DNA methylome and transcriptome landscapes of cancer-associated fibroblasts reveal a smoking-associated malignancy index

Sheng-Fang Su,[1,2,3] Hao Ho,[1] Jia-Hua Li,[1] Ming-Fang Wu,[1,4] Hsu-Chieh Wang,[5] Hsiang-Yuan Yeh,[6] Shuenn-Wen Kuo,[7] Huei-Wen Chen,[4] Chao-Chi Ho,[5] and Ker-Chau Li[1,8]

[1]Institute of Statistical Science, Academia Sinica, Taipei, Taiwan. [2]Graduate Institute of Oncology, National Taiwan University, College of Medicine, Taipei, Taiwan. [3]YongLin Institute of Health, National Taiwan University, Taipei, Taiwan. [4] Graduate Institute of Toxicology, National Taiwan University, College of Medicine, Taipei, Taiwan. [5]Department of Internal Medicine, National Taiwan University Hospital and National Taiwan University, College of Medicine, Taipei, Taiwan. [6]School of Big Data Management, Soochow University, Taipei, Taiwan. [7]Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan. [8]Department of Statistics, UCLA, Los Angeles, California, USA.

Unlike the better-studied aberrant epigenome in the tumor, the clinicopathologic impact of DNA methylation in the tumor microenvironment (TME), especially the contribution from cancer-associated fibroblasts (CAFs), remains elusive. CAFs exhibit profound patient-to-patient tumorigenic heterogeneity. We asked whether such heterogeneity may be exploited to quantify the level of TME malignancy. We developed a robust and efficient methylome/transcriptome co-analytical system for CAFs and paired normal fibroblasts (NFs) from non–small-cell lung cancer patients. We found 14,781 CpG sites of CAF/NF differential methylation, of which 3,707 sites showed higher methylation changes in ever-smokers than in nonsmokers. Concomitant CAF/NF differential gene expression analysis pointed to a subset of 54 smoking-associated CpG sites with strong methylation-regulated gene expression. A methylation index that summarizes the β values of these CpGs was built for NF/CAF discrimination (MIND) with high sensitivity and specificity. The potential of MIND in detecting premalignancy across individual patients was shown. MIND succeeded in predicting tumor recurrence in multiple lung cancer cohorts without reliance on patient survival data, suggesting that the malignancy level of TME may be effectively graded by this index. Precision TME grading may provide additional pathological information to guide cancer prognosis and open up more options in personalized medicine.

Introduction

The multistep process of tumor development accompanies the coevolution of the tumor microenvironment (TME) that engages critical participants during tumorigenesis through direct and indirect interaction with tumors (1, 2). The longstanding seed and soil theory highlights the TME in tumor metastasis and many aspects of oncogenic behaviors (3, 4). Located in the vicinity of the tumors, cancer-associated fibroblasts (CAFs) possess multifaceted capacity for supporting tumor growth both in vitro and in vivo (5–7), promoting cancer cell migration and metastasis possibly through traveling with circulating tumor cells, remodeling the extracellular matrix (ECM), and being involved in angiogenesis, inflammation, immunosuppressive, and CAF-tumor reciprocal crosstalk (8–12). CAFs maintain cancer stemness via paracrine signaling communication with cancer stem cells, contributing to tumor recurrence, drug resistance, and immune response modulation (12–15). Tar-

geting CAFs to inhibit tumor-promoting activities has arisen as a novel therapeutic strategy to complement common approaches of targeting the tumor itself (16–21).

CAFs exhibit differential gene expression patterns from their normal counterparts (22–25), and genome-wide analysis has also identified prognostic signatures in non–small-cell lung cancer (NSCLC) (26) and breast cancer (27). However, a rare incidence of point mutations or loss of heterozygosity in CAFs of breast and ovarian cancer has been demonstrated, suggesting the relatively stable genetic characteristics of nontumorous populations in the TME (28, 29). Global DNA hypomethylation and promoter hypermethylation in tumor suppressor genes have been used as prognostic markers (30, 31) and were found to correlate with altered regulation of gene expression (32). Although the role of aberrant DNA methylation has been much appreciated in cancer (33, 34), global profiling of DNA methylation in CAFs (35–37) simultaneously characterized with gene expression has been less comprehensive.

Previous studies have reported how DNA methylation aberrations in CAFs are correlated with disease resistance and prognosis (38–40). In this study, we aimed to characterize the behavior of CAFs in primary cultures replicating the lung cancer microenvironment and investigate the clinical impact of the TME through the evaluation of genome-wide DNA methylation profiling of CAFs in combination with mRNA expression analy-

sis. Given that CAFs are the major contributor within the TME, we asked if the malignancy level of the TME can be quantitated effectively via patterns of CAF methylation without the use of patients' survival data. We hypothesized that a poor-graded TME (like a bad soil) leads to poor survival of a patient, not the other way around. Our approach was to develop a methylation index that could distinguish CAFs from their matched normal fibroblasts (NFs) through a concise summary of the genome-wide methylation profile. We then illustrated how this index can be used to predict patient survival directly and confirmed the validity of its prognostic ability in 3 independent cohorts.

Results

Phenotypic heterogeneity of individual CAFs from NSCLC patients. We generated a cell library consisting of lung cancer patient–derived CAFs and matched NFs (Supplemental Figure 1 and Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/JCI139552DS1) using a previously established primary culture system (13). To examine cell-type purity of our CAF lines via flow cytometry, we performed immunofluorescence staining of 2 CAF markers, CD90 and FAP, and 1 epithelial marker, EPCAM. These CAFs were CD90 positive and FAP positive but EPCAM negative, in contrast with the A549 cancer cells, indicative of no cancer cell contamination. In addition, no *EGFR* mutation signal was detected in the CAF culture of a patient with somatic tumor cell mutation of *EGFR* (T>G mutation, L858R) using targeted sequencing, confirming no cancer cell contamination in these CAFs (Supplemental Figure 2). Proceeding with the investigation of phenotypic heterogeneity, we treated lung cancer cells (CLS1) (13) with conditioned medium (CM) from paired CAFs and NFs to examine their contributions to the tumor-promoting ability in cancer migration, invasion (Supplemental Figure 1, D and E), and sphere formation (Supplemental Figure 1F). The results reflect the considerable diversity of CAFs across patients.

Transcriptome and DNA methylome landscapes of CAFs from NSCLC patients. We developed distinct methodologies comprising several steps, each addressing a specific aim to characterize the clinical behavior of CAFs (Figure 1). Firstly, to investigate transcriptomic differences between CAFs and NFs, we profiled gene expression of the CAF/NF pairs from 25 NSCLC patients (Supplemental Table 1, discovery cohort) using the Affymetrix GeneChip Human Genome U133 Plus 2.0 array. Significance analysis of microarrays (SAM) led to the identification of 614 differentially expressed (DE) probes between CAFs and NFs with a difference in fold change greater than 1.5 at a false discovery rate control Q value of less than 0.1 (Supplemental Data File 1). Among them, 242 upregulated probes were annotated to 189 genes while 372 downregulated probes were annotated to 272 genes (Figure 2A). Figure 2B shows the distinct expression levels of DE probes for CAFs and NFs. Top DE genes were further analyzed by quantitative real-time PCR (qPCR) (Figure 2, C and D, and Supplemental Figure 3, A and B). To evaluate the biological significance of expression alteration in CAFs, pathway enrichment analysis showed ECM-receptor interaction, PI3K/Akt signaling pathway, focal adhesion, and TGF-β signaling to be highly enriched (Figure 2E and Supplemental Data File 2), suggesting a prominent mediator role for

CAFs (16). The top-ranked ECM-receptor interaction pathway featured upregulated *CD36* and *COL11A1* and downregulated *TNC* and *TNXB*, among others (Figure 2F). As a component of the TME, the ECM plays a key role in fibroblast activation and phenotypic heterogeneity, and thus its alteration can influence cancer development and progression.

Additional analyses of DE probes were conducted to investigate the correlation with clinical variables. Patients were stratified separately into different groups according to histology, cancer stage, sex, and smoking status. We compared the between-group differences in the fold change of each DE probe to determine the correlated probes, using Welch's t test ($P < 0.05$). For age, we treated it as a continuous variable and used Pearson's correlation coefficient ($P < 0.05$). The sizes of intersections among the 5 sets of DE probes correlated with clinical variables were visualized by UpSet plot (Figure 2G). Notably, the set of smoking-correlated DE probes overwhelmingly outnumbered other histological variables. Out of the 614 DE probes, 187 (30%) were significantly correlated with smoking status.

In parallel with the gene expression study, to investigate methylomic differences between CAFs and NFs, we conducted genome-wide DNA methylation profiling on the primary cultured CAF/NF pairs from 26 NSCLC patients (Supplemental Table 1, discovery cohort) using the Infinium Human Methylation 450K array. The β values of CpG sites were found to follow a bimodal distribution (Supplemental Figure 4A). Requiring the β-value difference to be greater than 0.1, we identified 14,781 differentially methylated (DM) CpG sites between CAFs and NFs with Q values less than 0.1 (Figure 3A and Supplemental Data File 3). Among them, 8,830 CpG sites (60%) exhibited hypomethylation, while 5,951 (40%) were hypermethylated in CAFs relative to NFs (Supplemental Figure 4B). Figure 3B shows the heatmap of β values of DM CpG sites for CAFs and NFs. The columns of the heatmap were firstly stratified by CAF/NF and then by patients' smoking status. Methylation patterns of some NF samples appeared similar to those of CAF samples, rendering them more susceptible to developing premalignancy in these patients. Pyrosequencing was further performed on selected DM CpG sites to quantify methylation differences between CAFs and NFs (Figure 3C).

Smoking status represents an important confounder that determines the unique CAF-specific DNA methylation pattern. We investigated how the diversity in the CAF/NF methylation difference may be correlated with the clinical characteristics of lung cancer patients. Histology, cancer stage, sex, and smoking status were used to stratify patients. Separately for each clinical variable, we compared the between-group differences in β-value change (Δβ) of each DM CpG site to determine the correlated sites, using Welch's t test ($P < 0.05$). For age, we treated it as a continuous variable and used Pearson's correlation coefficient ($P < 0.05$). The sizes of intersections among the 5 sets of DM CpG sites correlated with clinical variables were visualized by an UpSet plot (Figure 3D). Strikingly, consistent with the expression profiling, smoking had a strong correlation among clinical variables. A total of 4,972 DM sites (4,972/14,781, 34%) were significantly correlated with smoking status (never vs. ever; $P < 0.05$). Even after controlling Q less than 0.1, we still retained 3,707 smoking-associated DM sites (3,707/14,781, 25%) with larger Δβ in
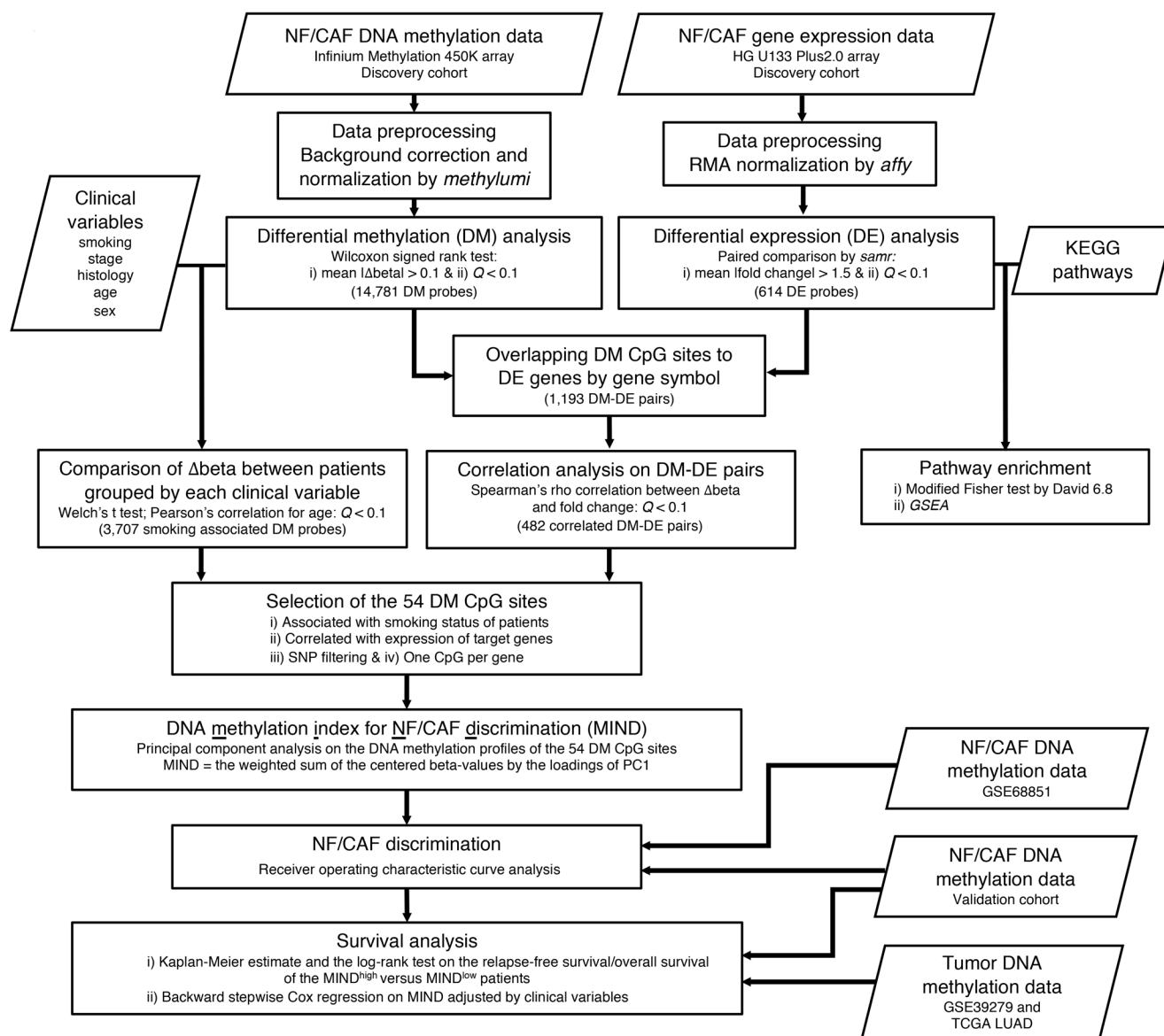
Figure 1. Flow chart of the integrative analysis of DNA methylation and gene expression profiles for the identification of MIND and its assessment in the lung cancer validation cohorts. The boxes and their corresponding notes indicate the analysis, criteria, and data set used at each step. MIND, DNA methylation index for NF/CAF discrimination; RMA, robust multiarray average; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, gene set enrichment analysis.

ever-smokers than in never-smokers, while no DM sites were retained for the other 4 clinical variables, suggesting smoking as an important player in the determination of the unique CAF-specific DNA methylation pattern.

Integration of gene expression with DNA methylation for cis-regulation analysis. To investigate how DNA methylation at CpG sites of CAF/NF may *cis*-regulate gene expression, we matched the 614 DE probes obtained earlier with 14,781 DM sites by gene symbol, leading to a total of 1,193 DM-DE pairs (Supplemental Data File 4). For each DM-DE pair, Spearman's ρ was used to evaluate the correlation between the methylation changes ($\Delta\beta$) and the expression fold changes. A total of 482 DM-DE pairs showed significant correlations ($Q < 0.1$) with the signs consistent with the ratios of the mean $\log_2$(fold change) to

the mean $\Delta\beta$. Among them, 340 (70.5%) pairs were negatively correlated (Figure 4A).

By intersecting the 482 *cis*-correlated DM-DE pairs with the 3,707 smoking-associated DM CpG sites (Figure 1), we reached a subset of 54 CpG sites after imposing exclusion criteria described in the Supplemental Methods. Hierarchical DNA methylation clustering of these 54 CpGs was conducted on 52 CAF/NF samples from 26 patients, and 4 subgroups were identified. Group I consisted of CAFs and Group IV mostly consisted of NFs (Figure 4B). Notably, we found that certain NFs or CAFs possessed specific methylation patterns, indicating that varied clinical status, such as recurrence, might be encoded. Together with the gene expression profile of the matched 54 genes shown in parallel (Figure 4B), these findings illustrate a distinct pattern of CAFs
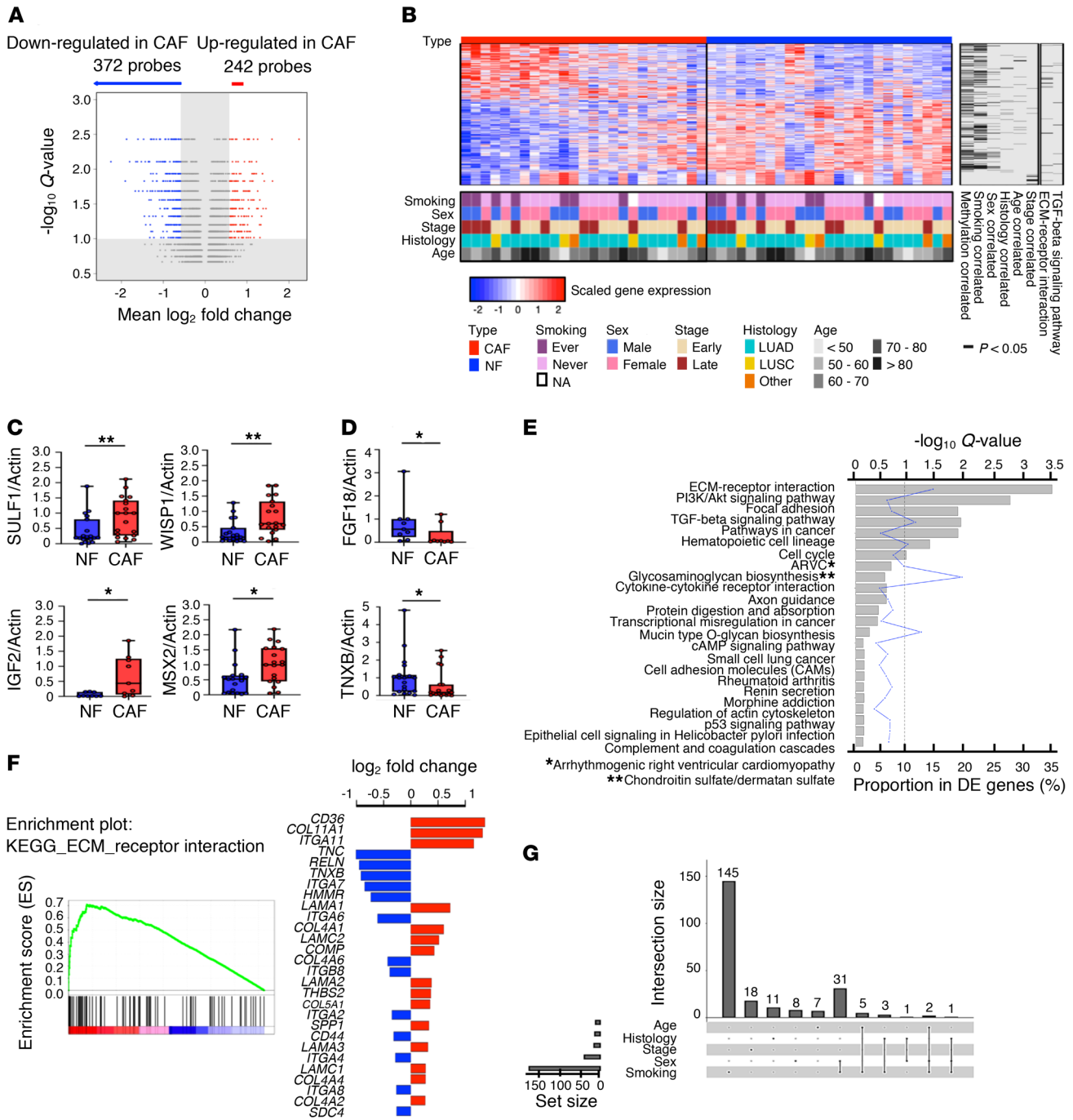
Figure 2. Identification of differential gene expression between lung NFs and CAFs. (**A**) Volcano plot of the differentially expressed (DE) probes between NFs and CAFs. A probe is significant if fold change > 1.5 and Q < 0.1. (**B**) Heatmap of 614 DE probes on 25 pairs of primarily cultured NF/CAF samples from NSCLC patients. The standardized $\log_2$ expression values are displayed. Rows represent probes and columns represent samples. The clinical characteristics are encoded in the bottom and the significantly (P < 0.05 by Welch's t test; for age, Pearson's correlation coefficient was tested using the t test) correlated DNA methylation data, clinical variables, and related pathways are indicated in the right columns. (**C** and **D**) Analysis of DE genes, upregulated (**C**) or downregulated (**D**) in CAFs, in NF/CAF pairs (n = 19 or n = 9 per gene) by quantitative real-time PCR (qPCR). Actin was used as the internal control. Data presented as mean ± SD; symbols represent individual samples. *P < 0.05, **P < 0.01 by Mann-Whitney U test. Box plots display the median, first and third quartiles with whiskers as maximum and minimum values. (**E**) Using DAVID analysis, 7 KEGG pathways were significantly enriched in the DE genes at the false discovery rate of 0.1. (**F**) GSEA enrichment plot for ECM-receptor interaction pathway in comparing CAFs (red) to NFs (blue). The $\log_2$ fold changes of the core genes are shown in the right panel. (**G**) UpSet plot for the intersections among the 5 sets of the DE probes correlated (P < 0.05 by Welch's t test; for age, Pearson's correlation coefficient was tested using the t test) with tumor histology, stage, age, sex, and patients' smoking status. The set of smoking-correlated DE probes overwhelmingly outnumbered other variables.
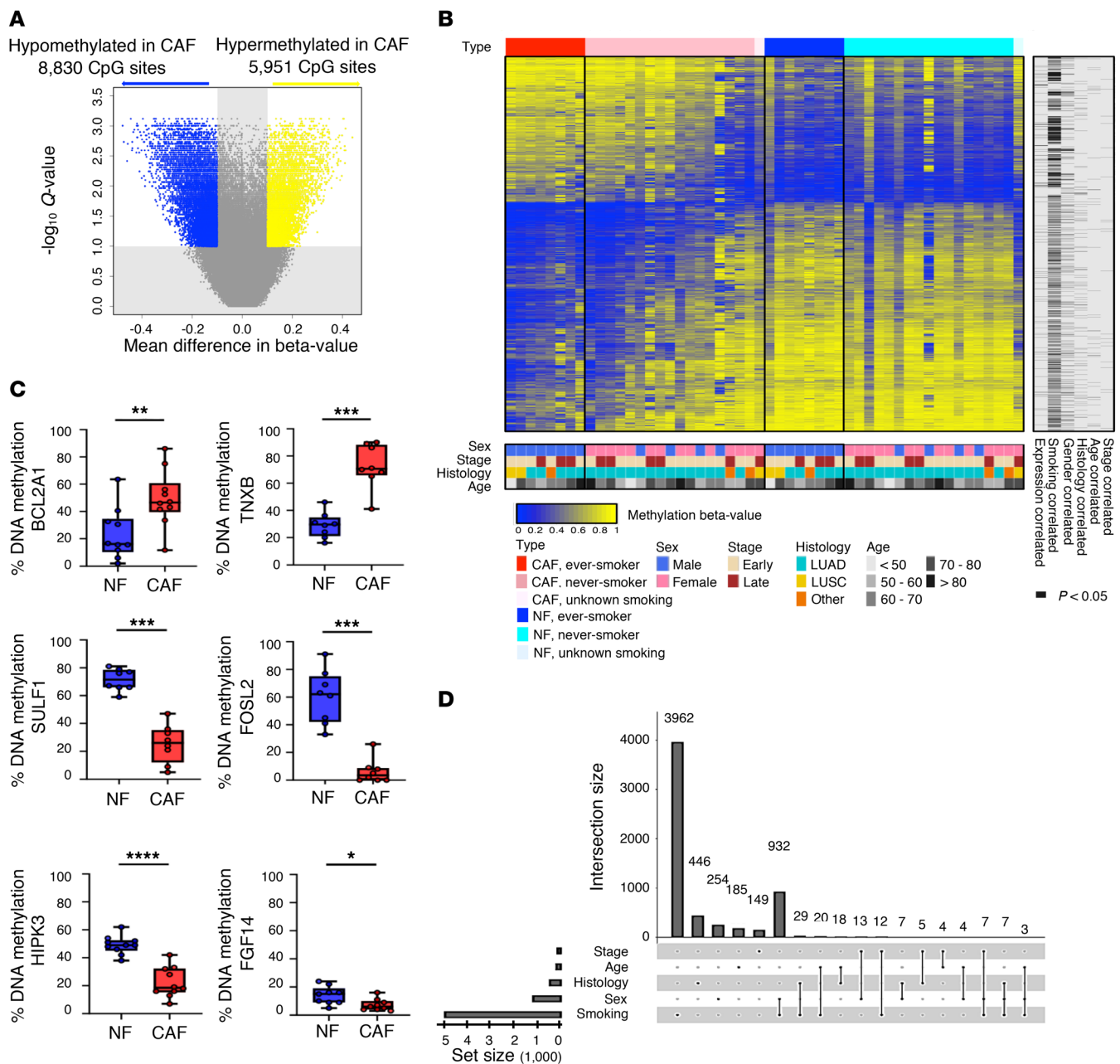
Figure 3. DNA methylation analysis of primary cultured NFs and CAFs from NSCLC patients. (**A**) Volcano plot of the differentially methylated (DM) CpG sites between NFs and CAFs. A probe is significant if the difference in β value is greater than 0.1 and Q is less than 0.1. (**B**) Heatmap of the DM probes on 26 pairs of primary cultured NF/CAF samples from NSCLC patients. The methylation level of each CpG site is presented by a β value ranging from zero to one. Unmethylated is indicated in blue, methylated is in yellow. Rows represent probes and columns represent samples. Clinical characteristics are indicated in the bottom. Samples with the same smoking status were ordered by the aggregated methylation changes of these DM sites defined by the sum of Δβ values multiplied by the sign of the average Δβ across all samples. The same ordering was applied for the matching CAF and NF samples. The significantly ($P < 0.05$ by Welch's t test; for age, Pearson's correlation coefficient was tested using the t test) correlated expression data and clinical variables are indicated in the right columns. (**C**) Analysis of DM genes hyper- or hypomethylated in CAFs in NF/CAF pairs ($n = 8$–10 per gene) by pyrosequencing. Data presented as mean ± SD; symbols represent individual samples. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ by Mann-Whitney U test. Box plots display the median, first and third quartiles with whiskers as maximum and minimum values. (**D**) UpSet plot for the intersections among the 5 sets of the DM CpG sites correlated significantly ($P < 0.05$ by Welch's t test; for age, Pearson's correlation coefficient was tested using the t test) with tumor histology, stage, age, sex, and patients' smoking status. The set of smoking-correlated DM sites overwhelmingly outnumbered other variables.

versus NFs both in methylation and expression. The β value–to–gene expression correlations for 54 genes are shown individually (Supplemental Figure 5 and Figure 4C).

DNA methylation index for NF/CAF discrimination. We performed principal component analysis (PCA) on the β values mea-

sured at the 54 CpG sites across 26 CAF/NF pairs. From the first PC (PC1, the horizontal axis; Figure 4D), a separation between CAFs and NFs was visible. We then took PC1 as a concise summary of genome-wide methylation profiles and assessed its ability to discriminate CAFs from NFs. More specifically, we constructed
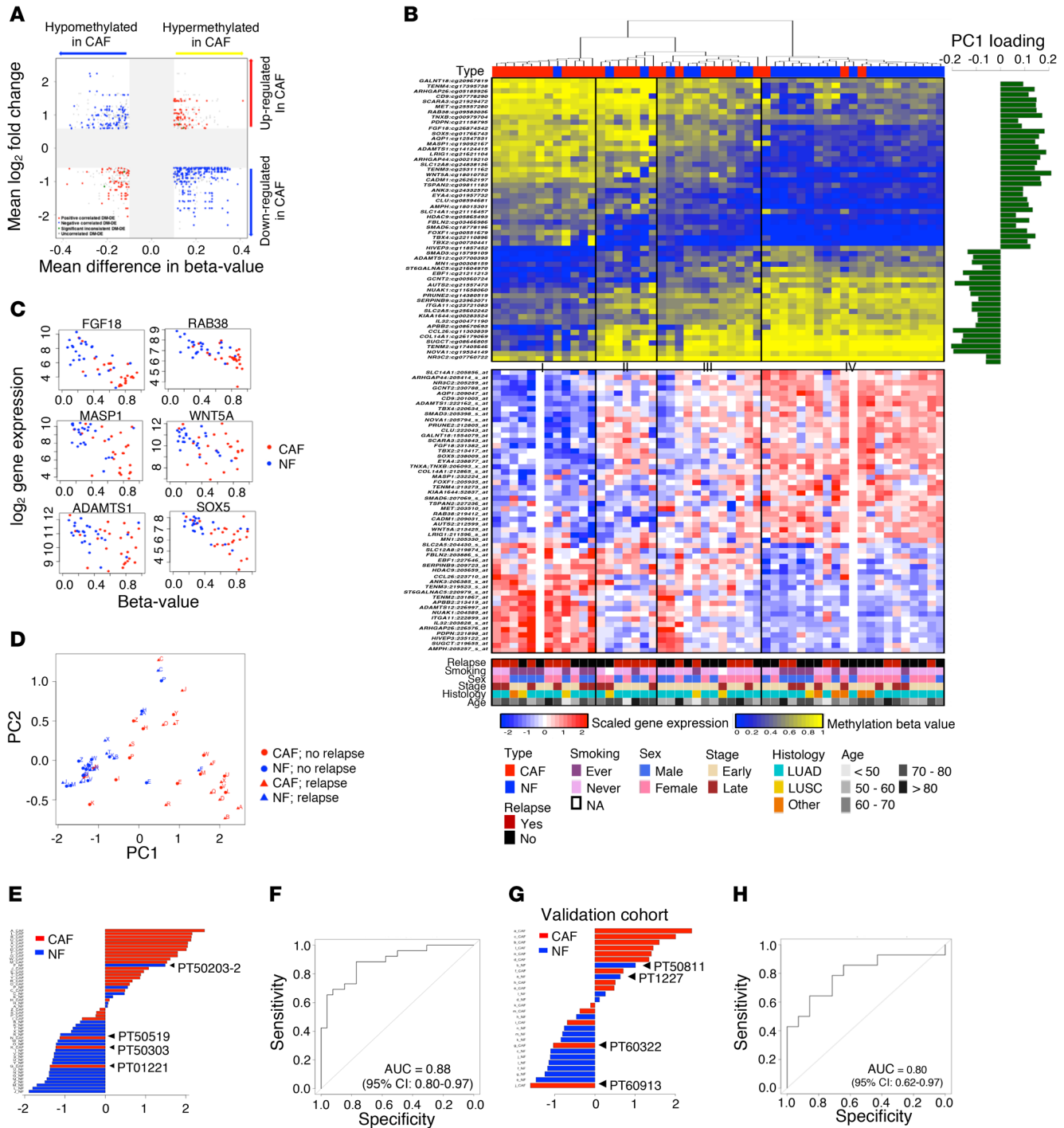
Figure 4. A distinct DNA methylation index for NF/CAF discrimination (MIND). (**A**) Scatter plot showing the mean β-value difference in DNA methylation versus the mean log₂ fold change in gene expression of the 1193 DM-DE pairs. The gray zone reflects the selective threshold. Among the 482 significantly *cis*-correlated DM-DE pairs, 340 (70.5%) pairs were negatively correlated. (**B**) Hierarchical DNA methylation clustering of the 52 NF/CAF samples from NSCLC patients based on the 54 smoking-associated CpGs (top). Unmethylated is indicated in blue, methylated is in yellow. Principal component analysis (PCA) was performed and the loading of the first component (PC1) is shown on the right. The expression profile of the matched 54 genes is also shown (bottom). Rows represent probes and columns represent samples. Clinical parameters including relapse are indicated in the bottom. (**C**) Scatter plots of DNA methylation versus gene expression in NF/CAF samples for selected genes. (**D**) Scatter plot of the first and second PCs showing DNA methylation profiles of 52 NF/CAF samples at the 54 CpG sites. Red, CAF; blue, NF. Triangles indicate samples from patients with relapse, while those with no relapse are shown as circles. (**E**) The methylation index for NF/CAF discrimination (MIND) was constructed as the weighted sum of the centered β values weighted by the loadings of PC1. The distribution of MIND, ordered from the largest value to the smallest, showed a clear partition between CAFs (red) and NFs (blue). (**F**) ROC curve showing the performance of MIND in NF/CAF classification with AUC of 0.88 (95% CI = 0.80–0.97), sensitivity of 88%, and specificity of 77% (Youden's index). (**G**) The distribution of MIND was applied to the validation cohort of NF/CAF pairs from 14 NSCLC patients. (**H**) ROC curve showing the performance of MIND in NF/CAF discrimination with AUC of 0.80 (95% CI = 0.62–0.97).
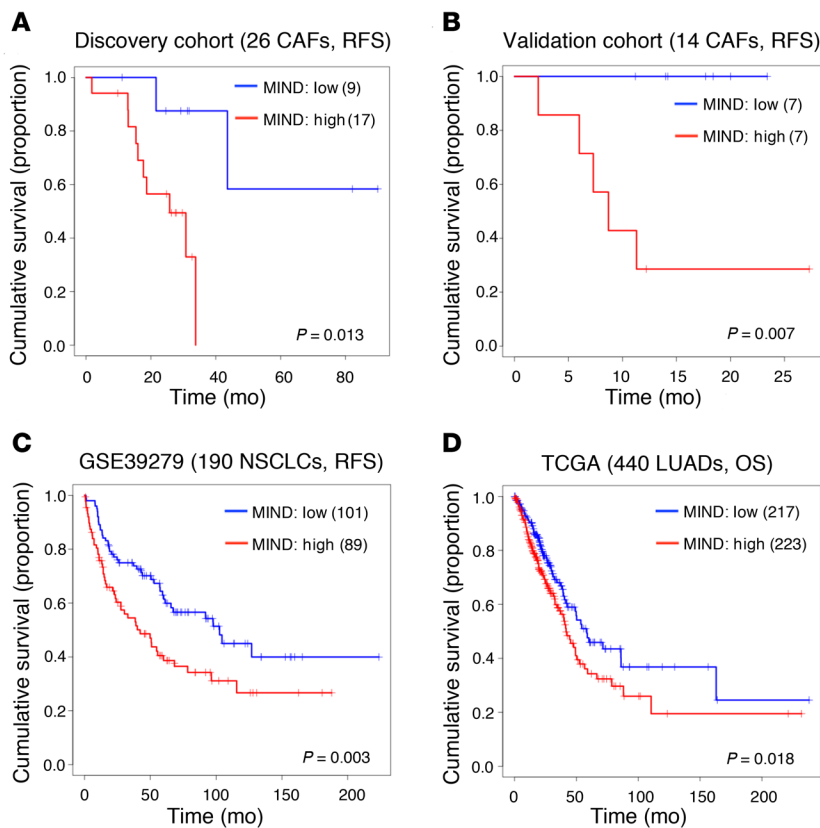
Figure 5. Prognostic significance of the smoking-associated DNA methylation signature. (A) Kaplan-Meier analysis for relapse-free survival (RFS) prediction by MIND in the 26 CAF samples. Patients were stratified by the MIND^high and MIND^low groups without reliance on survival data (*P* = 0.013, log-rank test). **(B)** Significance of MIND in recurrence prediction in the validation cohort of 14 CAF samples (*P* = 0.007, log-rank test). Patients were stratified into the MIND^high and MIND^low groups using the median as cutoff. **(C and D)** Significance of MIND for its prognostic power by applying to DNA methylation profiling of tumor samples from 2 independent cohorts, GSE39279 and TCGA-LUAD. **(C)** The 431 NCSLC patients in GSE39279 were split evenly into the MIND^high and MIND^low groups using the median as cutoff. After removing 241 patients with missing survival data, the log-rank test was performed on the remaining 190 patients (*P* = 0.003). **(D)** The 449 patients in the TCGA-LUAD cohort were split evenly into the MIND^high and MIND^low groups using the median as cutoff. After removing 9 patients with missing survival data, the log-rank test was performed on the remaining 440 patients (*P* = 0.018). OS, overall survival.

a distinct methylation index for NF/CAF discrimination (MIND) (Supplemental Data File 5) by simply summing the centered β values of the 54 CpG sites with weights determined by the loadings of PC1, as shown in Figure 4B. A discrimination between CAFs and NFs was evident using MIND (Figure 4E). Interestingly, further inspection of the NF sample that was mixed with CAFs revealed that this patient had developed recurrence (PT50203-2), indicating suspicious malignancy of the NFs. In contrast, the CAFs that were mixed with NFs came from relapse-free patients (PT50519, PT50303, and PT01221). Our findings suggest that MIND may have the potential to robustly detect premalignancy across individual patients.

The performance of MIND was evaluated by receiver operating characteristic (ROC), showing an AUC of 0.88 (95% CI = 0.80–0.97) with 88% sensitivity and 77% specificity at the optimal cutoff by Youden's index (Figure 4F). To validate the CAF/NF discriminatory ability of MIND, we used a validation cohort of paired CAFs/NFs cultured from 14 NSCLC patients (Figure 4G

and Supplemental Table 1), yielding an AUC of 0.80 (95% CI = 0.62–0.97) (Figure 4H). Another validation cohort used was a public data set with paired CAFs/NFs from 12 NSCLC patients (GSE68851) (37) and MIND also yielded good discrimination (AUC = 0.83, 95% CI = 0.66–1.0, 73% sensitivity and 92% specificity; Supplemental Figure 6).

Furthermore, we also performed PCA on the mRNA expression profile of the same 54 genes and produced a gene expression index (MIND-GE), which is the sum of the standardized expression with weights determined by the loadings of PC1. MIND-GE also showed high power in distinguishing CAFs from NFs (AUC = 0.83, 95% CI = 0.70–0.95; Supplemental Figure 7A) and this finding was consistent in another independent cohort of CAF/NF pairs from 15 NSCLC patients (GSE22874) (26) (AUC = 0.93, 95% CI = 0.84–1.0; Supplemental Figure 7B).

The potential of MIND in quantifying TME malignancy. To explore whether MIND could serve as a useful summary of the CAF methylome for grading the degree of TME malignancy across individual patients, we performed phenotypic studies to evaluate how cancer cells may react differently to the heterogeneous cell culture environment contributed by CAFs with different MIND scores. We collected CM from the CAFs with higher (MIND^high) or lower (MIND^low) MIND scores and applied them to 2 lung cancer cell lines (A549 and CL1-0). Assays of cell viability and invasion were performed to better clarify the objective and potential impact of MIND. The results showed that both the viability and invasive abilities of the cancer cells were highly promoted with the treatment of CM from MIND^high CAFs compared with that from the MIND^low CAFs (Supplemental Figure 8), suggesting that the malignancy level of CAFs can be quantitated effectively via MIND.

According to the seed and soil theory, a poor-graded TME (like a bad soil) leads to the poor survival of a patient; therefore, we used an objective way to separate the patients of our discovery cohort into a poor-TME group and a good-TME group by MIND (without the use of patient survival data). We computed the cutoff point for MIND to separate CAFs from NFs by controlling the probability of misclassifying NFs at no more than 5% while maximizing the probability of correct classification of CAFs. Using this cutoff point, we assigned the patients with CAF MIND scores that were higher than the cutoff point (MIND^high) to the poor-TME group and kept the rest (MIND^low) in the good-TME group.

Prognostic performance of MIND. Using the longitudinal patients' follow-ups of tumor recurrence in the discovery cohort, we compared the relapse-free survival (RFS) curves of the poor-TME and good-TME groups, stratified by the objective split described above, to assess the clinical significance of MIND. The result showed that MIND^high patients had poor outcomes (*P*

Table 1. Cox regression analysis of MIND for lung cancer recurrence

	Discovery cohort (CAF, RFS; $n = 26$)					
	Univariate Cox model			Backward stepwise multivariate Cox model		
Variable	HR	95% CI	P	HR	95% CI	P
MIND	9.35	1.15–75.86	0.036	9.29	1.14–75.44	0.037
Stage	1.58	0.44–5.69	0.484	-	-	-
Age	1.03	0.98–1.09	0.214	-	-	-
Sex	0.94	0.30–2.98	0.917	-	-	-
Smoking	1.89	0.55–6.52	0.315	-	-	-
	Validation cohort (CAF, RFS; $n = 14$)[A]					
	Univariate Cox model			Backward stepwise multivariate Cox model		
Variable	HR	95% CI	P	HR	95% CI	P
MIND	16.15	1.8–2127.90	0.009	29.17	2.19–6,520.53	0.006
Stage	6.52	1.17–66.22	0.032	13.9	1.41–1,796.42	0.021
Age	0.97	0.86–1.09	0.628	-	-	-
Sex	1.81	0.35–10.96	0.471	-	-	-
Smoking	2.63	0.51–15.97	0.243	-	-	-
	GSE39279 (tumor, RFS; $n = 190$)					
	Univariate Cox model			Backward stepwise multivariate Cox model		
Variable	HR	95% CI	P	HR	95% CI	P
MIND	1.83	1.22–2.75	0.004	2.00	1.32–3.04	0.001
Stage	4.20	2.63–6.71	<0.001	4.40	2.74–7.08	<0.001
Age	0.98	0.97–1.00	0.088	0.98	0.96–1.00	0.056
Sex	1.40	0.93–2.11	0.108	1.6	1.05–2.41	0.027
Smoking	1.05	0.57–1.92	0.883	-	-	-
	TCGA-LUAD (tumor, OS; $n = 440$)					
	Univariate Cox model			Backward stepwise multivariate Cox model		
Variable	HR	95% CI	P	HR	95% CI	P
MIND	1.46	1.07–2.01	0.018	1.42	1.02–1.97	0.037
Stage	2.45	1.74–3.41	<0.001	2.27	1.60–3.22	<0.001
Age	1.01	0.99–1.02	0.412	-	-	-
Sex	1.06	0.78–1.45	0.707	-	-	-
Smoking	1.01	0.64–1.59	0.968	-	-	-

[A]Firth's penalized likelihood method was used for this small data set. n, number of patients with nonmissing value of MIND and survival data; RFS, relapse-free survival; OS, overall survival; HR, hazard ratio; CI, confidence interval.

= 0.013, log-rank test; Figure 5A). Multivariate Cox regression confirmed the significance of MIND adjusted by cancer stage, age, sex, and smoking status (HR = 9.29, 95% CI = 1.14–75.44, P = 0.037; Table 1). We also applied MIND to our validation cohort of 14 patients. We ranked patients according to the MIND scores of their CAFs and split them into 2 groups of equal size. RFS was analyzed by log-rank test (P = 0.007; Figure 5B). Multivariate Cox regression further confirmed the prognostic ability of MIND adjusted by clinical variables (HR = 29.17, 95% CI = 2.19–6,520.53, P = 0.006; Table 1).

By searching public databases for clinical validation of MIND, we found that the GSE68851 we used earlier did not have survival data, and no large cohort of CAF methylome data with survival was available. Instead, 2 cohorts, GSE39279 (31) and The Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD), have methylome data on the tumors, instead of CAFs. Due to the impurity of tumor samples, some CAF content may be retained in the resected tumor tissue and profiled together with the bulk of tumor cells. Although the signals from CAFs would have deteriorated and the prognostic performance of MIND could be compromised, we still applied MIND to the 2 cohorts. For each patient, the MIND score was calculated. With the median cutoff, patients were split evenly into 2 groups (MIND[high] vs. MIND[low]) without the input of survival data. After that, those with no survival data were removed from further survival analysis. Promisingly, the results showed that patients with high MIND scores had significantly shorter RFS (P = 0.003 in GSE39279; Figure 5C) and shorter overall survival (OS) (P = 0.018 in TCGA-LUAD; Figure 5D) than those with low MIND scores. This finding was further confirmed by univariate and multivariate Cox regression (Table 1). Conclusively, the MIND index could be used to define high-risk populations for early detection of recurrence. It has the potential to detect premalignant TMEs in patients with different clinical status.

In addition to DNA methylation, MIND-GE also showed significant performance in recurrence prediction in our discovery cohort using the median as cutoff (P = 0.041; Supplemental Figure 9A). Applying MIND-GE to the gene expression profiling data of the TCGA-LUAD cohort found P = 0.069 in OS prediction (Supplemental Figure 9B). The result showed that MIND outperformed MIND-GE in the prognostic assessment.

Discussion

The integrated DNA methylome and transcriptome landscape study of CAFs we present here shows that the malignancy level of the TME can be quantified via a concise summary of CAF DNA methylation. A set of 54 smoking-associated CpG sites was found and MIND, a weighted sum of methylation levels on these sites, can be used to assign TME malignancy grades to patients. During the discovery stage of identifying MIND, no patient survival data were employed. At the validation stage, we investigated the clinical relevance of this index and found that higher MIND scores significantly correlated with shorter survival in our patient cohort and another 2 independent cohorts. Our study via exploring and exploiting the definitive behavior of CAFs helps score patients with the risk of tumor relapse. Poor-graded patients are more likely to have poor prognosis. The ability of MIND to predict cancer patient survival is important supportive evidence that MIND does reflect the malignancy of the TME.

During the course of investigation, we successfully established cell libraries of paired CAFs/NFs from NSCLC patients for reproducible interrogation. In spite of the increasing interest in studying genome-wide profiling of patient-derived CAFs (26, 37) for potential biomarker development (41) and in targeting CAFs for personalized therapy (42), progress has been impeded by a deficit in our understanding of CAF heterogeneity at the molecular level among patients. A clear and comprehensive molecular portrait of the tumorous field defect is relatively lacking (43). As epigenetics has a role in driving TME function (44), we developed the unique lung CAF/NF molecular landscape by conducting global DNA methylation and gene expression profiling, and elucidated how the cross-patient heterogeneity in the aberrant methylome and transcriptome correlated with clinical variables. In particular, significant DM CpG sites were found to be correlated with smoking status.

Our results point to the special impact of smoking on TME, suggesting that smoking not only could incur disarrangement in cancer itself, but also may disrupt the physiological functions of surrounding CAFs. The findings are consistent with earlier reports. Vizoso et al. observed a correlation of *SMAD3* hypermethylation in CAFs of 20 current-smoker NSCLC patients with TGF-β1 and ECM expression (37). Increased *SMAD3* methylation in CAFs of squamous cell carcinoma was also reported as a possible mechanism of resistance to nintedanib, which targets TGF-β1/SMAD3 (39). Smoking-associated DNA methylation changes could contribute to lung cancer development (45), thus acting as predictors in tumor screening (46, 47). Moreover, smoking could induce extensive damage in normal-appearing tissue adjacent to neoplastic lesions, indicating a phenomenon of smoking-related field cancerization (48). Although most studies focus on smoking effects on cancer cells, Teschendorff et al. demonstrated that DNA methylation changes in buccal cells exposed to smoke correlated with smoke-related epithelial cancers, notably lung cancer, indicating a correlation between the smoking-associated DNA methylation signature and preinvasive lung lesion development (49). In a meta-analysis of genome-wide methylation of blood DNA, comparison between current- versus never-smokers identified DM CpG sites that were annotated to genes involved in pulmonary function, cancers, and inflammatory diseases (50). These results suggest that DNA methylation sites could serve as biomarkers or therapeutic targets for tobacco-related diseases.

The connection between smoking and functions of the genes annotated to the 54 CpG sites of MIND can also be investigated from the perspective of chemical exposure. As a preliminary attempt, we applied the 54 genes to the Comparative Toxicogenomics Database (CTD), a database of curated information that includes chemical-gene interactions. The results showed a correlation between these genes and tobacco smoke pollutants, nicotine, the carcinogen benzo[a]pyrene, and particulate matter, indicating the sensitivity of these genes to the smoking-related TME (Supplemental Table 2).

In addition, we investigated the 54 genes in terms of their contribution to the biological mechanisms underlying tumor progression and found the participation of *COL14A1*, *FGF18*, *FBLN2*, *ADAMTS1*, *ITGA11*, and *IL32* in ECM/integrin signaling; *SMAD3*, *PDPN*, *WNT5A*, *ANK3*, *HIVEP3*, *TBX2*, *TENM3*, *FBLN2*, *NUAK1*, and *AUTS2* in TGF-β/Wnt/β-catenin signaling; *CADM1*, *TNXB*, and *AMPH* in PI3K/Akt signaling, contributing to early tumorigenesis, malignant transformation, angiogenesis, tumor cell migration/invasion, and metastasis; as well as immune modulation (Supplemental Table 2). PDPN has been reported as an immune suppressive factor in the TME through suppressing CD4+ effector T cell responses (51), and CCL26 could increase tumor-associated macrophages (52). In particular, 4 (*CLU*, *TBX4*, *COL14A1*, and *NOVA1*) and 10 (*TBX4*, *COL14A1*, *NOVA1*, *TENM3*, *AQP1*, *ARHGAP26*, *TBX2*, *EYA4*, *APBB2*, and *ST6GAL-NAC5*) of the 54 genes were in agreement with a study of the CAF gene expression signature in NSCLC (26) and a study of aberrant DNA methylation in lung CAFs (37), respectively. We also found that 14 (*TBX2*, *TBX4*, *WNT5A*, *EYA4*, *CLU*, *ADAMTS1*, *SCARA3*, *AQP1*, *TNXB*, *CADM1*, *LRIG1*, *CD9*, *SLC14A1*, and *FOXF1*) of the 24 hypomethylated and upregulated genes in CAFs have been reported as tumor suppressors (Supplemental Table 2). WNT5A, a cell migration and invasion factor secreted by gastric CAFs, has been associated with poor prognosis, further advancing epigenetic targets for anticancer therapy (40). Importantly, COL14A1, WNT5A, CD9, and PDPN are defined as specific markers for lung stromal cell subtypes and may play critical roles in lung fibrosis, tumor malignancy, and antitumor immunity (53–55). As for how these constituent genes of MIND may multiplicatively contribute to TME function, protein-protein interaction analysis was also performed to elucidate their potential connections (Supplemental Figure 10). Furthermore, we benchmarked MIND against other molecular signatures, including DNA methylation– and gene expression–based CAF-associated signatures (Supplemental Table 3). The result showed that MIND, derived without survival data, can reflect the malignancy level of CAFs more accurately than other molecular signatures derived from survival data.

In conclusion, we took advantage of the availability of high-quality DNA and RNA from the same samples for the methylome and transcriptome assays and succeeded in quantifying the protumorigenic potency of CAFs in NSCLC patients. By mapping the epigenetic profiles of individual CAFs, we showed that MIND could reveal the varying CAF characteristics by our integrative analysis of CAFs/NFs and demonstrated

how it could help to detect the disordered or premalignant TME. The identification of DNA methylation–based predictive markers may open a new direction in TME-targeted personalized medicine.

Methods

Study population. This study recruited 26 NSCLC patients (discovery cohort) and 14 NSCLC patients (validation cohort) who underwent surgical resection at the National Taiwan University Hospital. Written informed consent was obtained from each patient, along with the histoclinical information, including cancer type and stage, smoking behavior, and recurrence status by pathologist's assessment. The TNM staging system was applied according to the 7th edition of the American Joint Committee for Cancer (AJCC) staging system (56) (Supplemental Table 1).

Tissue collection and primary culture system. We established a cell library of CAFs from surgically resected lung tumor tissues of NSCLC patients. The normal tissue counterparts were obtained at least 5 cm away from the neoplastic lesions. The primary cultures of lung CAFs and the paired NFs were processed as previously described (13). Briefly, the tissues were minced and digested with deoxyribonuclease I (BioShop) and collagenase I (Life Technologies) at 37°C for 1 hour. After centrifugation, filtered cells were cultured in 6-well plates with DMEM (Life Technologies) containing 10% FBS (Gibco) at 37°C under 5% CO_2. In addition, corresponding CM from NFs and CAFs was carefully collected for the analysis of tumor-promoting capacity.

Cell lines. The human lung adenocarcinoma cell line A549 was purchased from American Type Culture Collection (ATCC, CCL-185), CL1-0 was provided by Pan-Chyr Yang (Department of Internal Medicine, National Taiwan University Hospital), and CLS1 was established in our previous study (13). The human bronchial epithelial cell line BEAS-2B was obtained from ATCC (CRL-9609).

Migration and invasion assays. Cells (5×10^4) were seeded onto Transwells (Corning) covered by the CM from cultured NFs and CAFs. After 14–16 hours, the number of cells that migrated/invaded through the basement membrane extract was quantified. For the invasion assay, the membrane was coated with Matrigel (R&D Systems).

Sphere-forming assay. Eight hundred cells per well were seeded on ultra-low-adherence 24-well plates (Corning) with MCDB201 medium (Invitrogen) and exposed to the CM from cultured NFs and CAFs. The medium was supplemented with epidermal growth factor (EGF, 20 ng/mL; Sigma-Aldrich) and basic fibroblast growth factor (bFGF, 20 ng/ml; Invitrogen). Images of the spheres were acquired and sphere sizes were calculated using the high-throughput imaging-based high-content platform, ImageXpress Micro4 (Molecular Devices).

Flow cytometry assay. Cells (1×10^6) from cultures were freshly prepared and incubated with fluorochrome-labeled antigen-specific antibodies or isotype controls in PBS at 4°C for 30 minutes. Cells were washed twice with PBS, and then analyzed by flow cytometry (LSR-Fortessa, BD Biosciences). Live and dead cells were distinguished by a live/dead fixable dead cell stain kit (Invitrogen). The following mouse anti-human antibodies were used: PE-CD90 (BD Biosciences, 555596, clone 5E10), Alexa Fluor 647–α/FAP (R&D Systems, FAB3715R-100UG, clone 427819), and BV421-CD326 (BD Biosciences, 563180, clone EBA-1). PE–mouse IgG1, κ (BD Biosciences, 559320, clone MOPC-21), Alexa Fluor 647–mouse IgG1 (R&D Systems,

IC002R, clone 11711), and BV421–mouse IgG1, κ (BD Biosciences, 562438, clone X40) were used as isotype controls.

Sample preparation. Genomic DNA and total RNA of collected samples were isolated with an AllPrep DNA/RNA Mini kit (Qiagen). One microgram of genomic DNA was subjected to bisulfite conversion using EZ DNA Methylation kits (Zymo Research) according to the manufacturer's protocol. Total RNA was extracted using a standard TRIzol (Invitrogen) protocol.

Gene expression analysis. Purified RNAs from the same NF/CAF samples as DNAs from 25 NSCLC patients were processed on the Affymetrix GeneChip Human Genome U133 Plus 2.0 Array, carrying 54,675 probes and covering 21,649 genes. Robust multiarray average (RMA) normalization was applied by using the package "affy" (57) in R. Probes on the sex chromosomes and chromosomes or gene symbols annotated as "-" were removed, yielding 41,158 probes. SAM was performed for identifying DE genes between paired CAF and NF samples using the package "samr" (58) in R.

Pathway enrichment analysis. We used DAVID Bioinformatics Resource 6.8 (59) to identify pathways annotated by Kyoto Encyclopedia of Genes and Genomes (KEGG) (60) enriched in the set of DE genes. We also applied gene set enrichment analysis (GSEA) (61) on KEGG pathways for comparing CAFs and NFs.

Genome-wide DNA methylation analysis. After passing the bisulfite conversion quality control, 52 bisulfite-treated DNAs from 26 lung NF/CAF pairs were processed onto the Illumina Infinium HumanMethylation450 BeadChip, which covers 485,460 CpG sites. The validation cohort consisting of 28 bisulfite-treated DNAs from 14 lung NF/CAF pairs was also processed. Background correction and normalization were performed using the package "methylumi" (62) in R. CpG sites on the sex chromosomes and with missing values were removed, yielding 472,676 CpG sites. The β value (0 to 1.0) of each CpG site represents the level of methylation as determined by the ratio of the methylated signal intensity (*M*) to the total intensity of unmethylated (*U*) and methylated (*M*) signals ($M/[U + M]$). Wilcoxon's signed rank test was used for DM CpG sites between paired NFs and CAFs. The Benjamini-Hochberg method was used to control the false discovery rate.

Correlation analysis of DM and DE. A DM probe and a DE probe were linked as a DM-DE pair when they were annotated to the same gene symbol. Spearman's rank correlation (*ρ*) was computed for each DM-DE pair to examine the correlation between the β-value differences and the expression fold changes across patients.

Pyrosequencing. Ten nanograms of bisulfite-converted DNA was used for PCR amplification with biotin-labeled primers (PyroMark PCR Kit, Qiagen). After binding to streptavidin-Sepharose beads (GE Healthcare), sequencing primers were used to anneal to the single-stranded PCR products and methylation was quantified at single-base resolution using the PyroMark Q96 Pyrosequencing System (Qiagen). The degree of methylation was defined as 0% to 100% at each CpG site by calculating the percentage of methylated cytosine. The primer sets for PCR amplification and sequencing were designed using the PyroMark Assay Design Software 2.0.

Quantitative RT-PCR analysis. cDNA was prepared and reverse transcribed from total RNA with iScript cDNA (Bio-Rad). RT-PCR was performed using SYBR green (Bio-Rad) on a Bio-Rad CFX Connect Real-Time Thermal Cycler. The mRNA expression was measured and normalized to the internal control.

TruSight Tumor 15 targeted next-generation sequencing. Fifteen of the commonly mutated genes in solid tumors, including EGFR, were sequenced by Illumina TruSight Tumor 15. We collected the tumor part tissue, the normal part tissue, together with the corresponding primary cultured NF and CAF cell lines from a NSCLC patient with somatic tumor cell mutation of EGFR L858R (T > G mutation). Twenty nanograms of DNA for each sample was prepared and sequenced on the Illumina MiniSeq System. A signal of T to G mutation on EGFR in the tumor DNA was detected.

Statistics. Welch's *t* test was used to assess methylation differences and gene expression differences between different patient groups, as stratified by the category variables: sex, cancer stage (early/late), smoking status (never/ever), histology (LUAD vs. lung squamous cell carcinoma [LUSC]); and Pearson's correlation coefficient was used for the continuous variable, age. PCA was used for finding the leading methylation indices. ROC curve analysis was performed and the optimal cutoff was determined by Youden's index. Backward stepwise selection for Cox's proportional hazard model was applied to test the prognostic ability of the methylation index under the adjustment of age, cancer stage, smoking status, and sex. The Kaplan-Meier method was used to estimate the RFS of patients grouped by the methylation index and the differences between the groups were analyzed with the log-rank test. All the statistical tests were 2-sided, with *P* less than 0.05 considered statistically significant. For all the variable selection steps, the Benjamini-Hochberg method was used to control the false discovery rate for multiple testing. The package "cox-phf" in R was used to fit the Cox regression with Firth's penalized likelihood with the best dichotomization for the validation cohort of 14 NF/CAF pairs (63). The package "UpSetR" in R was used for the UpSet plot (64).

Public data sets. DNA methylation data of 12 pairs of lung NFs/CAFs were downloaded from the NCBI's Gene Expression Omnibus (GEO) database with accession number GSE68851 (37). DNA methylation data of tumor samples from NSCLC patients were downloaded from GEO with accession number GSE39279 (31). Gene expression data of 15 pairs of lung NFs/CAFs were downloaded from the GEO database (GSE22874) (26). DNA methylation/gene expression data of tumor samples from lung adenocarcinomas were downloaded from TCGA data portal (65). DNA methylation data and newly curated clinical data (66) were downloaded from TCGA GDC data portal (https://gdc.cancer.gov/about-data/publications/pancanatlas).

Study approval. This study was approved by the National Taiwan University Hospital IRB (IRB201106046RC). Written informed consent was obtained from each patient.

Author contributions

SFS, HH, CCH, and KCL designed research studies. SFS, JHL, and MFW conducted experiments. HCW, SWK, and CCH acquired clinical samples and patient records. HH and HYY conducted bioinformatic and statistical analyses. SFS, HH, HWC, CCH, and KCL prepared and revised the manuscript. All authors gave final approval of the manuscript.

Acknowledgments

Address correspondence to: Ker-Chau Li, Institute of Statistical Science, Academia Sinica, 128 Academia Road, Section 2, Taipei 11529, Taiwan. Phone: 886.2.2787.5696; Email: kcli@stat.sinica.edu.tw. Or to: Chao-Chi Ho, Department of Internal Medicine, National Taiwan University Hospital and National Taiwan University College of Medicine, 7 Chung-Shan South Road, Taipei 10051, Taiwan. Phone: 886.2.23562905; Email: ccho1203@ntu.edu.tw.

1. Hu M, Polyak K. Microenvironmental regulation of cancer development. *Curr Opin Genet Dev*. 2008;18(1):27–34.
2. McAllister SS, Weinberg RA. Tumor-host interactions: a far-reaching relationship. *J Clin Oncol*. 2010;28(26):4022–4028.
3. Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat Rev Cancer*. 2003;3(6):453–458.
4. Langley RR, Fidler IJ. The seed and soil hypothesis revisited--the role of tumor-stroma interactions in metastasis to different organs. *Int J Cancer*. 2011;128(11):2527–2535.
5. Shiga K, et al. Cancer-associated fibroblasts: their characteristics and their roles in tumor growth. *Cancers (Basel)*. 2015;7(4):2443–2458.
6. Madar S, et al. 'Cancer associated fibroblasts'- more than meets the eye. *Trends Mol Med*. 2013;19(8):447–453.
7. Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer*. 2016;16(9):582–598.
8. Duda DG, et al. Malignant cells facilitate lung metastasis by bringing their own soil. *Proc Natl Acad Sci U S A*. 2010;107(50):21677–21682.
9. Kakarla S, et al. Cancer-associated fibroblasts as targets for immunotherapy. *Immunotherapy*. 2012;4(11):1129–1138.
10. Vicent S, et al. Cross-species functional analysis of cancer-associated fibroblasts identifies a critical role for CLCF1 and IL-6 in non-small cell lung cancer in vivo. *Cancer Res*. 2012;72(22):5744–5756.
11. Karagiannis GS, et al. Cancer-associated fibroblasts drive the progression of metastasis through both paracrine and mechanical pressure on cancer tissue. *Mol Cancer Res*. 2012;10(11):1403–1418.
12. Monteran L, Erez N. The dark side of fibroblasts: cancer-associated fibroblasts as mediators of immunosuppression in the tumor microenvironment. *Front Immunol*. 2019;10:1835.
13. Chen WJ, et al. Cancer-associated fibroblasts regulate the plasticity of lung cancer stemness via paracrine signalling. *Nat Commun*. 2014;5:3472.
14. Malanchi I, et al. Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature*. 2011;481(7379):85–89.
15. Saforo D, et al. Primary lung cancer samples cultured under microenvironment-mimetic conditions enrich for mesenchymal stem-like cells that promote metastasis. *Sci Rep*. 2019;9(1):4177.
16. Gascard P, Tlsty TD. Carcinoma-associated fibroblasts: orchestrating the composition of malignancy. *Genes Dev*. 2016;30(9):1002–1019.
17. McAllister SS, Weinberg RA. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. *Nat Cell Biol*. 2014;16(8):717–727.
18. Ohlund D, et al. Fibroblast heterogeneity in the cancer wound. *J Exp Med*. 2014;211(8):1503–1523.
19. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013;501(7467):346–354.
20. Gonda TA, et al. Molecular biology of cancer-associated fibroblasts: can these cells be targeted in anti-cancer therapy? *Semin Cell Dev Biol*. 2010;21(1):2–10.

21. Liu T, et al. Cancer-associated fibroblasts: an emerging target of anti-cancer immunotherapy. *J Hematol Oncol*. 2019;12(1):86.

22. Ma XJ, et al. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res*. 2009;11(1):R7.

23. Berdiel-Acer M, et al. Differences between CAFs and their paired NCF from adjacent colonic mucosa reveal functional heterogeneity of CAFs, providing prognostic information. *Mol Oncol*. 2014;8(7):1290–1305.

24. Orr B, et al. Identification of stromally expressed molecules in the prostate by tag-profiling of cancer-associated fibroblasts, normal fibroblasts and fetal prostate. *Oncogene*. 2012;31(9):1130–1142.

25. Herrera M, et al. Functional heterogeneity of cancer-associated fibroblasts from human colon tumors shows specific prognostic gene expression signature. *Clin Cancer Res*. 2013;19(21):5914–5926.

26. Navab R, et al. Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. *Proc Natl Acad Sci U S A*. 2011;108(17):7160–7165.

27. Finak G, et al. Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med*. 2008;14(5):518–527.

28. Hosein AN, et al. Breast carcinoma-associated fibroblasts rarely contain p53 mutations or chromosomal aberrations. *Cancer Res*. 2010;70(14):5770–5777.

29. Qiu W, et al. No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas. *Nat Genet*. 2008;40(5):650–655.

30. Brock MV, et al. DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med*. 2008;358(11):1118–1128.

31. Sandoval J, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J Clin Oncol*. 2013;31(32):4140–4147.

32. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012;13(7):484–492.

33. Baylin SB, Jones PA. A decade of exploring the cancer epigenome — biological and translational implications. *Nat Rev Cancer*. 2011;11(10):726–734.

34. Jones PA, et al. Targeting the cancer epigenome for therapy. *Nat Rev Genet*. 2016;17(10):630–641.

35. Hu M, et al. Distinct epigenetic changes in the stromal cells of breast cancers. *Nat Genet*. 2005;37(8):899–905.

36. Pidsley R, et al. Enduring epigenetic landmarks define the cancer microenvironment. *Genome Res*. 2018;28(5):625–638.

37. Vizoso M, et al. Aberrant DNA methylation in non-small cell lung cancer-associated fibroblasts. *Carcinogenesis*. 2015;36(12):1453–1463.

38. Mishra R, et al. Epigenetic changes in fibroblasts drive cancer metabolism and differentiation. *Endocr Relat Cancer*. 2019;26(12):R673–R688.

39. Ikemori R, et al. Epigenetic *SMAD3* repression in tumor-associated fibroblasts impairs fibrosis and response to the antifibrotic drug nintedanib in lung squamous cell carcinoma. *Cancer Res*. 2020;80(2):276–290.

40. Maeda M, et al. Cancer cell niche factors secreted from cancer-associated fibroblast by loss of H3K27me3. *Gut*. 2020;69(2):243–251.

41. Nurmik M, et al. In search of definitions: cancer-associated fibroblasts and their markers. *Int J Cancer*. 2020;146(4):895–905.

42. Sahai E, et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer*. 2020;20(3):174–186.

43. Dotto GP. Multifocal epithelial tumors and field cancerization: stroma as a primary determinant. *J Clin Invest*. 2014;124(4):1446–1453.

44. Sylvestre M, et al. Epigenetic mechanisms driving tumor supportive microenvironment differentiation and function: a role in cancer therapy? *Epigenomics*. 2020;12(2):157–169.

45. Fasanelli F, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat Commun*. 2015;6:10192.

46. Zhang Y, et al. Smoking-associated DNA methylation markers predict lung cancer incidence. *Clin Epigenetics*. 2016;8:127.

47. Ambatipudi S, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics*. 2016;8(5):599–618.

48. Kadara H, et al. Pulmonary adenocarcinoma: a renewed entity in 2011. *Respirology*. 2012;17(1):50–65.

49. Teschendorff AE, et al. Correlation of smoking-associated DNA methylation changes in buccal cells with DNA methylation changes in epithelial cancer. *JAMA Oncol*. 2015;1(4):476–485.

50. Joehanes R, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 2016;9(5):436–447.

51. Peters A, et al. Podoplanin negatively regulates CD4$^+$ effector T cell responses. *J Clin Invest*. 2015;125(1):129–140.

52. Lan Q, et al. CCL26 Participates in the PRL-3-induced promotion of colorectal cancer invasion by stimulating tumor-associated macrophage infiltration. *Mol Cancer Ther*. 2018;17(1):276–289.

53. Lambrechts D, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med*. 2018;24(8):1277–1289.

54. Barrett RL, Pure' E. Cancer-associated fibroblasts and their influence on tumor immunity and immunotherapy. *Elife*. 2020;9:e57243.

55. Biffi G, Tuveson DA. Diversity and biology of cancer-associated fibroblasts. *Physiol Rev*. 2021;101(1):147–176.

56. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17(6):1471–1474.

57. Gautier L, et al. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;20(3):307–315.

58. Tibshirani R, et al. samr: SAM: Significance Analysis Of Microarrays. http://statweb.stanford.edu/~tibs/SAM/sam.pdf. Accessed July 6, 2021.

59. Huang DW, et al. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.

60. Kanehisa M, et al. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. 2010;38(Database issue):D355–D360.

61. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.

62. Davis S, et al. methylumi: Handle Illumina methylation data. https://www.bioconductor.org/packages/release/bioc/html/methylumi.html. Accessed July 6, 2021.

63. Heinze G, Ploner M. coxphf: Cox regression with Firth's Penalized Likelihood. https://search.r-project.org/CRAN/refmans/coxphf/html/coxphf.html. Accessed July 6, 2021.

64. Conway JR, et al. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 2017;33(18):2938–2940.

65. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511):543–550.

66. Liu J, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018;173(2):400–416.