis shows that expression of three matricellular factors (TNC, S100A10, and S100A11) can predict survival in patients with lung adenocarcinoma. These factors could serve as disease markers that could be exploited for better diagnosis of lung cancer, and their future study could be used to inform the design of more potent treatments for patients.


**Reviewers:** L.M., Pancreatic Cancer Action Network; and G.O., INSERM.

The authors declare no conflict of interest.

Data deposition: The raw mass spectrometry data have been deposited to the Proteome Xchange Consortium via the PRIDE partner repository (dataset PXD003517).

Present address: Department of Physiology and Biophysics, University of Illinois at Chicago, Chicago, IL 60612.

*Present to whom correspondence may be addressed. Email: anaba@uiuc.edu or tjacks@mit.edu.

Present address: Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707054114/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1707054114
correlate with higher tumor recurrence rates and more advanced disease (7), yet their contributions to disease progression or utility as biomarkers remain unknown.

Because the complexity of the tumor microenvironment cannot be faithfully recapitulated in cell culture, genetically engineered mouse models of cancer are well suited to answer questions about the role of ECM components in tumor development and progression. In this study, we investigated the role of the ECM in the underlying biology of lung adenocarcinoma in the context of an autochthonous mouse model that recapitulates the in vivo complexity of cancer initiation and progression. This model is based on infection of a subset of adult lung epithelial cells with viral vectors expressing Cre recombinase in mice harboring a LoxP-Stop-LoxP Kras
t

mice and microdissected primary lung tumors (hereinafter referred to as primary tumors) and lymph node metastases were characterized by desmosomes staining for the 113 ECM proteins across all three biological replicates (Fig. 2 and Dataset S2C). Pairwise comparisons of the level of detection of the 113 ECM proteins in the different biological replicates allowed us to quantify the level of detection of 113 ECM and ECM-associated proteins in all 10 samples (Figs. 1D and 2 and Dataset S2). Annotation of these results using existing in silico murine matrisome data (9, 15) revealed the presence of 79 core matrisome proteins (43 ECM glycoproteins, 30 collagens, and 6 proteoglycans) and 34 ECM-associated proteins (14 ECM-affiliated proteins, 14 ECM regulators, and 6 secreted factors) (Fig. 2 and Dataset S2C). Pairwise comparisons of the level of detection of the 113 ECM proteins across all three biological replicates within each tissue type revealed significant intersample reproducibility (Fig. S2/).

Independent Component Analysis Identifies ECM Signatures of Fibrotic Lung, Primary Lung Tumor, and Metastasis. Unsupervised independent component analysis (ICA) identified three statistically significant ECM signatures within the dataset, which differentiated between the three disease states: fibrosis, primary tumor, and metastasis (Fig. 3A). This approach allowed us to identify the ECM proteins that were specific to each state (Fig. 3B). We observed increases in fibronectin (Fn) and the fibrinogen γ and β chains, but a decrease in osteopontin, in fibrotic tumors compared with lung tumors. Similar changes were consistent among independent samples from different animals, suggesting that these changes are consistent alterations of likely functional relevance. Primary lung tumors and lymph node metastases were characterized by decreased levels of nephroblastin and fibronectin, but only primary tumors demonstrated enrichment of previously unidentified peptides, we constructed unique isobaric tandem mass tag (TMT) (Dataset S2A) before MS analysis, which allowed us to obtain precise quantitative information from the samples. Because only 10 TMT tags exist at present, and because the MS analysis revealed similar ECM changes in primary lung tumors and lymph nodes (Dataset S1), we subsequently pooled these control samples and used this pool as a reference for quantitation in later analyses.

Because the mass spectrometer was operated in a data-independent manner, we performed two technical replicates (replicates 1 and 2); the overlap between the two replicates was 63%, which is satisfactory for this type of analysis (Figs. S1A and B). In both replicates, >40% of the ∼35,000 identified spectra belonged to ECM proteins, and resulted in the detection of 25 collagens and 75 additional ECM and ECM-associated proteins (Fig. 1C). To decrease the frequency of selecting collagen peptides and other abundant peptides for fragmentation, and thus increase the likelihood of selecting previously unidentified peptides, we constructed an exclusion list and performed two additional technical replicates (replicates 3 and 4; Fig. S1A and B). This analysis revealed an additional 590 unique peptides (Fig. S1A), 124 of which corresponded to known ECM proteins (Fig. S1B). Although we observed a 22% decrease in the number of spectra that were on the exclusion list, it is worth noting that the number of additional unique peptides identified was similar to that identified in a technical replicate performed without using the exclusion list. The union of all four replicates allowed us to quantify the level of detection of 113 ECM and ECM-associated proteins in all 10 samples (Figs. 1D and 2 and Dataset S2). Annotation of these results using existing in silico murine matrisome data (9, 15) revealed the presence of 79 core matrisome proteins (43 ECM glycoproteins, 30 collagens, and 6 proteoglycans) and 34 ECM-associated proteins (14 ECM-affiliated proteins, 14 ECM regulators, and 6 secreted factors) (Fig. 2 and Dataset S2C).
lung tumors contained increased laminin and elastin. Finally, the metastatic samples showed significant and specific elevations in the expression of S100A6, S100A10, and S100A11. Some tumors were uniformly positive for the S100 factors, whereas others exhibited patchy expression (Fig. S3). 

We used volcano plots to highlight ECM proteins that were detected in significantly altered abundance in pairwise comparisons of diseased and normal lungs (Fig. 3C and Dataset S3). Tnc was present in very low amounts in normal lungs, but was markedly increased levels in fibrotic and tumor samples (Figs. 2 and 3C). Fibronectin 1 (Fn), a binding partner of Tnc, was also significantly more abundant in the disease samples. In addition, three members of the S100 family of proteins were detected in greater abundance in advanced KP tumors compared with normal lung. Analysis of S100 protein expression showed increased levels of collagen I and Fn (ECM markers), actin (cytoskeletal marker), GAPDH (cytosolic marker), and histones (nuclear marker). The insoluble fraction remaining after serial extraction (highlighted in blue) was enriched for ECM proteins and largely depleted for intracellular components. 

Validation of MS Data by Immunohistochemistry. To confirm the MS data, we performed immunohistochemistry (IHC) for a subset of the identified ECM proteins to determine whether their expression was altered in fibrotic lungs, tumors, and metastases compared with normal lungs. Analysis of S100 protein expression patterns in normal lungs showed that these proteins were largely absent from the extracellular space. In the fibrotic lung samples, the overall expression for the three S100 proteins was relatively low, which is consistent with the MS analysis (Fig. 3C). In contrast, primary tumors had a marked increase of all three proteins in the tumor cells and in the extracellular spaces (Fig. 4 and Fig. S3). Some tumors were uniformly positive for the S100 factors, whereas others exhibited patchy expression (Fig. S3), but higher-grade tumors and aggressive areas were overall positive. Similarly, S100A6, S100A10, and S100A11 were also more abundant in KP lung metastases to the mediastinal lymph node (Fig. 4, Right), whereas metastases and fibrotic lungs (Fig. S2C). Moreover, analysis of the overlap between the ECM and ECM-associated proteins identified in significantly altered abundance in all three conditions revealed three proteins: Fn, Tnc, and S100A11 (Fig. 2D). There were also 13 proteins with similar changes between KP tumors and metastases, and 8 between fibrosis and tumors (Fig. 2D and Dataset S3F). In summary, analysis of the MS data revealed common and unique changes in specific ECM components that characterize the three disease states.
transcript levels were markedly up-regulated and Met cells, but not in during Lung Cancer Progression. and Met cell lines were harvested expression in cell lines isolated from lower- locus at four distinct regions expression. Although lines (Fig. 5 0.0001) (19). To confirm this Gocheva et al. -seq) data (20) that identified www.pnas.org/cgi/doi/10.1073/pnas.1707054114 List of 113 quantified ECM proteins in normal was significantly up-regulated in T tumors and their associated metastases (Fig. 4). In the normal lung, Fn staining was less pronounced, with some expression found tumors of diverse origins (17). In agreement with those reports and markers of disease development.

We next examined the expression of Fn, an ECM glycoprotein commonly bound to integrins, which are known proproliferation factors (16). IHC analysis revealed that compared with normal lungs, Fn levels were increased in fibrillar lungs as well as in lung tumors and their associated metastases (Fig. 4). In the normal lung, Fn staining was less pronounced, with some expression found closely associated with blood vessels. In the fibrillar samples, Fn expression was significantly increased and in a stromal pattern. The tumor and metastatic samples also showed increased staining that was associated primarily with the stroma and showed distinct fibrillar networks (Fig. 4 and Fig. S3).

Finally, we examined the expression of Tnc, an ECM protein that is highly expressed in the developing embryo but absent in most adult tissues. Numerous studies have shown reexpression of Tnc at sites of wound healing or inflammation, as well as in malignant tumors of diverse origins (17). In agreement with those reports and with our MS data, there was minimal Tnc staining in the normal lung, but Tnc expression was dramatically increased at sites of fibrosis as well as in tumors and metastases (Fig. 4 and Fig. S3).

In summary, the unique changes identified in each condition suggest that these ECM proteins might have specific roles in mediating the phenotypes associated with the disease state with which they are associated. Furthermore, the differential staining patterns raise the possibility that ECM proteins can serve as faithful biomarkers for each of these disease states.

**Regulation of Tnc During Lung Cancer Progression.** We focused our efforts on investigating the ECM protein Tnc, because its IHC pattern suggested that it may have a role in driving the pathology of lung fibrosis, as well as during tumor progression and the development of metastasis. A functional role for Tnc in mice with bleomycin-induced fibrosis has already been demonstrated; Carey et al. (18) showed that Tnc-knockout animals are protected against fibrosis and exhibit significantly lower accumulation of collagen in the lungs. Thus, we chose to focus on the role of Tnc in lung adenocarcinomas using the KP model.

To dissect some of the molecular details of the regulatory networks that are affected during cancer progression and lead to up-regulation of expression of Tnc, we used a collection of cell lines isolated from KP mice that are representative of different stages of tumor progression. T_nonmet cells were isolated from nonmetastatic primary KP tumors, whereas T_met and Met cell lines were harvested from metastatic primary tumors and their metastases to the lymph nodes or liver, respectively (19). Analysis of existing gene expression array data from these cell lines identified that, among other changes, Tnc was significantly up-regulated in T_met and Met cells, but not in the nonmetastatic T_nonmet lines (P = 0.0001) (19). To confirm this observation, we isolated mRNA from these cell lines and performed quantitative RT-PCR (qPCR) analysis for Tnc expression. Although there was almost no Tnc expression in cell lines isolated from lower-grade T_nonmet cells, Tnc transcript levels were markedly up-regulated in both the T_met and Met lines (Fig. 5). These observations suggest a possible role for Tnc in tumor progression to metastasis.

The transcription factor Nkx2-1 has been shown to inhibit tumor progression by repressing genes involved in metastasis (19). Given our observed anticorrelation between the patterns of expression of Nkx2-1 and Tnc in the cell lines (Fig. 5), we hypothesized that Tnc also might be subject to Nkx2-1 repression. To examine this possibility, we analyzed Nkx2-1 chromatin immunoprecipitation sequencing (ChIP-seq) data (20) that identified Nkx2-1 binding sites in Kras-driven lung tumors, and discovered that Nkx2-1 binds the murine Tnc locus at four distinct regions.
near the transcription start site (Fig. 5B). We confirmed these data by ChIP and quantitative PCR (ChIP-qPCR) with primers specific to the four putative binding regions in the Tnc locus, and observed that Nkx2-1 bound to these regions with a strength comparable to that of a canonical Nkx2-1 target gene, Sfpn1 (Fig. 5C). These results suggest that Nkx2-1 represses Tnc in KP tumor cells.

To determine whether Nkx2-1 is necessary and sufficient to control Tnc expression, we performed loss-of-function and complementary gain-of-function experiments. We observed that shRNA-mediated silencing of Nkx2-1 in two different Tnonmet cell lines indeed led to the derepression of Tnc (Fig. 5D). Analysis of previously published gene expression data on Tnonmet cell lines expressing shRNA for Nkx2-1 showed that Tnc levels increase following Nkx2-1 knockdown in three independent Tnonmet cell lines (19). Conversely, Tnc levels were strongly down-regulated in a T

sections of high-grade KP tumors for Tnc and smooth muscle actin revealed that despite the widespread presence of fibroblasts in these high-grade tumors, the presence of Tnc was restricted to the KP tumors and metastatic states. (A) Analysis of proteomic data reveals three distinct statistically significant signatures (P < 0.01) characterizing fibrosis, primary lung tumor, and metastatic samples. Although each signature in the row-normalized heatmap is characterized by low protein levels (blue), each signature is two-sided, allowing for identification of proteins with high levels that characterize each of the states. Blue indicates lower protein levels compared with yellow (higher levels). (B) Heat maps for each of the three signatures show representation of enriched or depleted proteins (|z| > 1.75). Rows represent standard-
protein induction by immunofluorescence (Fig. 6B) and Western blot analysis (Fig. S6A). Although Tnc expression was mostly intracellular shortly after cell seeding (Fig. 6B), over time the majority became secreted (Fig. S6A). Overexpression of Tnc did not affect the proliferation rate of cells in culture (Fig. S6B).

We next used the Tnc-overexpressing cell lines to test whether increased Tnc levels could promote tumor progression in a mouse model of metastasis that assays for the ability of cells to disseminate to distant organs (Fig. 6C). Whereas Tnc overexpression had no effect on primary tumor growth after subcutaneous transplantation (Fig. 6D), it had a significant effect on promoting the metastatic colonization of the lungs (Fig. 6E). Increased tumor burden in the lungs was observed with both Tnc-specific gRNAs (5- to 35-fold increase), and all inoculated animals developed metastases. TNC staining in the primary tumors and lung metastases was fibrillar (Fig. S6 C and D).

To further investigate the role of Tnc during the metastatic process, we performed tail vein metastasis assays, which test for the ability of tumor cells to extravasate, colonize, and grow in secondary sites (Fig. 6F). Tnc overexpression led to high levels of the protein as shown by IHC (Fig. 6G and Fig. S7), along with a significant increase in metastasis to the lungs (Fig. 6F). This set of experiments establish a direct role for Tnc in promoting metastasis in lung adenocarcinoma and provide further evidence that the overexpression of Tnc observed in advanced and metastatic tumor cells play an important role in the aggressive phenotype of these tumors.

**Gene Expression of Specific Matrisome Factors Is Associated with Poor Prognosis for Lung Adenocarcinoma Patients.** We next addressed how the findings from the mouse models relate to human lung cancer. We first investigated whether similar changes in the expression levels of TNC, as well as the other four factors identified and validated through our analysis, were found in patients with lung adenocarcinoma. To this end, we analyzed RNA-seq gene expression profiles of primary tumors, matched normal samples, and relevant clinical data obtained from The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD) cohort. TNC levels were significantly higher in tumors compared with normal lungs across the entire dataset as well as in matched tumors compared with normal tissue from the same patient (Fig. 7A and Fig. S8A). Similarly, expression of S100A6 and S100A11, two other factors found to be increased in the KP mouse model, was also significantly higher (Fig. 7A and Fig. S8A). In contrast, expression of S100A10 was significantly lower in matched tumor samples compared with normal lungs. FN1 expression was lower across the entire dataset, although matched tissues from the same patient exhibited a significant up-regulation (Fig. S8A).

We next explored whether the increased levels of any of these factors carries prognostic information. Kaplan–Meier 5-y survival analyses revealed a significantly poorer prognosis in patients with high TNC expression (top 25%) compared with those with low TNC expression (Fig. 7B). Similarly, high expression of S100A10 and S100A11 were correlated with significantly poorer patient prognosis, whereas high expression of S100A6 or FN1 did not (Fig. 7B and Fig. S8C). We next investigated whether a combined metric based on the expression values of the three genes, for which high expression is correlated with poorer prognosis, would have prognostic value. We used the geometric mean of the expression values of TNC, S100A10, and S100A11 (three-gene signature) to score and rank patients. Higher signature scores were indeed significantly associated

**Fig. 4.** Validation of significantly up-regulated ECM proteins by IHC. Representative images of IHC for the indicated proteins in normal lung, primary lung tumor, and lung metastases to the lymph node, stained under identical conditions. Positive signals are shown in brown; hematoxylin (blue) was used as a counterstain. LN indicates the normal lymph node region, and Met is the area occupied by lung metastasis. All pictures were taken under the same magnification. (Scale bar: 50 μm.)

**Fig. 5.** Nkx2-1 represses Tnc expression. (A) qRT-PCR analysis of Tnc expression in Tnc knockdown in two different cell lines for Tnc (Left) and Nkx2-1 (Right) expression relative to GAPDH used as control. *P < 0.05, **P < 0.01, unpaired t test. (B) Analysis of ChIP-Seq data (20) reveals binding of Nkx2-1 in the Tnc locus (Fig. 4C) with significant enrichment of Nkx2-1 binding at the Tnc genomic locus. Data represent mean ± SEM of three independent experiments. SftpA serves as a positive control. Negative control mapping to a gene desert region on murine chromosome 8 (GD8). The Tnc peak numbers correspond to those in Fig. 4C. (C) Western blots showing that Nkx2-1 knockdown significantly decreased Tnc protein levels in Tnc-overexpressing cell lines. (D) Quantification of Nkx2-1 and Tnc expression in early-stage (4–6 wk after initiation) and late-stage KP tumors (>12 wk after initiation).
with worse patient outcome (Fig. 7C and Fig. SSD) and exhibited stronger statistical significance ($P = 0.00009$) than either of the Kaplan-Meier analyses based on expression of individual genes (Fig. 7B and Fig. S8E). Interestingly, there was no association between the three-gene signature and survival in patients with colorectal cancer, suggesting potential organ and tumor-specific differences in ECM composition and function (Fig. S8F).

To further evaluate the prognostic value of the three-gene signature in predicting patient survival, in the context of other clinical covariates, we used the Cox proportional hazards model to perform univariate and multivariable survival analyses on the TCGA lung adenocarcinoma patient cohort (Fig. 7D). Univariate analysis indicated that an increasing three-gene signature score was significantly associated with poor patient survival ($P = 0.0209$). Multivariable analysis, controlling for other covariates (age, sex, smoking history, and mutational load), also showed a significant correlation between the three-gene signature and worse survival (hazard ratio, 1.30; $P = 0.00624$). Taken together, our findings provide important information for the prognosis of patients with lung adenocarcinoma, and suggest that this TNC expression-based gene signature could serve as a useful biomarker.

**Discussion**

Although previous studies have examined the changes in individual components of the ECM in lung cancer and fibrosis, a comprehensive approach to characterizing the composition of the ECM in these disease states has been lacking. In the present study, we used quantitative proteomics to characterize the global changes in ECM protein abundance that occur in fibrosis and during lung cancer development. This analysis was performed in the context of well-established mouse models that recapitulate the complexity of the in situ changes that accompany fibrosis and cancer progression. We compared the ECM changes that occur during tumorigenesis with those that occur in a mouse model of pulmonary fibrosis to delineate the similarities and differences between these conditions.

**Changes in ECM in Lung Fibrosis.** Two recent studies have used proteomics to report changes occurring in bleomycin-induced fibrosis (22, 23). Along with some overlapping findings, our study of the insoluble ECM has uncovered and quantified additional proteins. This can be attributed in part to our use of different proteomics technologies. Collectively, these studies, together with our study, provide a more complete understanding of the ECM changes that occur in this model, and the findings possibly could lead to the identification of novel antifibrotic agents and strategies.

We identified and validated two ECM proteins, Fn and Tnc, as highly abundant in fibrosis. These findings are consistent with previous reports in which both proteins were found to be up-regulated in mouse and rat models of bleomycin-induced pulmonary fibrosis (22–24). Schiller et al. (23) showed that Tnc is increased at both
protein and mRNA levels in the lungs of mice treated with bleomycin, and demonstrated that higher Tnc expression is associated with stiffer lung tissue, a feature of fibrosis. Furthermore, Tnc not only is a marker of pulmonary fibrosis, but also has been implicated in the pathogenesis of the disease. Tnc-knockout animals are protected against bleomycin-induced fibrosis and have lower accumulation of collagen, reduced fibroblast infiltration, and reduced activity of TGF-β in the lungs (18). Importantly, patients with pulmonary fibrosis show increased levels of Tnc, suggesting that the role of Tnc in pulmonary fibrosis is not limited to animal models of the disease (25, 26). Whether Tnc contributes to the development and severity of the disease in humans, and what factors cause its up-regulation in that setting, remain open questions.

**ECM Changes in Cancer.** Through delineating the changes that occur during pulmonary fibrosis and cancer progression, we have identified unique protein signatures that independently set apart each condition. Because changes in the composition of the ECM can have a regulatory role during cancer development and metastasis formation, we compared primary tumors and metastases both to normal lung and to each other. Although many factors were significantly altered in the tumorigenic state compared with healthy tissue, there were few statistically significant differences in ECM between the primary lesions and their associated lymph node metastases. Thus, the ECM at these local metastases closely resembled that of the primary tumors.

Interestingly of the 20 proteins that we found in higher abundance in the ECM signature of KP lung tumors (compared with normal lung), nine (Col12a1, Col14a1, Col18a1, Col8a1, Fbln2, Fn1, Ltbp2, Nid1, and Tnc) were encoded by genes that are part of the AngioMatrix signature, an ensemble of matrisome genes whose up-regulation at the mRNA has been associated with the induction of angiogenesis in the RIP1-Tag2 mouse model of pancreatic neuroendocrine cancer (27). Tnc was identified as an important component of the AngioMatrix in driving the angiogenesis in the neuroendocrine tumor model; the AngioMatrix also has negative predictive value in patients with glioma and colorectal carcinoma (27). The partial overlap between the lung tumor matrisome factors we have identified and the AngioMatrix raises the possibility of

---

**Fig. 7.** Prognostic value of matrisome factors within the LUAD patient cohort. (A) Gene expression values (RNA-seq normalized counts standardized for mean = 0, SD = 1) for a subset of the validated matrisome factors in matched normal lung tissue and primary lung tumors of patients with lung adenocarcinoma (n = 57). Two-sided P values (Kolmogorov–Smirnov test) are shown. (B) Kaplan-Meier 5-y survival analysis comparing patients in the top 25th percentile of expression for each gene (n = 114; red) and those in the bottom 75th percentile (n = 444; blue). Log-rank test P values are shown. (C) Kaplan-Meier 5-y survival analysis in TCGA LUAD using an expression metric to quantify the combined expression levels of S100A10, S100A11, and TNC (three-gene signature). Specifically, the geometric mean of the expression levels was used to score and rank patients. Shown are the top 45% scoring patients (n = 206) vs. the rest (n = 252). Log-rank test P value is shown (median survival, 1,043 d for the high-scoring patient subpopulation and 1,725 d for the remainder of the cohort). (D) Results of univariate and multivariable Cox proportional hazards model on overall survival in the LUAD cohort (all patients). Increasing three-gene signature score shows a significant association with poorer survival after controlling for other characteristics.

---

**Table:**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Univariate</th>
<th></th>
<th>Multivariable</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>p</td>
<td>HR (95% CI)</td>
<td>p</td>
</tr>
<tr>
<td>TNC, S100A10, S100A11 expression metric (high vs low)</td>
<td>1.20(1.04-1.40)</td>
<td>0.0209</td>
<td>1.30(1.08-1.57)</td>
<td>0.00624</td>
</tr>
<tr>
<td>Gender (male vs female)</td>
<td>0.96(0.66-1.39)</td>
<td>0.83088</td>
<td>0.96(0.63-1.47)</td>
<td>0.86102</td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.01(0.99-1.03)</td>
<td>0.05811</td>
<td>1.01(0.99-1.04)</td>
<td>0.31931</td>
</tr>
<tr>
<td>Smoking History (reformed &gt; 15yrs vs non-smoker)</td>
<td>0.97(0.51-1.84)</td>
<td>0.91765</td>
<td>1.15(0.54-2.45)</td>
<td>0.71345</td>
</tr>
<tr>
<td>Smoking History (reformed &lt; 15yrs vs non-smoker)</td>
<td>1.01(0.56-1.85)</td>
<td>0.965</td>
<td>1.56(0.75-3.26)</td>
<td>0.23273</td>
</tr>
<tr>
<td>Smoking History (current smoker vs non-smoker)</td>
<td>0.72(0.37-1.42)</td>
<td>0.342</td>
<td>1.12(0.50-2.53)</td>
<td>0.77761</td>
</tr>
<tr>
<td>Mutational Load (mutations per coding Mb)</td>
<td>0.98(0.96-1.00)</td>
<td>0.02314</td>
<td>0.97(0.95-0.99)</td>
<td>0.01091</td>
</tr>
</tbody>
</table>

HR = Hazard ratio; CI = Confidence interval
S100 proteins are regulated by binding of calcium ions, which allows them to act as calcium sensors and thus translate fluctuations in intracellular calcium levels into the appropriate cellular responses (28). The S100 proteins can act as extracellular factors as well. This family of proteins comprises 21 members, each with distinct functional properties. Although their precise roles are not well understood, multiple S100 family members exhibit deregulated expression in various cancer types.

Our analysis of patient data revealed that some of the S100 factors that we have identified in this study are highly expressed in lung adenocarcinoma tissue, and that this higher expression is correlated with poor patient prognosis. Several other studies have reported similar observations, supporting the idea that our findings may be clinically relevant (29-31). S100A10 expression has been significantly correlated with higher TNM stage, more frequent vascular invasion, and a poorer overall prognosis (30). Similarly, in a study of S100A11 expression in 179 tumor samples from patients with lung adenocarcinoma, Woo et al. (31) found significantly higher S100A11 levels in adenocarcinomas with KRAS mutations, as well as in poorly differentiated tumors. Moreover, strong S100A11 expression was correlated with shorter disease-free survival. Those results suggest that this protein might be involved in the tumorigenic process specifically in KRAS-mutant lung adenocarcinomas, the subset of human tumors that the KP model is designed to represent. Previous work using RNAi to knock down S100A11 has established that this factor promotes proliferation of human lung adenocarcinoma cell lines in vitro and s.c. growth in vivo using xenografts (32). Moreover, several studies also have suggested that this protein contributes to metastasis and invasion, and in fact up-regulation of S100A11 in patients with non-small-cell lung cancer is significantly associated with the presence of lymph node metastasis (33). Although the precise biological role of S100A11 in cancer remains unclear, the autochthonous KP model of lung adenocarcinoma represents an ideal setting for studying whether S100A11 has a functional role in the pathogenesis of lung cancer.

Role of Tnc in Tumorigenesis. Along with the S100 proteins, we also identified up-regulation of Tnc in high-grade KP tumors and metastases. In the developing embryo, Tnc expression is restricted to areas of active migration and epithelial-to-mesenchymal transition, whereas after birth its levels are down-regulated in all tissues (34). However, in adults, Tnc expression is increased at sites of inflammation and wound healing, as well as in cancers. Numerous clinical reports have shown up-regulation of Tnc in patients with diverse cancer types. Consistent with our results, high Tnc expression also has been observed in patients with lung cancer (35).

Whereas the roles of Tnc in tumor cell lines have been studied extensively, the results have led to sometimes conflicting findings, and relatively few studies have used mouse cancer models. To better understand the contribution of Tnc to breast cancer, Talts et al. (36) crossed Tnc-knockout mice to the MMTV-PyMT model, and reported that a lack of Tnc had no effect on primary tumor growth or metastatic dissemination to the lungs. Similarly, knockdown of Tnc in a human breast cancer cell line implanted into the mammary fat pad of immunodeficient mice had no effect on primary tumor growth, but did reduce metastasis to the lungs (37). In this context, tumor formation in Tnc-deficient mice has been found to promote the outgrowth of pulmonary lesions by enhancing Wnt and Notch signaling, thereby promoting the viability of cancer cells. Similar results were reported in a mouse model of pancreatic neuroendocrine cancer, in which Tnc was found to promote tumorigenesis and lung metastasis (38). These differing effects of Tnc on cancer may reflect the tumor type-specific or oncogene-specific roles of Tnc in cancer progression.

Here, using the CRISPR/Cas9 SAM system to overexpress Tnc from its endogenous promoter in tumor cells, we have shown that Tnc can promote the spread of lung adenocarcinoma cells without affecting primary tumor growth. Previous studies have identified tumor-associated fibroblasts (TAFs) as a major source of Tnc in tumors (39), but whether stroma-derived Tnc has a role in tumorigenesis in this model is unclear. Although the precise mechanism of action remains to be elucidated, we have shown that Tnc is a target of the transcription factor Nkx2-1, which has been shown to suppress lung cancer progression and metastasis through the suppression of embryonic genes (19). Thus, Nkx2-1 appears to exert its effects through various factors, including the ECM protein Tnc. Whereas the effect of Nkx2-1 on Tnc is likely direct, this possibility requires further investigation using promoter luciferase assays. Whether the same relationship is true in humans remains to be proven, although it is likely, given that the human and mouse Nkx2-1 proteins share 98% identity, and that two of the Nkx2-1 binding sites that we identified in the murine Tnc promoter are conserved in the human Tnc promoter (peak 2, 75% identity; peak 3, 80% identity). Moreover, Tnc might not be the sole ECM factor regulated by Nkx2-1; analysis of existing Nkx2-1 ChIP-seq data (20) identified potential Nkx2-1 binding sites within 4 kb of the transcriptional start sites of 14 of the 36 matrisomal proteins detected in differential abundance in KP lung tumors (Agrn, Col12a1, Col11a1, Ctds, Fbln5, Hspg2, Lgals3, Lbp2, Nid1, Nid2, Npnt, S100a11, Sftp1, Sftpd, Tnc). Further work is needed to confirm whether these genes are also transcriptionally regulated by Nkx2-1, and whether they affect metastasis.

Because Tnc expression is absent in normal adult tissues but highly up-regulated in many solid cancers, it represents an ideal diagnostic marker and therapeutic target in cancers (40). In several compounds currently in clinical trials, antibodies against Tnc have been conjugated to either radioactive compounds that can inhibit tumor growth or IL-2, aimed at improving the efficacy of chemotherapy. Another area showing promise is the development of therapeutic or preventive cancer vaccines in which Tnc could be targeted alone or in combination with other factors (41).
Mice were killed by CO₂ asphyxiation, and lungs were inflated with 10% formic acid (Polysciences), fixed overnight in zinc/formalin at room temperature, and then transferred to 70% ethanol and embedded in paraffin. Masić’s trichrome staining was performed following standard procedures. IHC was performed using the ABC Vectorstain Kit (Vector Laboratories) with antibodies to Tnc (AB19011), Nkx2-1 (1:400; Abcam, ab181975), S100A10 (1:500, Abcam, ab181975), and the slides were counterstained with hematoxylin. Table S1 shows the number of samples (each sample is an individual mouse) that were stained.

CRISPR Activation. Nondonal 1233 KP cells (43) stably expressing dCas9-P64A-Blast (Addgene; 61425) and MS2-P65-HSF1-Hygro (Addgene; 61426) were generated via sequential lentiviral transduction and selection with blasticidin and hygromycin, respectively. To overexpress Tnc, we designed and cloned five independent gRNA sequences targeting the Tnc promoter into a lentiviral vector (Lenti-sgRNA-M522-Zeocin; Addgene; 61427) and subsequently transduced and zeocin-selected the aforementioned cell lines to generate KP cell lines stably expressing all three components. The target gRNA sequences were designed using the SAM algo-

### ACKNOWLEDGMENTS

We thank Nadya Dimitrova for a critical review of the manuscript; the Koch Institute Swanson Biotechnology Center for technical support, specifically Kathleen Cormier at the Hope Babette Tag (2013) for Multiplexed Quantification and Roderick Bronson for pathological analysis. This work was supported by the National Cancer Institute’s Tumor Microenvironment Network (Grants U54 CA126515 and U54 CA163109), by the Howard Hughes Medical Institute (at which R.O.H. and T.J. are investigators), the Ludwig Center for Molecular Oncology at MIT, and the National Cancer Institute (Koch Institute Cancer Center Support Grant P30-CA14051). V.G. was supported by a Jane Coffin Childs Memorial Fund Postdoctoral Fellowship. N.J. is supported by a Mazumdar-Shaw International Oncology Fellowship. R.O.H. is a Daniel K. Ludwig Professor for Cancer Research. G.T. is a David H. Koch Professor of Biology and a Daniel K. Ludwig Scholar at MIT.

### Clinical Data Analyses

RNA-seq gene expression profiles of primary tumors, matched normal samples, and relevant clinical data of 488 patients with lung adenocarcinoma with primary tumor samples were obtained from TCGA (https://cancergenome.nih.gov). Details of the bioinformatics analyses are provided in SI Materials and Methods.