

Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer

Roya Navab^{a,1}, Dan Strumpf^{a,1}, Bizhan Bandarchi^{a,1}, Chang-Qi Zhu^{a,1}, Melania Pintilie^a, Varune Rohan Ramnarine^a, Emin Ibrahimov^a, Nikolina Radulovich^a, Lisa Leung^a, Malgorzata Barczyk^{a,b}, Devang Panchal^a, Christine To^a, James J. Yun^a, Sandy Der^a, Frances A. Shepherd^{a,c}, Igor Jurisica^{a,d,e}, and Ming-Sound Tsao^{a,e,f,2}

^aThe Campbell Family Institute for Cancer Research, Ontario Cancer Institute at Princess Margaret Hospital, University Health Network, Toronto, ON, Canada M5G 2M9; ^bDepartment of Biomedicine, Jonas Lies vei, N-5009 Bergen, Norway; Departments of ^cMedicine, ^dComputer Science, and ^eMedical Biophysics, and ^fLaboratory of Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada M5A 2N4

Edited* by Tak Wah Mak, The Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute at Princess Margaret Hospital, University Health Network, Toronto, ON, Canada, and approved March 10, 2011 (received for review September 28, 2010)

The tumor microenvironment strongly influences cancer development, progression, and metastasis. The role of carcinoma-associated fibroblasts (CAFs) in these processes and their clinical impact has not been studied systematically in non-small cell lung carcinoma (NSCLC). We established primary cultures of CAFs and matched normal fibroblasts (NFs) from 15 resected NSCLC. We demonstrate that CAFs have greater ability than NFs to enhance the tumorigenicity of lung cancer cell lines. Microarray gene-expression analysis of the 15 matched CAF and NF cell lines identified 46 differentially expressed genes, encoding for proteins that are significantly enriched for extracellular proteins regulated by the TGF- β signaling pathway. We have identified a subset of 11 genes (13 probe sets) that formed a prognostic gene-expression signature, which was validated in multiple independent NSCLC microarray datasets. Functional annotation using protein-protein interaction analyses of these and published cancer stroma-associated gene-expression changes revealed prominent involvement of the focal adhesion and MAPK signaling pathways. Fourteen (30%) of the 46 genes also were differentially expressed in laser-capture-microdissected corresponding primary tumor stroma compared with the matched normal lung. Six of these 14 genes could be induced by TGF- β 1 in NF. The results establish the prognostic impact of CAF-associated gene-expression changes in NSCLC patients.

integrin α 11 | NAViGaTOR | adenocarcinoma | extracellular matrix

Lung cancer is the leading cause of cancer deaths worldwide (1). Non-small cell lung carcinoma (NSCLC) accounts for 85% of all lung cancers. One of the most consistent histological features of cancer cell invasion is the appearance of desmoplasia: stromal changes characterized by the activation of stromal fibroblasts into carcinoma-associated fibroblasts (CAFs), increased matrix protein disposition, new blood vessel formation, and immune cell infiltration (2, 3). These changes have been reported to promote tumor cell growth, invasion, metastases, and resistance to treatment as well as to mediate immune reaction against tumor cells (2, 4, 5). Published prognostic gene signatures for NSCLC have included genes encoding for ECM proteins such as collagens 1A1, 1A2, and 9; glypican 3; intercellular adhesion molecule 1 (ICAM-1); laminin B1; selectin L; selectin P; and secreted protein acidic and rich in cysteine (SPARC) (3, 5). Furthermore, stromal elements, including PDGF-C, VEGF-A, IL-8, endothelin-1 (EDN1), osteopontin (SPP1), and the chemokine CXCL1, recently have been suggested as potential therapeutic targets in skin and breast cancers (4, 6).

To gain greater insight into the gene-expression characteristics in CAFs and tumor stroma of NSCLC and more direct evidence for their clinical role, we conducted microarray analyses on paired CAF and normal fibroblasts (NFs) cultured from 15 resected NSCLC specimens and their corresponding laser-capture-microdissected (LCM) tumor stroma and histologically normal lung parenchyma. We identified a NSCLC stromal prognostic signature that could be validated in multiple independent published expression datasets of primary NSCLCs. The results establish the clinical relevance of stromal gene-expression changes in lung cancer.

Results

Cultured CAFs Display Features of Myofibroblasts. By using a study protocol approved by the Institutional Research Ethics Board, CAFs and NFs were cultured from 15 surgically resected primary NSCLCs, and the histologically confirmed normal lung tissue was obtained from the same lobe (Table S1A). Both the primary cultured CAFs and tumor stromal fibroblasts expressed α -smooth muscle actin (α -SMA) (Fig. 1A), a marker of myofibroblasts (7). These primary CAFs and NFs could be maintained in culture for up to 20 population doublings. Primary fibroblasts preserved their ability to induce collagen gel contraction (8), which was greater with CAFs than NFs (Fig. 1B). These results showed that CAFs can maintain the phenotypic properties of myofibroblasts even in the absence of continuing interaction with carcinoma cells.

To facilitate further the *in vitro* and *in vivo* studies using CAFs and NFs, one pair of primary cultured CAFs and NFs was immortalized with lentivirus expressing human telomerase (hTERT). The resulting immortalized CAF 094^{YFP^{hTERT}} and NF 094^{YFP^{hTERT}} cell lines showed loss of senescence (Fig. 1C).

CAFs Enhance Invasion and Tumorigenicity of NSCLC Cells. By using the coculture Matrigel invasion assay, CAFs increased the invasiveness of both NCI-H460 and A549 NSCLC cells compared with NFs (Fig. 1D and Fig. S1A). This effect was not attributable to an effect on proliferation of the tumor cells (Fig. S1B). We also observed that the invading tumor cells appeared elongated and fibroblast-like only when they were cocultured with CAFs but not with NFs (Fig. S1C). A similar alteration of tumor cell appearance was observed when they were exposed to conditioned media of CAFs but not NFs (Fig. S1D). By using gelatin zymography, an active lytic band (68 kDa) of matrix metalloproteinase 2 (MMP-2) was observed in CAFs but not in NFs cocultured with tumor cells (Fig. 1E), suggesting a role for MMP-2 activation in CAF-induced enhancement of tumor cell invasion.

Subcutaneous coimplantation of A549 lung adenocarcinoma cells with CAF 094 into severe combined immunodeficient (SCID) mice significantly enhanced tumor growth compared with controls (Fig. 1F). A similar effect was observed with the immortalized CAF 094^{YFP^{hTERT}} (Fig. 1F). These findings were validated further with the NCI-H460 cell line (Fig. 1F).

Author contributions: R.N., D.S., and M.-S.T. designed research; R.N., D.S., B.B., E.I., N.R., L.L., M.B., D.P., and J.J.Y. performed research; R.N., D.S., C.-Q.Z., M.P., V.R.R., C.T., I.J., and M.-S.T. analyzed data; and R.N., D.S., B.B., C.-Q.Z., M.P., N.R., S.D., F.A.S., I.J., and M.-S.T. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE22874).

¹R.N., D.S., B.B., and C.-Q.Z. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: ming.tsao@uhn.on.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1014506108/-DCSupplemental.

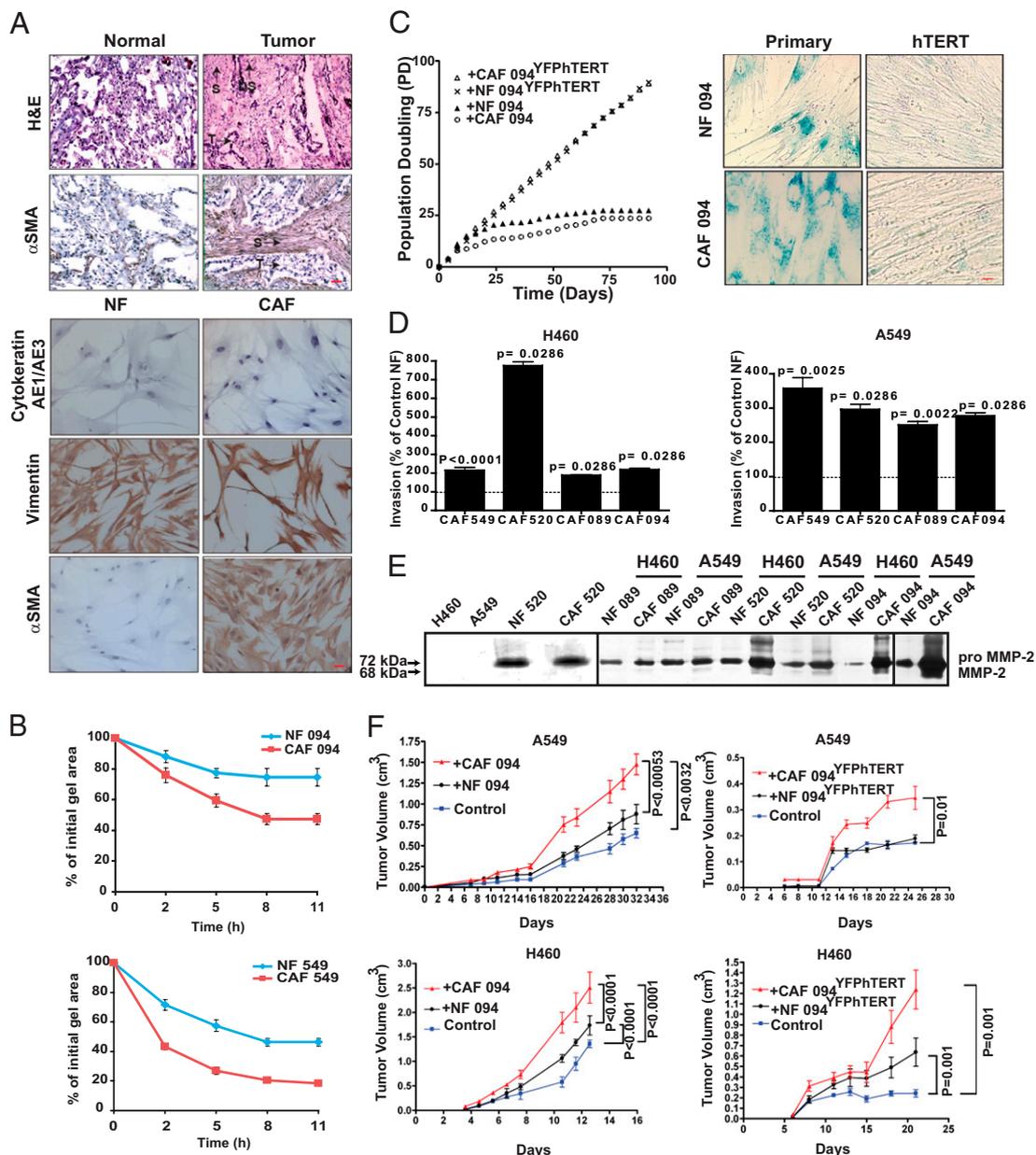


Fig. 1. Characterization of CAF and lung NF. (A) Representative H&E-stained sections of normal lung and a lung adenocarcinoma with prominent dysplasia (DS) and showing strong staining of the tumor stromal fibroblasts (S) but not tumor cells (T) for α -SMA. Both CAF and NF stain positive for vimentin but negative for cytokeratin (AE1–AE3 antibody). (Scale bar: 20 μ m.) (B) Time-dependent collagen gel contraction induced by NF and CAF. Each point represents means \pm SD of eight replicate samples. (C Left) In contrast to the parent cell lines, NF 094^{YFP^{hTERT}} and CAF 094^{YFP^{hTERT}} continue to proliferate beyond 10–20 population doublings. (Right) The senescence-associated acidic β -galactosidase enzyme activity (blue staining) was detected in primary fibroblasts that failed to continue doubling but not in hTERT-immortalized cells. (Scale bar: 40 μ m.) (D) Matrigel invasion ability of H460 and A549 cell lines was enhanced by coculture with four pairs of primary CAF. Significance was tested with the Mann–Whitney test. (E) Gelatin zymography shows activation of MMP-2 when tumor cells were cocultured with CAF compared with coculturing with NF. The bands show the lytic zones. (F) Both primary (CAF 094) and immortalized (CAF 094^{YFP^{hTERT}}) cells enhanced the in vivo tumorigenicity of both A549 and H460 cells in SCID mice (mean \pm SEs with eight mice in each group).

CAF-Specific Gene-Expression Profile in NSCLC. Using the Affymetrix Exon 1.0 ST oligonucleotide array and paired Significance Analysis in Microarray (SAM) (9), we identified 46 differentially expressed genes ($q < 10\%$; absolute fold change > 2) between CAFs and NFs of 15 patients (Table S1A, Dataset S1A and B, and Fig. 2). The expression levels and fold change for these genes vary across the 15 pairs of CAFs and NFs (Fig. 2, Dataset S1C, and Fig. S24). Using reverse-transcriptase/quantitative PCR (RT-qPCR), we demonstrated the stability of mRNA expression by high (≥ 0.8) intraclass correlation coefficients (ICCs) for each gene (Dataset S1D) in three pairs of NFs and CAFs

(549, 927, and 746), over three passages using triplicate samples. Gene Ontology (GO) analysis revealed that these 46 genes were enriched significantly for those encoding for extracellular proteins (extracellular region: 19/46 genes, $P = 8.7 \times 10^{-5}$, GO:0005576), and also a large proportion of membrane-bound proteins (22/46). These genes are involved in signal transduction (14/46, GO:0007165), response to stress (11/46, GO:0006950), cell adhesion (7/46, GO:0007155), and angiogenesis (3/46, GO:0001525). The differentially expressed genes were also annotated in several Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, including ECM–receptor interaction (hsa04512;

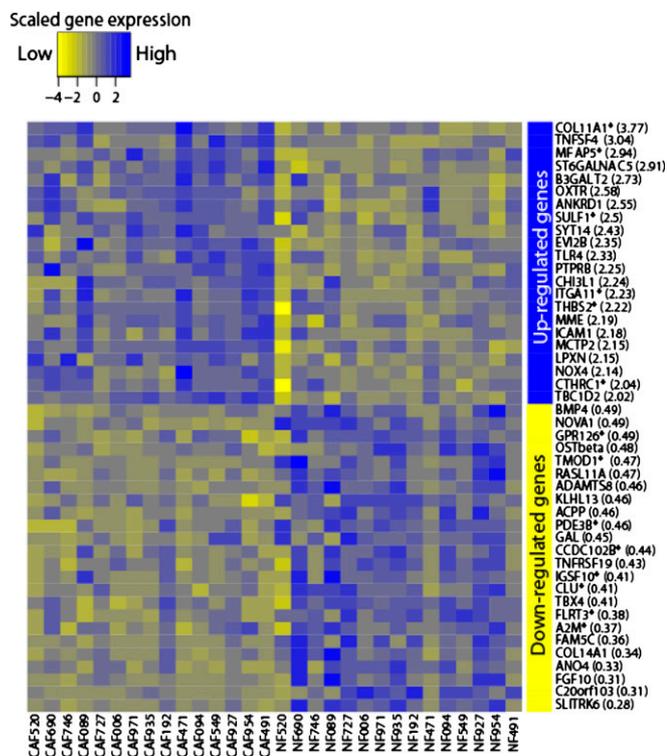


Fig. 2. Differentially expressed genes in 15 paired CAF vs. NF and in corresponding NSCLC stroma vs. normal lung. A total of 46 differentially expressed genes [22 up-regulated (blue sidebar) and 24 down-regulated (yellow sidebar)] were identified by using paired SAM analysis (absolute fold change > 2; $q < 0.1$). Heatmap plot of scaled gene-expression levels (mean centered for each gene and SD set to 1). Columns represent paired samples of CAF and NF. Rows are genes with mean fold change between CAFs and NFs (in parentheses). Asterisks denote 14 differentially expressed genes (6 up- and 8 down-regulated) overlapping with NSCLC stroma vs. normal lung analysis.

3/46; IGTA11, THBS2, and COL11A1), focal adhesion (hsa04510; 3/46; IGTA11, THBS2, and COL11A1), and TGF- β signaling pathway (hsa04350; 2/46; THBS2 and BMP4).

To evaluate the potential role of methylation in gene-expression changes occurring in CAFs vs. NFs, we conducted genome-wide methylation profiling in five pairs of CAFs and NFs using the Illumina Human Methylation 27K BeadChip (*SI Materials and Methods*). We identified a single gene in the 46 differentially expressed genes, Leupaxin (LPXN), as a candidate hypomethylated gene (*Table S2A*). We also assessed potential genomic copy number variation (CNV) or loss of heterozygosity (LOH) in CAF compared with NF, using the Illumina Human OmniExpress 12 (~700,000 SNPs) BeadChip (methodology is described in *SI Materials and Methods*). Apart from one CNV segment (single-copy loss: 46,599 bases) on chromosome 6, which was detected in two samples (CAF746 and CAF927), the remaining CNV were nonoverlapping between any of the CAF samples, ranging in size between 306 and 165,331 base pairs (*Table S2B* and *Fig. S2B–D*). Furthermore, none of the 46 differentially expressed genes between CAFs and NFs mapped in this region on chromosome 6.

Clinical Impact of CAF Genes on Prognosis. We explored the prognostic properties and clinical relevance of the genes differentially expressed in CAFs by using three published NSCLC microarray gene-expression datasets (*Table S1B*). As the training dataset, we used 218 randomly assigned patients from the National Cancer Institute Director’s Challenge Consortium (DCC) for the Molecular Classification of Lung Adenocarcinoma (10) profiled by the Affymetrix U133A chip (*Table S1C*). The other half of this dataset ($n = 218$) was used as a validation dataset. We were able to map 35 of the 46 transcript cluster IDs/genes differentially expressed in CAFs vs. NFs to 49 probe sets in U133A (*Materials*

and *Methods*). Univariate survival analysis on the training set showed that 7 of 49 probe sets (14.3%) were significantly associated with overall survival ($P < 0.05$), indicating that the CAF genes we identified were enriched for prognostic genes (*Table S1D*). To evaluate the aggregate effect of the CAF genes on patient survival, a prognostic expression signature selection algorithm [Maximizing \bar{R} Square Analysis (MARS)] was applied to the 49 probe sets to identify an expression signature (11) (*SI Materials and Methods*). For signature selection, risk score was the product of standardized gene-expression level and its univariate coefficient with survival (*Table S1D*). An 11-gene (13-probe set) combination was identified (*Table 1*). Principle component (PC) analysis identified PC1, PC2, and PC7 as being significantly associated with survival ($P < 0.05$; *Table S1E*); thus, only these three PCs were included in the final model. From here onward, and including the validation, the risk score was the product of these three PCs weighted by their coefficients in the final model (*Table S1E*). An optimal cutoff of risk score at -0.056 was identified in the training set (*Table S1E*) to distinguish the high-risk (HRISK) or low-risk (LRISK) groups and subsequently was used throughout the validation.

In the training set, the signature identified 117 LR and 101 HR patients with significantly different survival outcomes [hazard ratio (HR) = 3.38, 95% confidence interval (CI) 2.17–5.29, $P < 0.0001$, *Fig. 3A*]. In multivariate analysis, the signature remained significantly associated with survival and was independent of stage, adjuvant treatment, age, and sex (HR = 3.27, 95% CI 2.08–5.15, $P < 0.0001$, *Table S1C*). The prognostic effect of the signature was first tested in the DCC test set (*Table S1B* and *Fig. 3B*) and was able to classify 123 patients into LRISK and 95 into HRISK groups with significant difference in survival (HR = 1.57, 95% CI 1.06–2.35, $P = 0.026$), although multivariate analysis showed a trend only as an independent prognostic factor (HR = 1.37, 95% CI 0.91–2.05, $P = 0.131$, *Table S1C*). We tested the signature in two additional independent NSCLC datasets from Duke University (12) and Sungkyunkwan University (SKKU) (13) (*SI Materials and Methods* and *Table S1B*). The signature identified 43 LRISK and 46 HRISK Duke patients (HR = 1.96, 95% CI 1.05–3.62, $P = 0.031$, *Fig. 3C*), and 72 LRISK and 66 HRISK SKKU patients (HR = 1.65, 95% CI 1.03–2.65, $P = 0.039$, *Fig. 3D*). Multivariate analysis adjusting for stage, histology, age, and sex showed that the signature was independently prognostic in both datasets (Duke: HR = 2.12, 95% CI 1.11–4.02, $P = 0.022$; SKKU: HR = 1.78, 95% CI 1.08–2.93, $P = 0.024$, *Table S1C*).

Tumor Stroma Gene-Expression Profile in NSCLC. To explore the relevance of CAF-associated gene-expression changes further in the context of NSCLC stroma, we also profiled LCM tumor stroma and corresponding normal lung tissue from the same 15 NSCLCs that were used to establish CAFs and NFs. SAM analysis ($q < 10\%$; absolute fold change > 2) identified 53 up-regulated and 3,078 down-regulated genes in the tumor stroma vs. normal

Table 1. Thirteen probe sets representing 11 genes that constitute the CAF-associated prognostic gene signature

Probe set	Gene symbol	Gene title
202637_s_at	ICAM-1	Intercellular adhesion molecule 1
203083_at	THBS2	Thrombospondin 2
203434_s_at	MME	Membrane metallo-endopeptidase
206825_at	OXTR	Oxytocin receptor
208591_s_at	PDE3B	Phosphodiesterase 3B, cGMP-inhibited
208791_at	CLU	Clusterin
208792_s_at		
210121_at	B3GALT2	UDP-Gal:βGlcNAc β 1,3-galactosyltransferase, polypeptide 2
211742_s_at	EVI2B	Ecotropic viral integration site 2B
212865_s_at	COL14A1	Collagen, type XIV, α1
216866_s_at		
214240_at	GAL	Galanin prepropeptide
220603_s_at	MCTP2	Multiple C2 domains, transmembrane 2

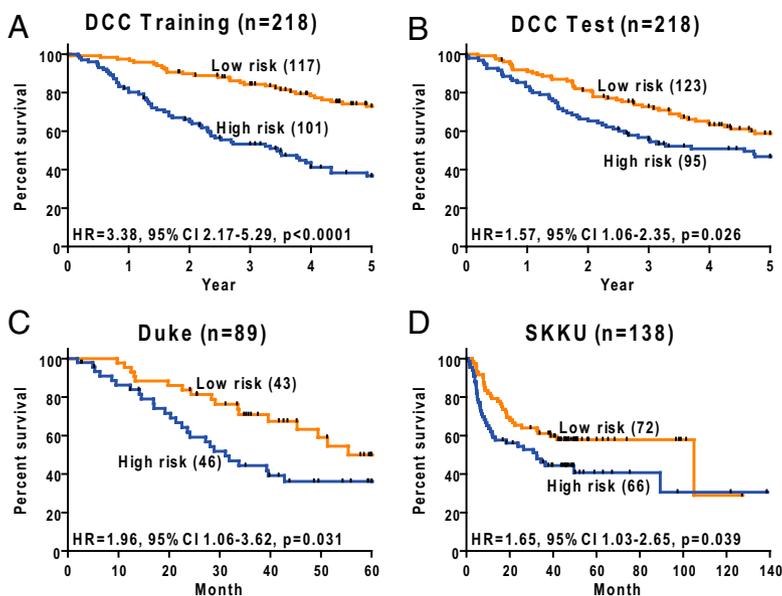


Fig. 3. Prognostic significance of CAF-associated gene-expression signature in multiple independent cohorts of NSCLC patients. (A) The 11-gene (13-probe set) signature is prognostic in 218 patients randomly assigned to training set in the DCC lung adenocarcinoma study. (B–D) By using the fixed algorithm, the prognostic value of the signature was tested on the microarray data of the remaining 218 DCC patients (assigned as testing set; B), NSCLC patients from Duke (C), and NSCLC patients from SKKU (D). Patients were classified as low-risk and high-risk groups by the signature. Significant differences in survival outcome between these two groups were observed in both the training and testing sets; 5-y overall survival was used for the DCC and Duke patients, and disease-free survival was used for SKKU patients.

lung tissue (Dataset S1 E and F). Similar to the CAF vs. NF differential gene annotation, the up-regulated genes also were enriched significantly for genes associated with the GO terms “extracellular region” (28/53, GO:0005576, $P = 2.08 \times 10^{-10}$) and “cell adhesion” (11/53, GO:0007155, $P = 0.004881$), and those associated with the KEGG pathway of ECM–receptor interaction (hsa04512, 8/46, $P = 1.5 \times 10^{-6}$), focal adhesion (hsa04510, 8/46, $P = 2.76 \times 10^{-4}$), and Reactome pathway Signaling by PDGF (REACT_16888, 5/53, $P = 0.0051$). Up-regulated genes included members of the metalloproteinase protein family (e.g., MMP-1, MMP-9, MMP-12, and MMP-14) and cytokines (e.g., CXCL13 and CXCL14). Down-regulated genes showed most significant enrichment in GO annotation for intracellular compartments and organelles [e.g., intracellular (2,234/3,078, GO:0005622, $P = 1.0 \times 10^{-84}$) and nucleus (1,075/3,078, GO:0005634, $P = 7.3 \times 10^{-24}$)], biological processes [e.g., protein metabolic process (632/3,078, GO:0019538, $P = 1.67 \times 10^{-13}$)], and RNA metabolic process (262/3,078, GO:0016070, $P = 2.8 \times 10^{-14}$).

Shared Differentially Expressed Genes of CAF and Tumor Stroma. Cross-selection of genes that were differentially expressed in NSCLC CAFs vs. NFs and tumor stroma vs. normal lung identified 14 genes in common to both analyses (six up-regulated and eight down-regulated; Dataset S1G and Fig. S3A). Fisher’s exact test showed that this overlap was significantly more than one would expect by chance ($P = 0.0045$); similarly, the overlap for up-regulated and down-regulated genes was larger than one would expect from chance alone ($P = 1.054 \times 10^{-11}$ and 0.013, respectively). Using RT-qPCR, we verified the expression of the 14 overlap genes and three of the genes at the protein level by Western blotting (Fig. S3B and SI Materials and Methods). Overall, RT-qPCR showed significant correlation with the microarray results in 13 of 14 CAF genes and 10 of 14 NF genes (Table S3 A and B). Clustering analysis based on Spearman’s correlation for gene expression in the microarray data of CAF and tumor stroma samples also demonstrated coclustering of *ITGAI1*, *THBS2*, *COL11A1*, and *CTHRC1* in both datasets (Table S3 B–D).

Although the number of overlapped genes was small, GO analysis still showed that these genes were enriched in genes of the extracellular region (9/14, $P = 0.006$) and KEGG pathways, including ECM–receptor interaction (3/14, $P = 0.005$) and focal adhesion (3/14, $P = 0.01$), which was in agreement with findings in the CAF vs. NF and tumor stroma vs. normal lung. Ten of the 14 genes were also mapped to protein–protein interactions (PPIs) in Interolog Interaction Database (I2D) v1.72 (14, 15). Nine of the 10 genes/proteins interact via shared neighboring proteins (Fig.

S3E). The resulting PPI network revealed association of these shared CAF/stroma gene products with components of signaling pathways involved in fibroblast activation, including PDGF ligands PDGFA and PDGFB interacting with A2M as well as TGF- β receptors TGF β R1 and TGF β R2 interacting with Clusterin (CLU). The proteins in this network are also enrichment for genes/proteins in the KEGG pathway of ECM–receptor interaction ($P = 0.021$) and GO biological process of cell migration ($P = 0.034$). Importantly, 7 of the 14 overlap genes were reported as transcriptional targets of the TGF- β signaling pathway (Net-Path) (16, 17). These included *ITGAI1* (18, 19), *MFAP5* (*MAGP-2/MP-25*) (20), *THBS2* (21), *A2M* (22), *CLU* (23), *CTHRC1* (24), and *SULF1* (17). To confirm this association, we demonstrate that TGF- β treatment of NF 094^{YFP^{hTERT}} cell line, which expresses low levels of these seven genes, induced the mRNA expression of all genes except *A2M* (Fig. S4).

Discussion

Several lung cancer microarray studies have reported gene-expression changes with potentially significant prognostic impact, encoding for proteins that are expressed by stromal cells (3, 25). These studies were performed on RNA isolated from whole-tumor tissues, without prior separation of stromal and tumor cells; thus, the cellular origin of these gene-expression changes was largely uncertain. To address this shortcoming directly, we profiled the gene-expression changes specifically in patient-matched paired primary cultured fibroblasts from NSCLC stroma (CAF) and the normal lung tissue (NF). To confirm that CAFs represent a valid model for this study, we first showed that CAFs differentially enhanced the invasiveness of cocultured NSCLC cells compared with NFs and also enhanced tumorigenicity of NSCLC cells lines in vivo. We demonstrated that an expression signature derived from CAF-associated genes is prognostic in multiple independent NSCLC microarray datasets. We also showed that genes differentially expressed between CAFs and NFs were also commonly differentially expressed in NSCLC tumor stroma compared with normal lung parenchyma. Furthermore, genome-wide methylation profiling of a subset of CAFs and NFs identified only 1 of 46 differentially expressed genes as a candidate site of hypomethylation. Thus, it is likely that control of gene-expression changes between CAFs and NFs would be attributed to other gene-expression control mechanisms. Moreover, analysis for CNV showed few alterations per chromosome across all autosomes in the tested samples, with only one altered segment detected in two CAF samples. These findings are consistent with previous studies conducted with SNP arrays for CAFs in breast, ovarian, and pancreatic cancers, showing no major CNV and LOH (26).

The ability to derive a tumor stroma gene expression–based prognostic predictor/signature has been reported in breast cancer (6). However, this stroma-derived prognostic predictor was not associated with a specific cell type in the stroma. To explore a gene signature that reflects the clinical relevance of a CAF gene expression, we applied the MARS algorithm (11) that we developed for identifying a minimal gene-expression set with association to survival outcome on the CAF-associated genes and tested in a published microarray dataset of a large cohort of NSCLC patients (10). The 11-gene (13-probe set) prognostic signature was significantly associated with patient survival and was validated in the testing dataset as well as two additional independent NSCLC cohorts with published microarray datasets. The demonstrated prognostic value of this NSCLC CAF gene signature underscores the biological and clinical relevance of CAF gene-expression changes in NSCLC.

One of the 14 genes shared between differentially expressed genes in CAFs vs. NFs and NSCLC stroma vs. normal lung tissue is integrin $\alpha 11$ (*ITGA11*), the protein product of which we reported previously as expressed mainly by tumor stromal fibroblasts with significant influence on the growth of NSCLC cells in vivo (27). *ITGA11*, together with other differentially expressed genes, including *CTHRC1*, *SULF1*, *MEF5*, *CLU*, and *THBS2*, are known to be regulated by the TGF- $\beta 1$ signaling pathway (18, 19), central to CAF differentiation and epithelial mesenchymal transition (28). The common regulation of these genes was further supported by the correlation in their expression in CAF (Table S4B). As seen in our comparisons of CAFs to NFs, genes that were differentially expressed in NSCLC stroma compared with normal lung were membrane-bound proteins and members of several extracellular protein families such as collagens and metalloproteases. We also found differential up-regulation of cytokines and members of the Ig protein family [*IGLV6-57*, *CD79A*, *KIR2DL3*, and I γ locus (*IGL@*)], possibly representing gene-expression changes caused by immune reactions not directly related to fibroblasts function occurring in the tumor stroma. In contrast, the down-regulated genes include mainly genes annotated to intracellular compartments (e.g., nucleus) and basic biological processes (such as GO terms “protein” and “RNA metabolic process”). The differences may reflect the difference in cell types of microdissected stroma and normal tissue because the stroma samples are selectively devoid of epithelial cells, whereas the normal lung includes both epithelial and stroma cells.

To characterize the specificity of our NSCLC stroma/CAF-associated genes, we compared them to CAF/stroma genes identified in six other studies (*SI Materials and Methods*) of breast cancer, cholangiocarcinoma (bile duct), and skin basal cell carcinoma (Table S4A and B). We found a small (23/3,110) gene overlap between NSCLC stroma/CAF differentially expressed genes and stroma/CAF characteristic genes in other tumors (Table S4B). Indeed, fibroblasts derived from different anatomical locations have been shown to be heterogeneous in gene expression (29). Nonetheless, the tumor stroma and CAF characteristic genes reported here and in previous studies (Table S4A) are largely components of the ECM and cell adhesion and potentially are involved in matrix remodeling and paracrine signaling, suggesting that similar processes occur in tumor stroma/CAF across multiple tumor types. Evaluation of the functional association of tumor stroma/CAF genes across different tumor types using a PPI network-based approach revealed high connectivity among them via shared neighboring proteins. Furthermore, we identified a subset of 55 proteins that interact with NSCLC stroma and CAF-associated proteins and proteins/genes from two or more other tumor stroma/CAF studies. Pathway annotation of these 55 interacting proteins reveals significantly enriched representation of genes from the MAPK pathway ($P = 7.7 \times 10^{-4}$) and genes/proteins involved in focal adhesion ($P = 1.7 \times 10^{-4}$) (Fig. 4). Some of these proteins are upstream components in signaling pathways known to regulate tumor stroma/CAF differentiation, including the TGF- β pathway (TGFBR1 and TGFBR2), the PDGF pathway ligands (PDGFA and PDGFB), or pathways involved in tumorigenesis in epithelial cells, e.g., ERBB receptor family-based signaling pathways (ERBB2 and EGFR).

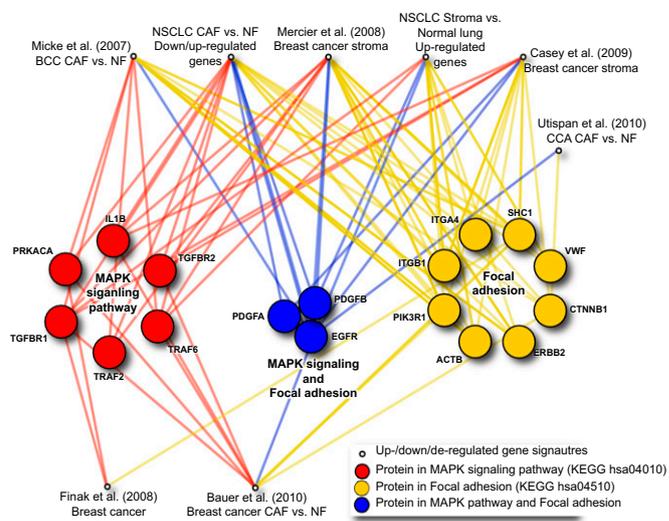


Fig. 4. Proteins encoding the CAF and tumor stroma-associated genes from seven different epithelial tumors share common involvement in the MAPK signaling pathway and focal adhesion. CAF/tumor stroma characteristic gene signatures from seven studies were mapped to PPIs. In the resulting PPI network, 55 proteins interacting with multiple NSCLC stroma and CAF proteins identified in this study (NSCLC; top) and additional tumor stroma/CAF studies from breast cancer, cholangiocarcinoma (CCA), and basal cell carcinoma (BCC) show significant enrichment in focal adhesion and the MAPK pathway ($P = 1.7 \times 10^{-4}$ and $P = 7.7 \times 10^{-4}$, respectively). Shapes (nodes) represent proteins, and lines (edges) indicate physical PPIs, colored based on annotation to KEGG focal adhesion (hsa04510; yellow). PPI network was visualized with NAViGaTOR 2.1.14; indirectly linked proteins and interactions are faded-out to reduce image complexity.

The results of our study provide direct evidence to support the important role of CAF and tumor stroma gene-expression changes in the biology and clinical outcome of NSCLC.

Materials and Methods

Methods for in vitro and in vivo assays are provided in detail in *SI Materials and Methods*. Similarly, mRNA expression and genome-wide methylation profiling as well as genomic CNV and LOH analyses are provided in *SI Materials and Methods*.

Prognostic Signature Selection. Transcript cluster IDs corresponding to differentially expressed genes in NSCLC CAF vs. NF were mapped to 49 probe set IDs in Affymetrix human chip array U133A via Affymetrix annotation and GeneAnnot v1.9 (30) (*SI Materials and Methods* and Table S1F).

Datasets used for prognostic signature discovery and validation were the adenocarcinoma (ADC) dataset from the DCC ($n = 442$) (10) and NSCLC datasets from Duke University ($n = 89$) (12) and SKKU ($n = 138$) (13). Six cases in the DCC with either unknown adjuvant treatment (adjuvant chemo therapy or radiation therapy) or unknown stage were excluded. The DCC dataset was randomized into two equal subgroups, one as training set and the other as a test set. Independent validation of the signature was carried out in the additional two NSCLC datasets. Demographic characteristics of the datasets are provided in Table S1B. The MARS algorithm was used to identify the prognostic signature (*SI Materials and Methods*).

Functional Annotation and PPI Analysis. Functional association and annotation was performed with the following annotation sources. (i) DAVID bioinformatics resources v6.7 (31) for GO terms (32) and KEGG (33) and Reactome (34) pathways annotation and enrichment analysis. Statistical significance was assessed by using a modified Fisher's exact test. Values of $P < 0.05$ (after Benjamini and Hochberg correction for multiple testing) were considered as significant enrichment. (ii) Annotation and enrichment analysis using 25 KEGG signaling pathways and cellular processes (Table S4C). (iii) Matching of differentially expressed genes to PPIs in I2D v1.72 (14, 15) with additional PPI updates. Experimental PPI networks were generated by querying I2D with the target genes/proteins to obtain their immediate interacting proteins. Relationships between the interacting proteins were added to the same network

(depth of 1 plus). The proteins in PPI networks were annotated with GO biological processes and tested for enrichment using DAVID, as described above. Significant KEGG pathways were determined for the proteins corresponding to differentially expressed genes and their interacting proteins in PPI networks, as previously described (35). PPI networks were annotated, visualized, and analyzed with NAViGaTOR v2.1.14 (36) (the complete method is described in *SI Materials and Methods*). Transcriptional targets of the TGF- β signaling pathway were annotated as per NetPath and specific literature (16, 17).

PPI Network for Stromal/CAF-Associated Genes from Other Tumors. The 46 differentially expressed genes identified in NSCLC CAF vs. NF, the 53 up-regulated in NSCLC stroma vs. normal lung, and differentially expressed stromal/CAF genes identified in six previously published studies in cholangiocarcinoma, basal cell carcinoma, and breast cancer were mapped to their corresponding Entrez gene ID (Entrez Gene 1/2010) and SwissProt accession (UniProt v57.12; total of 423 genes/proteins; *SI Materials and Methods*). Mapped proteins were used to query I2D and were matched to PPI (see *Functional Annotation and PPI Analysis*). A total of 347 PPI-matched proteins, corresponding to the tumor stroma/CAF genes (Table S4A), connected in a network composed of 3,231 nodes and 40,922 interactions. This PPI network was further analyzed in NAViGaTOR v2.1.14 to identify 55 proteins that interact with at least four proteins, corresponding to at least one gene/protein from the 46 differentially expressed genes identified in NSCLC CAF vs. NF, at least one from the 53 up-regulated in NSCLC stroma vs. normal lung, and at least two from two of the six other differentially expressed stromal/CAF genes sets. These 55 proteins were annotated by using 25 KEGG signaling and cellular processes pathways (Table S4C) to identify enrichment/overrepresentation of specific pathways compared with

their representation in the entire tumor stroma/CAF PPI network. Enrichment significance was determined with Fisher's exact test in R (v2.8.1).

Statistical Analysis. Differences in tumor growth rates of xenografts were tested by using mixed-effects model estimation (37), and a Mann-Whitney test was used to compare invasiveness levels between CAFs and NFs. The independent prognostic effect of the identified signature was tested in the training and validation datasets using the Cox proportional hazards model with the adjustments of stage, age, and sex.

Additional descriptions of the experimental procedures are described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank L. Waldron, M. Kotlyar, and A. Dubuc for discussion on data analysis, J. Xu and J. C. Ho for immunostaining, and Dr. D. Gullberg (University of Bergen, Bergen, Norway) for comments on the manuscript. This work was supported by Canadian Cancer Society Grant 019293, the Princess Margaret Hospital Foundation Investment in Research Fund and the Mary Vivian Endowment Fund, Ontario Research Fund RE-03-020 from the Ontario Ministry of Research and Innovation, Canada Foundation for Innovation Grants 12301 and 203383, Natural Science and Engineering Research Council of Canada Grant 104105, Genome Canada via Ontario Genome Institute, and IBM. B.B. was supported by the Terry Fox Foundation Strategic Training Initiative in Health Research at Canadian Institute of Health Research Grant TGT-53912 and the Ontario Institute of Cancer Research. M.-S.T. is the Choksi Chair in Lung Cancer Translational Research, I.J. a Canada Research Chair in Integrative Computational Biology, and F.A.S. the Scott Taylor Chair in Lung Cancer Research. This research is also funded in part by the Ontario Ministry of Health and Long Term Care.

- Jemal A, et al. (2009) Cancer statistics, 2009. *CA Cancer J Clin* 59:225–249.
- Demarchi LM, et al. (2000) Prognostic values of stromal proportion and PCNA, Ki-67, and p53 proteins in patients with resected adenocarcinoma of the lung. *Mod Pathol* 13:511–520.
- Nakamura H, et al. (2003) cDNA microarray analysis of gene expression in pathologic Stage IA nonsmall cell lung carcinomas. *Cancer* 97:2798–2805.
- Crawford Y, et al. (2009) PDGF-C mediates the angiogenic and tumorigenic properties of fibroblasts associated with tumors refractory to anti-VEGF treatment. *Cancer Cell* 15:21–34.
- Ramaswamy S, Ross KN, Lander ES, Golub TR (2003) A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33:49–54.
- Finak G, et al. (2008) Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med* 14:518–527.
- Serini G, Gabbiani G (1999) Mechanisms of myofibroblast activity and phenotypic modulation. *Exp Cell Res* 250:273–283.
- Hinz B, Celetta G, Tomasek JJ, Gabbiani G, Chaponnier C (2001) α -Smooth muscle actin expression upregulates fibroblast contractile activity. *Mol Biol Cell* 12:2730–2741.
- Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98:5116–5121.
- Shedden K, et al. (2008) Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat Med* 14:822–827.
- Zhu CQ, et al. (2010) Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* 28:4417–4424.
- Bild AH, et al. (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439:353–357.
- Lee ES, et al. (2008) Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. *Clin Cancer Res* 14:7397–7404.
- Brown KR, Jurisica I (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol* 8:R95.
- Brown KR, Jurisica I (2005) Online predicted human interaction database. *Bioinformatics* 21:2076–2082.
- Kandasamy K, et al. (2010) NetPath: A public resource of curated signal transduction pathways. *Genome Biol* 11:R3.
- Yue X, et al. (2008) Transforming growth factor- β 1 induces heparan sulfate 6-O-endosulfatase 1 expression in vitro and in vivo. *J Biol Chem* 283:20397–20407.
- Honda E, Yoshida K, Munakata H (2010) Transforming growth factor- β upregulates the expression of integrin and related proteins in MRC-5 human myfibroblasts. *Tohoku J Exp Med* 220:319–327.
- Lu N, et al. (2010) The human α 11 integrin promoter drives fibroblast-restricted expression in vivo and is regulated by TGF- β 1 in a Smad- and Sp1-dependent manner. *Matrix Biol* 29:166–176.
- Schedin P, O'Brien J, Rudolph M, Stein T, Borges V (2007) Microenvironment of the involuting mammary gland mediates mammary cancer progression. *J Mammary Gland Biol Neoplasia* 12:71–82.
- Kuhn I, et al. (2001) Identification of AKT-regulated genes in inducible MERAkt cells. *Physiol Genomics* 7:105–114.
- Kobie JJ, Akporiaye ET (2003) Immunosuppressive role of transforming growth factor β in breast cancer. *Clin Appl Immunol Rev* 3:277–287.
- Lee KB, et al. (2008) Clusterin, a novel modulator of TGF- β signaling, is involved in Smad2/3 stability. *Biochem Biophys Res Commun* 366:905–909.
- Pyagay P, et al. (2005) Collagen triple helix repeat containing 1, a novel secreted protein in injured and diseased arteries, inhibits collagen expression and promotes cell migration. *Circ Res* 96:261–268.
- Beer DG, et al. (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8:816–824.
- Campbell I, Qiu W, Haviv I (2011) Genetic changes in tumour microenvironments. *J Pathol* 223:450–458.
- Zhu CQ, et al. (2007) Integrin α 11 regulates IGF2 expression in fibroblasts to enhance tumorigenicity of human non-small-cell lung cancer cells. *Proc Natl Acad Sci USA* 104:11754–11759.
- Margadant C, Sonnenberg A (2010) Integrin-TGF- β crosstalk in fibrosis, cancer and wound healing. *EMBO Rep* 11:97–105.
- Kalluri R, Zeisberg M (2006) Fibroblasts in cancer. *Nat Rev Cancer* 6:582–602.
- Chalifa-Caspi V, et al. (2004) GeneAnnot: Comprehensive two-way linking between oligonucleotide array probesets and GeneCards genes. *Bioinformatics* 20:1457–1458.
- Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
- Ashburner M, et al. (2000) The Gene Ontology Consortium Gene ontology: Tool for the unification of biology. *Nat Genet* 25:25–29.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38(Database issue):D355–D360.
- Matthews L, et al. (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* 37(Database issue):D619–D622.
- Gortzak-Uzan L, et al. (2008) A proteome resource of ovarian cancer ascites: Integrated proteomic and bioinformatic analyses to identify putative biomarkers. *J Proteome Res* 7:339–351.
- Brown KR, et al. (2009) NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. *Bioinformatics* 25:3327–3329.
- Littell RC, Henry PR, Ammerman CB (1998) Statistical analysis of repeated measures data using SAS procedures. *J Anim Sci* 76:1216–1231.