

RESEARCH

Open Access



Extracellular matrix profiles determine risk and prognosis of the squamous cell carcinoma subtype of non-small cell lung carcinoma

Amelia L. Parker^{1,2*} , Elise Bowman³, Adriana Zingone³, Brid M. Ryan^{3,4}, Wendy A. Cooper^{5,6,7}, Maija Kohonen-Corish^{8,9,10}, Curtis C. Harris³ and Thomas R. Cox^{1,2*}

Abstract

Background: Squamous cell carcinoma (SqCC) is a subtype of non-small cell lung cancer for which patient prognosis remains poor. The extracellular matrix (ECM) is critical in regulating cell behavior; however, its importance in tumor aggressiveness remains to be comprehensively characterized.

Methods: Multi-omics data of SqCC human tumor specimens was combined to characterize ECM features associated with initiation and recurrence. Penalized logistic regression was used to define a matrix risk signature for SqCC tumors and its performance across a panel of tumor types and in SqCC premalignant lesions was evaluated. Consensus clustering was used to define prognostic matretypes for SqCC tumors. Matretype-specific tumor biology was defined by integration of bulk RNAseq with scRNAseq data, cell type deconvolution, analysis of ligand-receptor interactions and enriched biological pathways, and through cross comparison of matretype expression profiles with aging and idiopathic pulmonary fibrosis lung profiles.

Results: This analysis revealed subtype-specific ECM signatures associated with tumor initiation that were predictive of premalignant progression. We identified an ECM-enriched tumor subtype associated with the poorest prognosis. In silico analysis indicates that matrix remodeling programs differentially activate intracellular signaling in tumor and stromal cells to reinforce matrix remodeling associated with resistance and progression. The matrix subtype with the poorest prognosis resembles ECM remodeling in idiopathic pulmonary fibrosis and may represent a field of cancerization associated with elevated cancer risk.

Conclusions: Collectively, this analysis defines matrix-driven features of poor prognosis to inform precision medicine prevention and treatment strategies towards improving SqCC patient outcome.

Keywords: Extracellular matrix, Non-small cell lung cancer, Squamous

Background

The squamous subtype of non-small cell lung cancer (NSCLC) is the second most common lung cancer subtype [1, 2], contributing to a lung cancer survival rate of only 20.2% [3]. The lack of actionable mutations in this cancer type has hindered the development of precision medicine protocols that have substantially improved patient survival in subgroups of adenocarcinoma patients [4]. Platinum doublet therapy remains the first-line

*Correspondence: Am.parker@garvan.org.au; t.cox@garvan.org.au

¹ Matrix and Metastasis Lab, Cancer Ecosystems Program, Garvan Institute of Medical Research and The Kinghorn Cancer Centre, 384 Victoria St, Darlinghurst, NSW 2052, Australia

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

treatment for the majority of SqCC patients; however, it is not yet possible to predict which patients will respond. In early-stage patients, where surgery is the treatment of choice and potentially curable, up to 30% will have their disease recur [5].

Worldwide, lung cancer screening programs have demonstrated substantial improvements in patient mortality through early detection and intervention [6]. The increased detection of early, pre-invasive lesions also raises clinical challenges as 30% of pre-invasive lesions will spontaneously regress in the absence of intervention [7, 8]. There is currently no way of identifying which premalignant lesions will become invasive and progress to cancer. Robust identification of individuals at high risk of premalignant progression will be an important keystone for the successful implementation of screening strategies for lung cancer patients globally.

The high degree of subclonal tumor heterogeneity [9] and high mutational burden [10] in SqCC tumors make it difficult to achieve sustained clinical responses when treating cancer cells alone. This has been illustrated in recent precision medicine trials targeting specific cancer cell mutations in advanced stage patients who were ineligible for surgery or radiotherapy or those who had previously failed platinum doublet therapy, where only a small proportion of patients demonstrated clinical responses [11, 12]. The tumor microenvironment is increasingly acknowledged to play an important role in regulating the clonal evolution of the tumor, and the sensitivity of cancer cells to both chemotherapy and immunotherapies. Evidence in other cancer types indicates that the cell extrinsic factors from the tumor microenvironment have the potential to override cancer cell intrinsic signaling to regulate tumor progression independently of the clonal heterogeneity of the tumor [13]. Complementary treatment approaches that target the tumor-supporting components of the local microenvironment have the potential to simultaneously target multiple cancer cell clones to achieve more durable clinical responses in these highly heterogeneous tumors.

Fundamental to the success of a co-targeting approach is the ability to identify prognostic features of SqCC tumors and exploit these vulnerabilities for therapeutic benefit. A number of approaches have been employed to develop prognostic signatures in lung cancer, with the majority being developed in adenocarcinoma subtypes only [14–16], or of mixed NSCLC subtypes including both adenocarcinoma and squamous NSCLC [17–20]. Those developed for SqCC specifically have been developed by genome wide profiling and show little overlap [21–23]. Molecular subtyping of SqCC tumors has revealed distinct and heterogeneous SqCC tumor types at the transcript and protein level that are characterized

by varying degrees of immunological activation, proliferation, and differentiation [24, 25]. Importantly, these appear to have a distinct biology from lung adenocarcinomas [26–28], suggesting that prognostic programs operate differently between these major NSCLC subtypes. Improvements in SqCC treatment will require a thorough understanding of how SqCC biology differs from adenocarcinoma, and identification of therapies that may be more effective in this specific subtype.

The extracellular matrix (ECM) is an important regulator of cell behavior [29, 30] and is increasingly recognized to regulate tumor progression in multiple cancer types [31]. The ECM and its remodeling by cancer, stromal and immune cells have been shown to promote the growth, survival, and metastasis of cancer cells [32–34]. Associations between ECM expression and prognosis have been identified in NSCLC generally [17] and the expression and prognostic value of specific ECM proteins, such as tenascin-C [14], have been more extensively studied in adenocarcinoma than squamous NSCLC [35]. In SqCC tumors, for example, higher expression of the glycoprotein periostin was observed compared with adenocarcinomas, where it is expressed by activated fibroblasts [36] and is associated with survival [36, 37]. In addition, the glycoproteins thrombospondin 1 and 2 are overexpressed in SqCC [38]. However, it remains unclear how the SqCC ECM environment differs from that of the normal lung, and if ECM remodeling can inform diagnostic and therapeutic strategies in this poor-prognosis cancer. In addition, matrix molecules do not operate in isolation, but rather as integrated macromolecular networks [31]. A broad perspective of how these matrix networks are modulated in SqCC tumors, and the impact of this on cancer cell behavior is needed. The approval of the anti-stromal therapy nintedanib as a second-line therapy in adenocarcinoma patients [39] and clinical responses in Phase I trials in SqCC patients [40] attests to the potential for a co-targeting approach in the treatment of the lung cancer more generally. A more extensive examination of the ECM, and the role it plays in SqCC would be fundamental to the effective implementation of matrix co-targeting approaches specifically in this NSCLC subtype.

Here we present an unbiased examination of bulk and single-cell transcriptomic and proteomic analysis to examine the ECM landscape in SqCC, and to define matrix-associated signatures of risk and prognosis. Overlapping ECM remodeling with aging and chronic lung diseases suggests that early-stage changes in the lung ECM may promote tumor initiation and contribute to the increased incidence of lung tumors with age. Prognostic ECM subtypes (matreotypes [41]) identified in this study suggest that a subset of poor-prognosis SqCC patients

may benefit from the repurposing of existing stromal therapies as part of a co-targeting approach.

Methods

Clinical samples

Publicly available data was obtained as described (Table 1) and analyzed in accordance with the respective guidelines of each platform. TCGA data (RNA sequencing, whole exome sequencing, copy number variation, and reverse phase protein array data) from 223 SqCC patients (including 17 patients with matched non-involved non-tumor tissue) and 162 adenocarcinoma patients (including 37 patients with matched non-involved non-tumor lung tissue) were obtained as described in Table 1 for assessment of ECM gene expression, associated tumor biology, and patient outcome. SqCC lung tumor and non-tumor biospecimens from the NCI-MD cohort (30 patients from the greater Baltimore area with stage I–III tumors) were collected and processed for RNA sequencing and analyzed for associations with patient outcome according to approved procedures as described previously [52, 53] (National Cancer Institute, USA, IRB OH98-C-N027). SqCC tissue microarrays were generated from primary lung SqCC samples obtained from 94 patients with stage I–III tumors resected at Royal Prince Alfred Hospital, Sydney, NSW, Australia, between 1996 and 2002 as described previously [54]. Briefly, each core was 1mm in diameter and consisted of 3–6 tumor core replicates. These tissue microarrays were stained with picosirius red, imaged using polarized microscopy and analyzed for associations with

patient survival according to approved protocols (Royal Prince Alfred Hospital and Garvan Institute of Medical Research, HREC10/RPAH/491, X14-0359, Garvan GHRP 1423). SqCC scRNAseq data was obtained from Lambrechts et al. (see Table 1) derived from 3 patients who had undergone lobe resection as previously described [46]. The clinicodemographic features of these cohorts are outlined in Supplementary Tables 4, 5 and 6. Written informed consent was received prior to participation in these studies.

RNA sequencing

Lung tumor and adjacent non-tumor tissue samples were collected as part of the NCI- University of Maryland Study according to protocols approved by the institutional review board (OH98-C-N027, National Cancer Institute, USA) and processed as described previously [53, 55] and outlined in detail in [Supplementary Methods](#). Gene expression levels were quantified using RSEM [56] and batch corrected (Combat algorithm, SVA package [57]). Additional details are provided in [Supplementary Methods](#).

Bulk RNAseq analysis of squamous cell carcinomas

RSEM-quantified RNAseq abundances and relevant clinical information from publicly available datasets were accessed according to Table 1. Gene level counts were filtered and TMM normalized (EdgeR package [58]) before being \log_2 transformed.

Matrisomal genes were defined as core matrisome or matrisome-associated according to Naba et al. [29].

Table 1 Datasets accessed in this study

Datasets			
Cohort	Data type	Source	Reference
TCGA LUSC	RNAseq, WES, CNV, RPPA (Level 4 data)	GDAC Firehose, TCPA Portal	[42, 43]
TCGA LUAD	RNAseq	GDAC Firehose	[44]
NCI-MD Cohort SqCC	RNAseq	This study; Gene Expression Omnibus, GSE201221	
UHN Cohort	Microarray	Gene Expression Omnibus, GSE50081	[20]
TCGA Pan-Cancer Cohort	RNAseq	https://gdc.cancer.gov/about-data/publications/pancanatlas with survival data from [45]	[45]
SqCC cell types from scRNAseq	scRNAseq	Loom files and Signature Matrix	[46]
Premalignant lesions	RNAseq	Gene Expression Omnibus GSE108124	[7]
Aging lung	RNAseq	Gene Expression Omnibus GSE165192	[47]
Data and signatures			
Data type	Data format	Source	Reference
IPF fibrosis score	Expression Data	Manuscript	[48]
IPF fibroblast phenotypes from scRNAseq	scRNAseq	Gene Expression Omnibus	[49]
IPF cell types from scRNAseq	Signature matrix	Manuscript	[50]
Lung cancer-specific immunological cell types	Gene Signatures	Manuscript	[51]

Principal component analysis of tumor compared with non-tumor tissue was performed using the `prcomp` algorithm and default settings in base R. Differential gene expression analysis between tumor and non-tumor tissue was performed using the `Limma` package [59]. Heatmaps of gene expression matrices were generated using the `ComplexHeatmap` package [60] for visualization. Additional details are provided in [Supplementary Methods](#).

Proteomic analysis of squamous cell carcinomas

The proteome of TCGA SqCC samples were analyzed using the Level 4 Reverse Phase Proteomic Array (RPPA) data downloaded from the TCPA portal (<https://tcpaportal.org/tcpa/download.html>) [42]. Differential enrichment of proteomic data between ECM-High and ECM-Low Matreotypes was performed using Mann-Whitney U tests with Benjamini-Hochberg correction for multiple comparisons.

Tumor purity analysis

To assess the tumor purity and relative stromal and immune composition of bulk RNAseq data from the TCGA LUSC dataset, ESTIMATE scores of tumor purity, including stromal and immune admixture scores [61] for each sample were accessed from the MD Anderson Bioinformatics server (<https://bioinformatics.mdanderson.org/estimate/disease.html>) as the LUSC samples for RNAseq2 data.

Squamous cell carcinomas canonical molecular subtypes

Canonical SqCC molecular subtypes (Basal, Classical, Primitive, Secretory) defined by Wilkerson et al. [24] were assigned to TCGA LUSC bulk RNAseq samples as the canonical molecular subtype with the highest Pearson correlation coefficient using the published centroids [24]. Samples were classified as primitive (9.4%), classical (42.1%), secretory (21.5%), and basal (26.9%), a distribution that is consistent with previous reports [24, 43]. Enrichment of canonical molecular subtypes within ECM-High and ECM-Low matreotypes was assessed by Fisher's exact test.

Matrix risk signature generation

Risk signature feature selection was performed on differentially expressed core matrix genes between tumor and non-tumor tissue in the TCGA LUSC dataset. Genes were selected based on their association with tumor compared with non-tumor tissue using Elastic Net penalized logistic regression (`glmfit` algorithm, `glmnet` package with $\alpha = 0.5$ [62]), to account for the high degree of correlation of some matrix genes. Z -scaled RNAseq expression data was initially partitioned 80% to 20% into training and test datasets, respectively

(`createDataPartition` algorithm, `caret` package) and the model was developed using training data only. The coefficient of shrinkage (λ) was chosen as the value within one standard error of the λ that minimized the cross-validation prediction error rate (`cv.glmnet` algorithm, `glmnet` package). To generate a matrix risk score and for visualization purposes, odds ratios were calculated using Firth's correction (`logistf` algorithm, `logistf` package). The matrix risk score for each sample was summed as the product of the $\log(\text{odds ratio})$ (Supplementary Table 1) and the expression value for each z -scaled gene expression (Equation 1). Additional details are provided in [Supplementary Methods](#).

$$\text{Matrix Risk Score} = \sum_{i=1}^n z_i \beta_i \quad (1)$$

where i = gene in the matrix risk signature of length n
 z_i = z -scaled gene expression of gene i
 β_i = $\log(\text{odds ratio})$ of gene i

Matreotype identification

Identification of matreotypes was performed using Monte-Carlo reference-based consensus clustering (M3C command using the K-means clustering algorithm with default settings, M3C package [63]) applied to the TCGA expression matrix of significantly differentially expressed core matrix genes from tumor and non-tumor tissue. Cluster number was determined to give a maximal Relative Cluster Stability Index (RCSI), a Monte-Carlo p -value less than 0.05 and a minimal Proportional of Ambiguous Clustering (PAC) Score [63].

Centroids for each matreotype were calculated on the TCGA LUSC dataset as the mean z -scaled expression level for significantly differentially expressed core matrix genes between tumor and non-tumor tissue. Matreotypes were assigned as the minimum Euclidean distance between each sample and the matreotype centroids. The association of matreotypes with categorical clinicodemographic information (Fisher's exact test) and survival (survival and `survminer` packages) was assessed. Hazard ratios were calculated (`coxph` function, `survminer` package) with corrections for age and stage, which were clinical covariates significantly associated with outcome in univariate analyses. Gender, smoking status, and pack years were not statistically significantly associated with outcome in univariate analyses.

Matreotype associations with driver mutations were assessed using `maftools`. FGFR copy number variation analysis was performed on TCGA CNV data using probes mapped to the segment containing the FGFR cluster at chromosome 8p11.23 (Hg18; 38 387 813-38 445 509) with segment means < -0.7 and > 0.7 defining copy number losses and gains, respectively. The association

of FGFR amplifications with ECM-High and ECM-Low matretypes was assessed by Fisher's exact test. Additional details are provided in [Supplementary Methods](#).

Pathway analyses

Pathway analysis was performed on the 50 hallmark pathways and C2 oncogenic pathways described in the molecular signatures database (mSigDb package (<https://davislaboratory.github.io/msigdb>)). Pathway activity estimates were applied to each sample (gsva algorithm with default settings, GSVA package [64]) and differential pathway enrichment was assessed using Limma [59] as described in detail in the [Supplementary Methods](#).

Ligand-receptor interaction analysis

Genes within the RNAseq datasets were annotated as ligands and receptors based on the curated database of human ligand-receptor pairs previously published by Ramilowski et al. [65] based on supporting literature. Only ligands corresponding to core matrisome genes as defined by Naba et al. [29] were retained for further analysis. The strength of the interaction between core matrisomal genes and their receptors were calculated as the product of expression values from the ligand (i.e., core matrisomal gene) and its cognate receptor in each sample (Equation 2), as described previously [66].

$$\text{Interaction Score} = R_i L_i \quad (2)$$

where R_i = z-scaled gene expression of receptor i

L_i = z-scaled gene expression of ligand i

Receptors were then grouped into receptor classes, and interaction scores for each receptor class were calculated as the maximum interaction score for that receptor class. Receptors were also mapped to their corresponding mSigDb Hallmark pathways.

Differential enrichment of ligand-receptor interaction scores or ligand-pathway interaction scores in the different matretypes were tested using Limma (limma package) [59] and visualized as a circos plot (circos algorithm, circlize package [67]). Additional detail is provided in [Supplementary Methods](#).

Cellular composition analysis

Two main approaches were implemented to identify enriched cell types in SqCC matretypes using datasets described in Table 1. Deconvolution of bulk TCGA LUSC RNAseq data into NSCLC epithelial, fibroblast, and endothelial cell types was performed using the signature matrix derived from scRNAseq analysis of NSCLC tumors published by Lambrechts et al. [46] and CibersortX (absolute mode, batch correction, 100

permutations). Differential enrichment of cell types was tested using the limma package [59].

The relative enrichment of immune cells within the tumor microenvironment was assigned to each sample by applying NSCLC-specific immune cell signatures developed by Faruki et al. [51] using Gene Set Variation Analysis (gsva algorithm, GSVA package, default settings). Differential enrichment of immune cell type scores in the ECM-High vs ECM-Low matretypes were assessed by Kruskal-Wallis test with Benjamini-Hochberg multiple comparisons correction. Additional detail is provided in [Supplementary Methods](#).

Single-cell RNAseq analysis of squamous cell carcinoma

Raw gene expression matrices and cellular metadata including cell type assignments from scRNAseq data of NSCLC samples were obtained and processed as described [46, 68] (Seurat v4.0.1; see [Supplementary Methods](#) for details). Cell type scores for ECM genes were assigned using the AddModuleScore function (Seurat package, default settings). Matrix risk scores were calculated on the z-scaled gene expression matrix for each cell type using Equation 1 as described above (Matrix Risk Signature section).

Fibrosis score

A fibrosis score was calculated based on the idiopathic pulmonary fibrosis signature described by McDonough et al. [48] (see [Supplementary Methods](#)). Briefly, normalized gene expression was weighted in the positive and negative direction for genes up- and downregulated in IPF lungs, respectively (Equation 3).

$$\text{IPF Fibrosis Score} = \sum_i^n w_i z_i \quad (3)$$

where $w_i = 1$ for gene upregulated in IPF lungs and -1 for genes downregulated in IPF lungs for gene i in the n gene signature

z_i = scaled gene expression for gene i in the n gene signature

Transcription factor enrichment analysis

Transcription factor analysis on core matrisomal correlation clusters was performed on gene lists corresponding to each correlation cluster using the chEA transcription factor targets database and chEA3 web interface [69] (see [Supplementary Methods](#) for details). The integrated mean rank for each transcription factor was calculated for each correlation cluster.

Picrosirius red staining and quantitation in tissue microarray

Picrosirius red staining of SqCC TMAs was performed as described previously [70]. Stained sections were imaged on a Leica DMI 6000 fitted with posterior and anterior polarizing filters and picrosirius red signal relative to tissue area was quantified using an in-house script published previously [70]. The maximum picrosirius red signal over three to five cores was calculated per patient. The association of picrosirius red signal with survival was performed using Cox proportional hazards models (coxph algorithm and survival package). See [Supplementary Methods](#) for additional detail.

Statistics

All statistical analysis was performed in R (v3.6.3) and GraphPad Prism (v8). *P*-values were adjusted for multiple comparisons using the Benjamini-Hochberg method. Statistical testing of two groups was performed by the non-parametric Mann-Whitney *U* test and of more than two groups by the Kruskal-Wallis test. *P*-values less than 0.05 were considered significant.

Data and code availability

Publicly available data used in this study are accessible as listed in Table 1. The NCI-MD cohort RNAseq data analyzed in this study are accessible at the NCBI GEO website under the accession number GSE201221 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE201221>) [71]. Publicly available TCGA LUSC and LUAD RNAseq, WES, and CNV data are available from Broad GDAC Firehose (<https://gdac.broadinstitute.org>). Publicly available Pan-Cancer RNAseq and survival data is available from Genomic Data Commons repository (<https://gdc.cancer.gov/about-data/publications/pancanatlas>).

Publicly available TCGA RPPA data is available from the TCPA Portal (<https://tcpaportal.org/tcpa/download.html>). Publicly available data from the UHN cohort is available in the Gene Expression Omnibus repository, GSE50081 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50081>). Publicly available RNAseq of squamous carcinoma in situ is available in the Gene Expression

Omnibus repository, GSE108124 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108124>). scRNAseq data are available from <https://lambrechtslab.sites.vib.be/en/data-access>. RNAseq data of the aging lung is available in the Gene Expression Omnibus repository, GSE165192 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165192>). Code used in this analysis is available via GitHub (<https://github.com/AParkerLab>). All other relevant data supporting the conclusions of this article are included within the article and its [supplementary information files](#). Additional information may be obtained from the corresponding authors upon reasonable request.

Results

Cellular heterogeneity contributes to tumor-specific ECM remodeling in squamous carcinoma

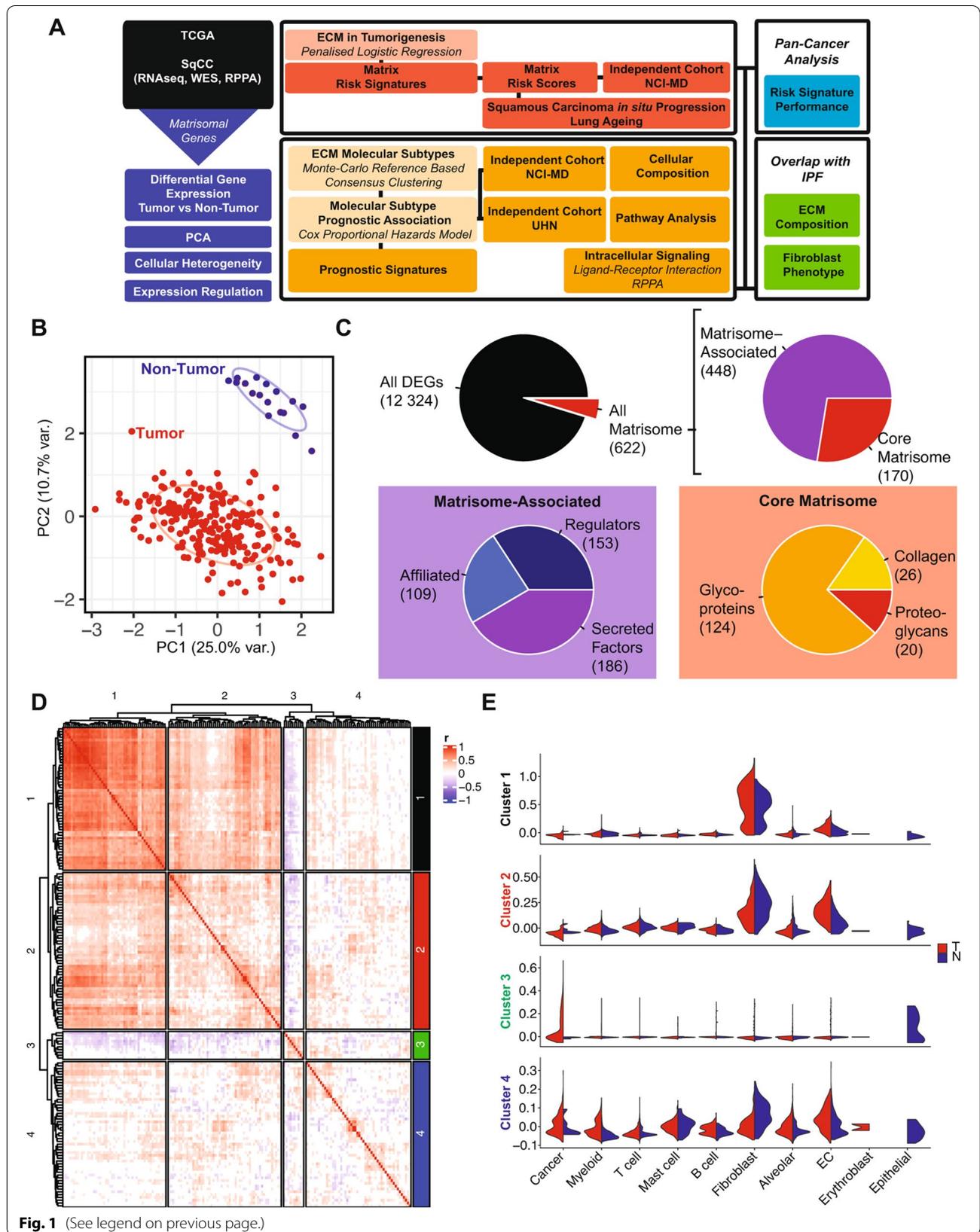
In order to comprehensively profile the ECM landscape in SqCC, we began by examining changes in the expression of ECM genes in tumor compared with non-tumor tissue using RNA sequencing data from The Cancer Genome Atlas [43] for which clinicodemographic and whole exome sequencing data was also available. Whole transcriptome RNA sequencing data from SqCC tumor ($n=223$) and matched, adjacent, non-involved non-tumor ($n=17$) tissues was filtered to include only matrisomal genes, as defined by Naba et al. [29] (Fig. 1A). These genes included 274 core matrisomal genes, which have clear structural roles in the ECM, as well as 753 matrisome-associated genes, which include enzymes involved in ECM remodeling and soluble factors that directly interact with the matrix [29].

Principal component analysis of core matrisomal gene expression indicated that the first two principal components of these genes were able to distinguish tumor from adjacent, non-involved lung tissue (Fig. 1B), suggesting that the expression of the core, structural components of the ECM in SqCC tumors substantially differs from that of normal lung.

Differential gene expression analysis of the global RNAseq data identified that 5.34% of all differentially expressed genes comparing tumor and non-tumor tissue were matrisomal genes (Fig. 1C). The majority of

(See figure on next page.)

Fig. 1 The ECM is significantly dysregulated in tumor compared with non-tumor tissue in SqCC. **A** Workflow describing the approach to characterizing the ECM landscape in SqCC. **B** Principal component analysis of core matrisomal gene expression in tumor (T) compared with non-tumor (NT) tissue. **C** Distribution of differentially expressed genes between tumor and non-tumor. 59.5% of matrisome-associated (488 genes) and 62.0% of core matrisomal (170) genes are differentially expressed in SqCC tumors. This represents 59.1% (26 genes), 57.1% (20 genes), and 63.6% (124 genes) of collagens, proteoglycans, and glycoproteins, respectively. It also represents 63.7% (109 genes), 64.3% (153 genes), and 54.1% (186 genes) of all ECM-affiliated proteins, ECM regulators, and secreted factors, respectively. Number of genes indicated in brackets. **D** Correlation analysis of differentially expressed core matrisomal genes identifies four major clusters of genes in SqCC tumors. r = Spearman correlation coefficient. Non-significant correlations are colored white; positively correlated genes are red and negatively correlated genes are blue. **E** Expression score of core matrisomal genes corresponding to the four correlational clusters shown in **D**, in specific lung cell types separated by tumor (T) and non-tumor (N) tissue. EC: endothelial cell



matrisomal genes were differentially expressed, with the expression of 62.0% of core and 59.5% of matrisome-associated genes differing between tumor and non-tumor tissue (Fig. 1C).

Comparative analysis with RNA sequencing data of 162 lung adenocarcinoma tissues (125 tumor, 37 non-tumor tissues) [44] revealed that 48.9% of all core matrisomal genes are differentially expressed between tumor and non-tumor tissues in both histological subtypes of NSCLC (Additional file 1: Fig. S1A). The majority of these genes were similarly regulated in adenocarcinoma tumors compared with matched non-tumor tissue suggesting that the dysregulation of many core matrisomal genes may be important in lung tumorigenesis in a subtype-independent manner (Additional file 1: Fig. S1B). However, the opposite expression of the basement membrane collagen *COL4A6*, basement membrane netrin *NTN5*, *VWA5B2*, and collagen fibrillogenesis regulator *AEBP1* in SqCC compared with adenocarcinoma (Additional file 1: Fig. S1B) points to histological-subtype-specific changes in the core matrisome that may be important in tumor development. Further examination of collagen alpha-6 type IV (*COL4A6*) expression in tumor tissue for the two subtypes indicated that *COL4A6* expression did not significantly correlate with the expression of the adenocarcinoma marker *TTI* in adenocarcinoma (Additional file 1: Fig. S1C) but was significantly correlated with the SqCC marker *TP63* (Additional file 1: Fig. S1D), suggesting it is upregulated during squamous differentiation associated with tumor development in a cell-type-specific manner [72].

In recent years, it has become clear that single matrix molecules rarely function in isolation and instead function as part of a dynamic 3D supramolecular network of structurally and functionally integrated matrix components [31]. To understand whether core matrisomal genes were coordinately regulated in SqCC tumors, we performed correlation analysis on all differentially expressed core matrisomal genes in tumor tissues (Fig. 1D). Unsupervised hierarchical clustering of this analysis identified 4 major correlated core matrisomal clusters (Fig. 1D), broadly corresponding to a fibrotic ECM remodeling cluster (correlation cluster 1; including fibrillar collagens *COL1A1*, *COL3A1*, *COL5A1*, and pro-fibrotic glycoproteins *THBS1* and *THBS2*); a basement membrane-enriched cluster (correlation cluster 2; including collagen type 4 (*COL4A1*, *COL4A2*, *COL4A3*, *COL4A4*, and basement membrane proteoglycan *HSPG2*); a glycoprotein-enriched cluster (correlation cluster 3; including *BGLAP*, *ZP3*, *FNDC8*, and *EMILIN3*); and with the fourth cluster consisting of genes which were not significantly correlated with one another (correlation cluster 4, including

COL4A6, and *CRELD1* and *CRELD2* glycoproteins) (Fig. 1D). Correlated matrisomal gene expression profiles in bulk RNAseq may reflect coordinated regulation of gene expression within cells, differences in cell type composition among samples or combinations of both these scenarios. To understand in more detail how these correlated matrisomal genes may reflect the contributions of individual cells within the tumor microenvironment, we assessed the expression of these correlation clusters in individual cell types in tumor and non-tumor tissue from scRNAseq data of SqCC patients (3 patients with tumor and adjacent non-involved non-tumor tissue [46]). This analysis identified that the fibrotic correlation cluster genes (correlation cluster 1) were expressed largely by fibroblasts, while the basement membrane-enriched cluster genes (correlation cluster 2) also had additional contributions from endothelial cells (Fig. 1E, Additional file 1: Fig. S1E-H). The small group of glycoprotein-enriched genes (correlation cluster 3) were significantly expressed by epithelial cells in non-tumor tissue, and by cancer cells in tumor tissue (Fig. 1E, Additional file 1: Fig. S1I), indicating epithelial origin. Finally, the fourth cluster of genes (correlation cluster 4) was contributed to by multiple cell types within both the normal lung and tumor microenvironment (Fig. 1E, Additional file 1: Fig. S1J). Significant differences between tumor and non-tumor scores for fibroblasts, endothelial and epithelial cells in correlation clusters 1 to 3 (Fig. 1E) suggest that the presence of tumor cells likely induces altered matrisomal gene expression within stromal cells within the tumor microenvironment. Together, these analyses indicate that not only do cancer cells as well as resident stromal and immune cells within the SqCC tumor microenvironment contribute to significant ECM remodeling compared with normal lung tissue, but also that cellular heterogeneity within the tumor microenvironment contributes to, and likely underlies, the matrisomal profiles of SqCC tumors.

Co-regulation of gene transcription can also result in highly correlated gene expression networks. To examine this in more detail, we performed transcription factor enrichment analysis on the correlated gene clusters using chEA3 [69]. Because the transcriptional regulation of many ECM genes has not yet been comprehensively defined, integrating multiple CHIP databases with expression analysis in this approach expands the scope for identifying upstream regulators of matrix gene expression. This analysis identified enrichment of overlapping as well as distinct transcription factors for each cluster that have been implemented in lung development and cancer (Additional file 1: Fig. S1K). Consistent with the high correlation between genes in these clusters, correlation clusters 1 and 2 were both enriched with

targets of AEBP1 and OSR1, which are expressed during fetal lung development [73] (Additional file 1: Fig. S1K). Conversely, correlation cluster 1 was selectively enriched for the EMT transcription factor TWIST2 while correlation cluster 2 was specifically enriched for FOXC2, which has been associated with mesenchymal development, and is known to be associated with platinum resistance and poor prognosis [74, 75]. Cluster 3 was enriched for LHX8 and FEV as well as DBX2, which is differentially expressed in hepatocellular carcinoma [76]. Finally, cluster 4 showed enrichment for SP7 and OLIG1, which has been previously shown to have prognostic value in NSCLC [77], as well as TP63, the master regulator of squamous differentiation [43] that correlates with *COL4A6* expression in this cluster (Additional file 1: Fig. S1D). While further experimental validation of these findings is required, these data highlight that common regulatory networks have the potential to co-regulate the expression of several matrix components within the tumor microenvironment. This, together with cell-type-specific matrix deposition from cancer cells, as well as normal and co-opted stromal and immune cells from the tumor microenvironment, is likely to contribute to dysregulated matrisomal gene expression in SqCC.

ECM remodeling predicts squamous lung cancer risk

To more precisely define those ECM components that distinguish between tumor and non-tumor tissue in an unbiased manner, we further refined our list of differentially expressed matrisomal genes between tumor and non-tumor SqCC tissue using Elastic Net penalized logistic regression. This ensures that only the most significant genes were retained in the model while accommodating the high degree of correlation observed for some of the ECM genes (Fig. 1D). The final matrix risk model was then corrected for clinical covariates including age, sex, tumor stage, smoking status, and pack years [78] to minimize cohort bias.

This analysis resulted in a 28-gene signature, with the microfibril protein microfibril-associated protein 4 (MFAP4) and the squamous-associated anchoring fibril collagen alpha1 type VII (COL7A1) having the lowest and highest odds ratios, respectively (Fig. 2A, Additional file 1: Fig. S2A and Table S1). In support of

the feature selection approach, our signature includes ECM components identified as differentially regulated in non-small cell lung cancer, including osteopontin (SPP1), which has been associated with lung adenocarcinoma prognosis [79–83], but not yet specifically SqCC. Similarly, COL11A1 has also been associated with patient survival in NSCLC [84] and other cancers [85, 86]. Conversely, COL7A1 is known to play an important role in squamous tumorigenesis in the skin and esophagus [87, 88], suggesting that our signature captures SqCC-specific ECM remodeling programs.

Collapsing our matrix risk signature into a gene-weighted matrix risk score confirmed that this specific matrix risk score is significantly higher in tumor compared with non-tumor tissue in a test dataset partitioned from the TCGA LUSC dataset (Fig. 2B) as well as an independent SqCC dataset from the NCI-MD cohort (Fig. 2C) and distinguished tumor from non-tumor tissue by ROC analysis in these cohorts (Fig. 2D, E, Additional file 1: Fig. S2B). Analysis of each gene identified substantial predictive power for each individual gene within this risk signature, and further model refinement highlighted that a minimal signature of TNXB, COL7A1, and SPP1 achieved comparable predictive power in the NCI-MD cohort as measured by ROC curve analysis (Additional file 1: Fig. S2C), suggesting the key importance of these matrix components in tumor microenvironment ECM remodeling.

Comparatively, an adenocarcinoma-specific matrix signature developed on TCGA samples in parallel [44] shared 15 genes in common with the squamous matrix risk signature (Additional file 1: Table S1). This indicates that our matrix risk signature captures squamous-specific ECM components (e.g., COL7A1), as well as ECM components that associate with lung tumor development independently of histological subtype (e.g., EMLIN2, SPP1, COL10A1, COL11A1, and CTHRC1). To examine if our SqCC matrix risk signature reflected ECM changes in the lung specifically, or was more broadly applicable to other cancer types, we also examined its performance in distinguishing tumor from non-tumor tissue among solid tumor types in the TCGA Pan-Cancer cohort [45]. Our matrix risk score was significantly elevated in tumor compared with non-tumor tissue in all cancer types

(See figure on next page.)

Fig. 2 ECM changes associated with increased lung cancer risk and premalignant progression. **A** Odds ratios for genes in the matrix risk signature. See Additional File 1: Table S1. **B,C** Matrix risk scores are significantly higher in tumor compared with non-tumor tissue in the TCGA (**B**, $p=6.7E-12$, Mann-Whitney U test) and NCI-MD cohorts (**C**, $p=4.6E-10$, Mann-Whitney U test). **D,E** ROC analysis of the matrix risk score in distinguishing between tumor and non-tumor tissue on the bootstrapped test data from the TCGA cohort (**D**, area under the curve (AUC)=1) and the NCI-MD cohort (**E**, area under the curve (AUC) = 0.89). **F,G** Expression scores for matrix risk signature genes with positive odds ratios (**F**) and negative odds ratios (**G**) in specific lung cell types from tumor and non-tumor tissue. **H** Matrix risk score in regressive (Reg.) and progressive (Prog.) pre-invasive squamous carcinoma in situ lesions. $p=0.023$, Mann-Whitney U test. **I** Matrix Risk Score in young and old lungs in the absence of tumors, $p=0.0020$, two-sided Student's t test

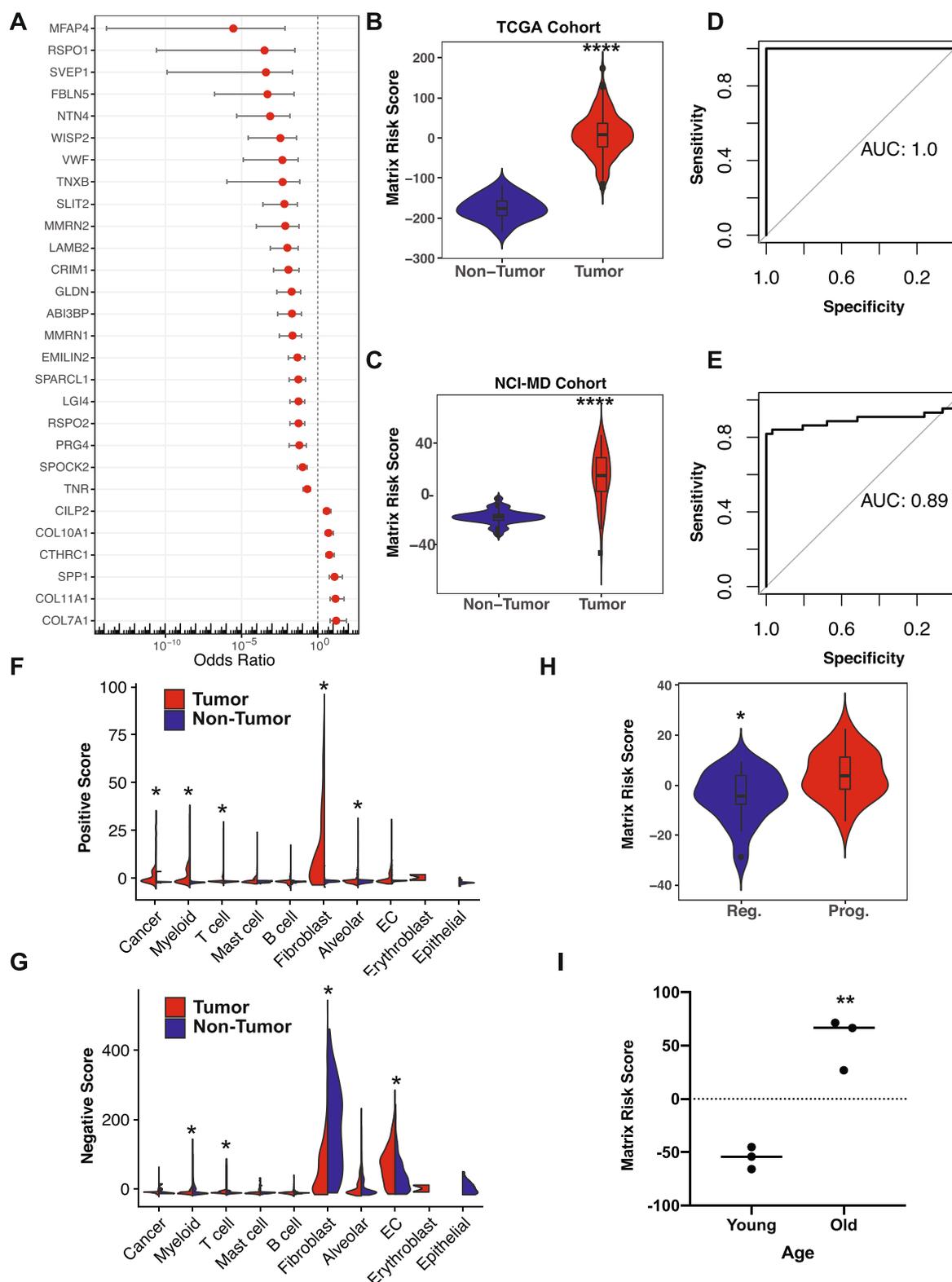


Fig. 2 (See legend on previous page.)

tested including lung adenocarcinoma (LUAD), except for pancreatic adenocarcinoma (PAAD) and sarcoma (SARC) (Additional file 1: Fig. S2D), and demonstrated predictive power (ROC analysis AUC>0.75) in multiple cancer types (Additional file 1: Fig. S2E and Table S2). Interestingly, the predictive value of our matrix risk signature was highest in the cervical SqCC and endocervical adenocarcinoma cohort, which has a large proportion of tumors with squamous characteristics [89]. This suggests that the matrix gene expression changes identified by our matrix risk signature may be more widely applicable to other cancer types, in particular those of squamous type, and warrants further investigation in these contexts.

To investigate which cell types express these key ECM genes, we calculated the enrichment of matrix risk genes with positive odds ratios (positive score) and negative odds ratios (negative score) in specific cell types identified by scRNAseq analysis of SqCC tumor and matched non-tumor tissue [46]. This analysis identified that core matrix genes that are positively associated with SqCC tumors (Additional file 1: Table S1; OR >1) are principally expressed by cancer-associated fibroblasts (CAFs) as well as cancer and myeloid cells (Fig. 2F, Additional file 1: Fig. S3A, B). In contrast, matrix genes that are negatively associated with SqCC tumors (Additional file 1: Table S1; OR <1) are expressed by alveolar and endothelial cells as well as a non-tumor-associated fibroblast population (Fig. 2G, Additional file 1: Fig. S3C). The contribution of these distinct fibroblast populations (cancer vs non-tumor-associated fibroblasts) to the ECM profile of SqCC tumors further supports the association of these matrix components with SqCC risk and may reflect phenotypic transdifferentiation of resident fibroblasts into CAFs during tumorigenesis [46]. Overall, these data suggest that changes in stromal cell gene expression within the tumor microenvironment, as well as matrix expression derived from cancer cells themselves, contribute to this risk score.

The early detection of pre-invasive SqCC in situ (CIS) lesions presents an opportunity to prevent the onset of lung cancer yet remains clinically challenging. While standard pathological assessment is able to distinguish between tumor and non-tumor tissue, there is currently no way of predicting which CIS lesions will progress to cancer, and which are likely to regress. This presents significant clinical challenges in deciding on the appropriate interventions while avoiding overtreatment [7]. Applying our matrix risk score to a recent dataset of pre-invasive CIS lesions [7] revealed that our matrix risk score was significantly higher in CIS lesions that progress to cancer compared to those that spontaneously regress (Fig. 2H). Further refinement of this matrix risk signature to

specifically predict which lesions will progress to cancer identified a minimum 6-gene risk signature of *RSPO1*, *CTHRC1*, *SPP1*, *MMRN1*, *COL10A1*, and *PRG4* as the most significant matrix predictors of premalignant progression with a ROC AUC of 0.99 (Additional file 1: Table S3, Fig. S3D). This suggests that not only are matrix changes occurring and detectable at the earliest pre-invasive stage of tumor development, but also that the ECM profile of CIS lesions has the potential to predict a high risk of tumor development.

Like many solid tumors, lung cancer incidence increases with age [78, 90, 91] and ECM of the lung continues to remodel with age [47, 92]. Age-related ECM remodeling is emerging as a significant contributor to the increased age-related cancer risk in other tumor types [93–95]. The matrix risk score was modestly although significantly correlated with age at diagnosis in lung cancer patients (Additional file 1: Fig. S3E). Further analysis of bulk RNAseq data from young and old lungs [47] revealed that older lungs have a higher SqCC matrix risk score than younger lungs (Fig. 2I). Therefore, matrix expression changes that occur during lung aging appear to overlap to some extent with matrix remodeling seen in lung tumors. In this way, age-associated ECM changes may prime aged lungs for SqCC initiation and progression, as was recently reported in melanoma [93–95], and thereby partially underpin the increased clinical incidence of these tumors with age.

Together, these data indicate that ECM remodeling occurs early in disease and is associated with malignant progression. Profiling these changes at the time of diagnosis may assist in identifying patients with high risk of developing aggressive lung cancer.

SqCC matretypes are prognostic

SqCC has a high rate of recurrence, even among early-stage patients following curative surgery. The ability to predict which patients are at a higher risk of recurrence would represent a significant milestone and has the potential to dramatically improve patient outcome. Expression-based molecular subtyping on the bulk tumor, encompassing cancer cells and other non-malignant cells within the tumor microenvironment, has revealed that SqCC tumors can be clustered into distinct tumor subtypes, with differences in biology and prognosis [24]. While these molecular subtypes broadly differ in cell adhesion and ECM pathways [24], recent studies indicate that they do not fully capture SqCC heterogeneity [27] and a detailed appraisal of the ECM landscape in SqCC may provide additional clinically actionable insight into SqCC biology. By focusing on the core matrix expression of SqCC tumors, we sought to define whether the matrix of the SqCC tumor microenvironment can be

grouped into distinct matrix molecular subtypes (matreotypes), and to determine whether these matreotypes have prognostic value.

Consensus clustering of matrisomal genes was performed to identify tumor and non-tumor tissue with similar ECM gene expression profiles. This approach robustly revealed three clusters, with one cluster corresponding exclusively to non-tumor tissue (Fig. 3A, Additional file 1: Fig. S4A, B and C).

Broad inspection of the two tumor matreotypes identified one as enriched in matrisomal gene expression (designated ECM-High, Additional file 1: Table S4) and one depleted of matrisomal gene expression (designated ECM-Low, Additional file 1: Table S4), as reflected in the significant enrichment and depletion of both the KEGG ECM (Fig. 3B) and Reactome ECM degradation pathways (Fig. 3C) in the ECM-High and ECM-Low groups, respectively. Differential gene expression analysis identified that pro-fibrotic ECM genes are the most highly upregulated marker genes for the ECM-High matreotype (including COL5A1, COL1A2, THBS2, Additional file 1: Fig. S4D), while the ECM-Low matreotype is enriched in a small number of glycoproteins (including FNDC6 and BGLAP, Additional file 1: Fig. S4D). Importantly, upregulation of some matrix risk signature genes, including COL11A1, CTHRC1, COL10A1, and CILP2, in the ECM-High matreotype suggests some overlap between ECM remodeling profiles associated with risk and progression in SqCC.

Analysis of patient outcome for these two tumor matreotypes also identified distinct prognostic associations, with the ECM-high matreotype having significantly worse disease-specific survival than the ECM-low matreotype (Fig. 3D). These matreotypes were not significantly associated with tumor stage (Supplementary Table 4), with the poor survival of the ECM-High matreotype retained in the early-stage setting (Supplementary Figure 4E and F) suggesting that differentially prognostic matreotypes are independent of stage. Similarly, matreotypes were not significantly associated with clinical covariates gender, race, or smoking packyears (Additional file 1: Table S4). However, while there was a statistically significant difference in smoking behavior between patients with ECM-High and ECM-Low tumors, the predominance of smokers in SqCC cohorts (Additional file 1: Table S4) precludes robust assessment of matreotype associations with smoking behavior. Matreotype associations with smoking behavior will require future validation in larger cohorts with a greater diversity of smoking behavior, particularly a higher proportion of never smokers.

The prognostic association of SqCC matreotypes were independently validated in the NCI-MD cohort

of largely early-stage SqCC (Fig. 3E, Additional file 1: Table S5) and a similar trend was also seen in the UHN cohort of early-stage tumors [20] (Additional file 1: Fig. 4G). This prognostic association was validated at the protein level in picosirius-red-stained tissue microarrays of SqCC tumors imaged by polarized microscopy (Supplementary Figure 4H). When bound to fibrillar collagens (collagen types I, II, III and V), picosirius red enhances collagen's birefringent properties, enabling specific imaging of fibrillar collagens [96]. This enables selective imaging of fibrillar collagens, which include five of the top ten upregulated core matrisomal genes in the ECM-High compared with the ECM-Low matreotype (Additional File 1: Fig. 4H). Importantly, across all stages and in the early-stage setting of stage I and IIA tumors, multivariate Cox proportional hazards modeling indicates that a high picosirius red birefringence signal is significantly associated with survival independently of stage (Additional file 1: Fig. S4I and J, Tables S6 and S7). Together, these data indicate that the ECM-high matreotype is significantly associated with poor overall survival in SqCC. Of note, analysis of the mutational landscape of these tumors did not identify any significant association of either matreotype with oncogenic driver mutations or FGFR amplifications (Additional file 1: Fig. S5A and B, Table S4). Thus, the matreotypes and their prognostic associations appear to be independent of the mutational landscape of the tumor.

Analysis of canonical SqCC subtypes defined by Wilkerson et al. [24], namely primitive, classical, secretory, and basal subtypes, indicated that the ECM-Low matreotype is composed primarily of classical and primitive molecular subtypes, while the ECM-High matreotype was significantly enriched for secretory and basal subtypes (Table 2), consistent with cell adhesion and ECM receptor interactions as known features of basal SqCC [24]. These data suggest that SqCC matreotypes represent unique features of SqCC tumors, related to, but not entirely captured by, whole genome molecular subtyping.

As pro-fibrotic ECM remodeling identified in the ECM-High matreotype has been associated with worse prognosis in a number of different cancer types, including pancreatic and breast cancer [97], we sought to examine the prognostic value of these matreotypes in other solid tumor types, including lung adenocarcinoma. This pan-cancer analysis revealed that our matreotypes are not prognostic in lung adenocarcinoma (Additional file 1: Figure S5C), attesting to the subtype specificity of these matreotypes in NSCLC. However, the ECM-High matreotype was significantly associated with disease-specific and progression-free survival in a number of other solid tumor types, including pancreatic ductal adenocarcinoma (PAAD) (Additional file 1: Fig. S5D and

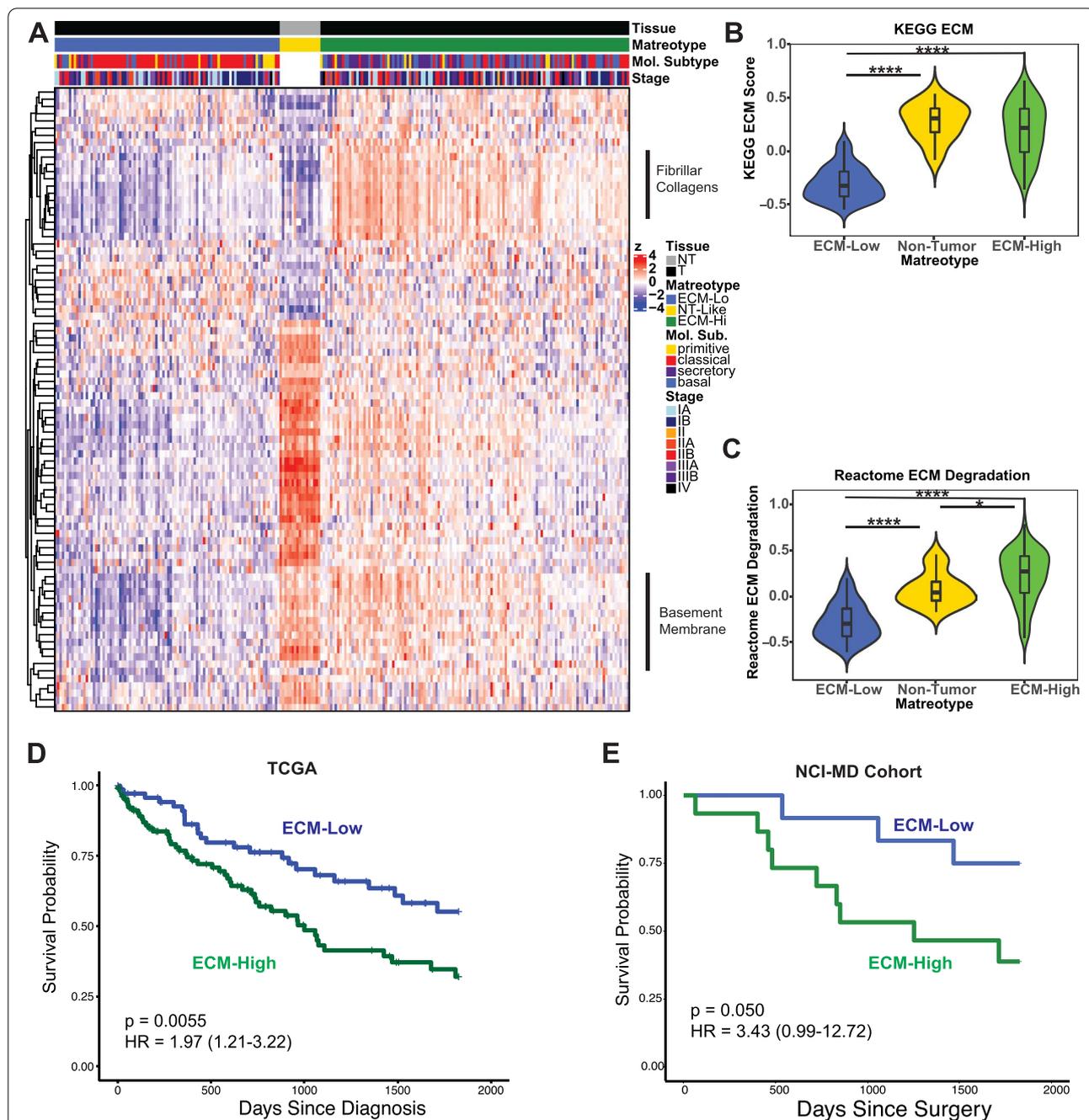


Fig. 3 The ECM-High matreotype is associated with poor prognosis. **A** Gene expression heatmap of core matrisomal genes differentially expressed between tumor and non-tumor tissue illustrating the three major matreotypes (ECM-Low, Non-Tumor, and ECM-High) identified by consensus clustering. **B,C** ECM-specific signatures assigned to the three matreotypes reveal ECM-High and ECM-Low matreotypes compared with non-tumor (NT) matreotype for the KEGG ECM (**B**, ECM-Low vs ECM-High $p < 2.2E-16$, NT-Like vs ECM-High $p = 0.27$, ECM-Low vs NT-Like $p = 1.3E-10$, Mann-Whitney U test) and Reactome ECM degradation pathways (**C**, ECM-Low vs ECM-High $p < 2.2E-16$, NT-Like vs ECM-High $p = 0.011$, ECM-Low vs NT-Like $p = 1.3E-7$, Mann-Whitney U test). **D,E** Disease-specific survival of tumors corresponding to the ECM-High (blue) and ECM-Low (green) matreotypes in the TCGA (**D**, ECM-Low $n = 73$; ECM-High $n = 107$; Log-rank $p = 0.0055$, Cox proportional hazards model HR (95% confidence interval) = 1.97 (1.21–3.22) of ECM-High compared with reference ECM-Low; multivariate model with age and stage covariates HR = 1.92 (1.17–3.17) $p = 0.010$) and NCI-MD cohorts (**E**, ECM-Low $n = 14$, ECM-High $n = 16$; Log-rank $p = 0.050$, Cox proportional hazards model HR (95% confidence interval) = 3.43 (0.99–12.72); multivariate model with stage covariate HR = 1.47 (1.01–18.55) $p = 0.048$)

E, Table S8). This suggests that ECM remodeling associated with these matreotypes may play an important role in the progression of a specific but diverse subset of solid tumors.

Prognostic SqCC Matreotypes have divergent cellular ecosystems

Canonical molecular subtypes are thought to reflect different cellular compositions [24, 25] and since these matreotypes were identified in bulk RNAseq data, gene expression values reflect the expression of multiple cell types within the tumor microenvironment. To determine how different cell types within the tumor microenvironment may be contributing to the different ECM compositions of these matreotypes, we utilized cellular deconvolution approaches to infer the cellular composition of these tumors.

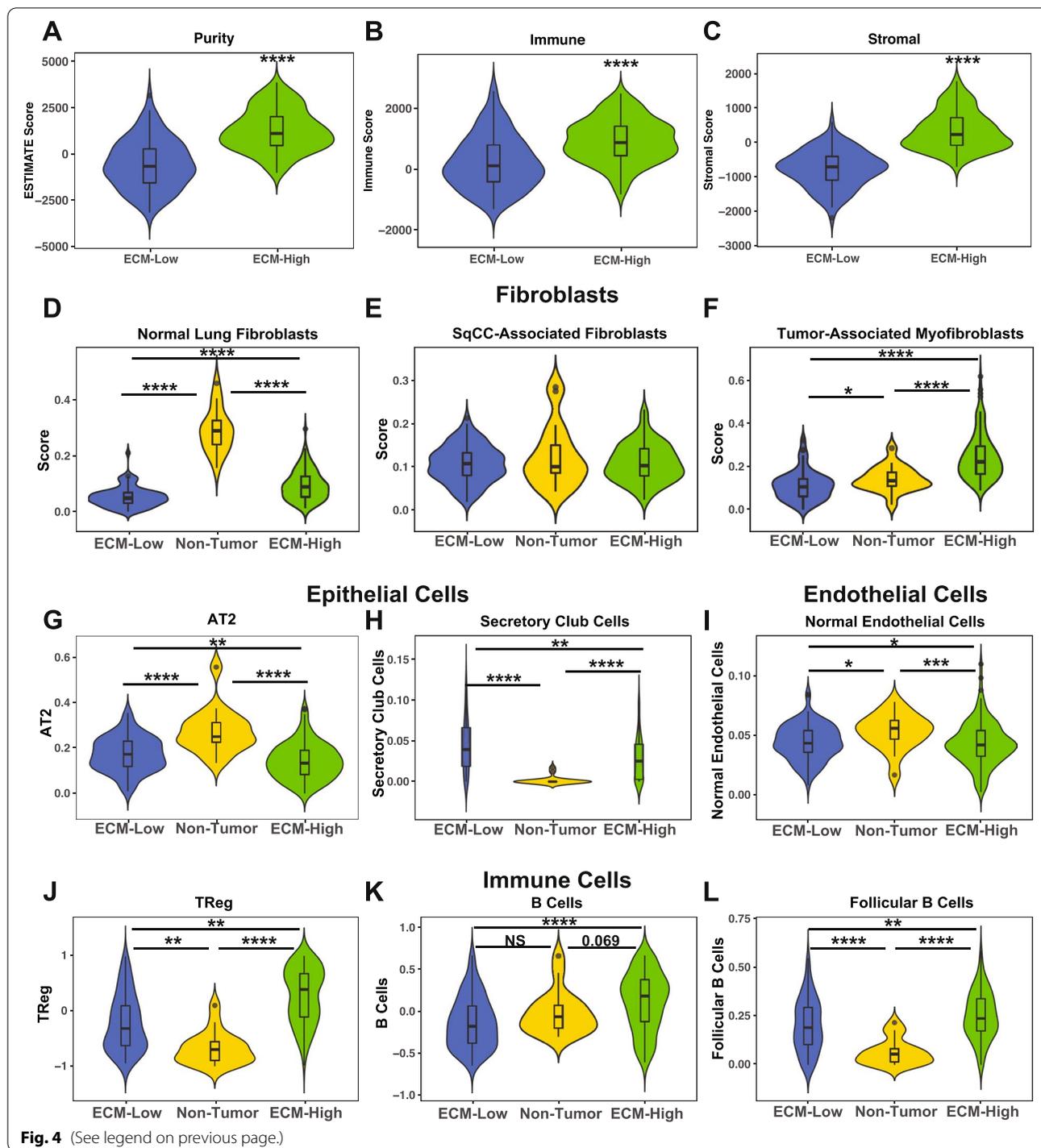
In silico estimation of tumor purity using ESTIMATE [61] revealed significantly lower tumor purity in the ECM-High matreotype compared with the ECM-Low matreotype (Fig. 4A). This lower tumor purity was due to significantly higher stromal (Fig. 4B) and immune scores (Fig. 4C) in the ECM-High compared with the ECM-Low matreotype, suggesting that ECM-High matreotype tumors may harbor a more diverse cellular ecosystem. Of note, tumor purity itself ($p=0.2$, HR = 1 [0.99–1], Cox proportional hazards model) and immune scores ($p=0.72$, HR = 1 [0.99–1]) were not significantly associated with survival, and the stromal score ($p=0.039$, HR = 1.0003 [1–1.0006]) was only marginally associated with survival. This indicates that broad enrichment of stromal cells within the tumor microenvironment does not explain the prognostic value of the matreotypes and highlights that the specific matrisomal profile of SqCC may be a stronger indicator of prognosis than tumor purity alone.

To further interrogate these differences in more detail, deconvolution approaches were applied using cell type-specific signatures from NSCLC scRNAseq data [46] using CIBERSORTx [98]. While several fibroblast subtypes (subtypes 2 (COX4I2-high), 4 (PLA2GA-high) and 5 (MMP3-high) as identified in [46]) were not detected in more than 10% of samples in this cohort, normal fibroblasts were significantly depleted in both tumor matreotypes compared with the non-tumor matreotype (Fig. 4D). Furthermore, the ECM-High matreotype was significantly enriched with tumor-associated myofibroblast cells (Fig. 4F) but not a fibroblast subset originally identified in a subset of SqCC tumors but currently with no known functional significance (Fig. 4E). Additionally, the ECM-High matreotype was significantly depleted of type II pneumocytes (AT2, Fig. 4G), and secretory club cells (Fig. 4H) relative to the ECM-Low matreotype, suggesting the presence of tumor-associated epithelial dysregulation in these poor-prognosis tumors. As expected, the ECM-High matreotype was significantly depleted of normal endothelial cells (Fig. 4I), consistent with high expression of correlation cluster 2 genes (Fig. 1D) in this subset of tumors.

Direct comparison of immune cell scores in the ECM-High and ECM-Low matreotypes using NSCLC-specific immune signatures [51] indicated that the ECM-High matreotype is relatively enriched in a variety of immune cells compared with the ECM-Low matreotype, but levels remain lower than non-tumor tissue (Additional file 1: Fig. S6). However, of particular note was that this matreotype was significantly enriched with regulatory T cells (Fig. 4J), B cells (Fig. 4K), and the follicular B cell subset (Fig. 4L) compared with both non-tumor tissue and the ECM-Low matreotype. Interestingly, T-follicular helper cells (Additional file 1: Fig. S6A) were also significantly enriched in the ECM-High matreotype, suggesting the potential enrichment of tertiary lymphoid structures in

(See figure on next page.)

Fig. 4 SqCC Matreotypes have distinct cellular ecosystems. **A–C** Comparison of tumor purity (**A**, $p<2.2E-16$, Mann-Whitney U test), immunological enrichment (**B**, $p=8.0E-10$, Mann-Whitney U test), and stromal enrichment (**C**, $p<2.2E-16$, Mann-Whitney U test) in the ECM-High (green) compared with ECM-Low (blue) matreotypes identifies significant enrichment of both stromal and immune cells in ECM-High matreotype tumors. **D–F** Fibroblast cell types in the three major SqCC matreotypes ECM-Low (blue), ECM-High (green), and non-tumor (yellow) with fibroblasts from non-tumor tissue (**D**, ECM-Low vs ECM-High $p=3.8E-10$, ECM-Low vs Non-tumor $p=7.3E-11$, ECM-High vs Non-Tumor $p=7.8E-11$, Mann-Whitney U test), SqCC-associated fibroblasts (**E**, ECM-Low vs ECM-High $p=0.85$, ECM-Low vs Non-tumor $p=0.59$, ECM-High vs Non-Tumor $p=0.66$, Mann-Whitney U test) and tumor-associated myofibroblasts (**F**, ECM-Low vs ECM-High $p=2.3E-22$, ECM-Low vs Non-tumor $p=0.036$, ECM-High vs Non-Tumor $p=1.8E-5$, Mann-Whitney U test). **G,H** Epithelial cell types in the three major SqCC matreotypes ECM-Low (blue), ECM-High (green), and non-tumor (yellow) corresponding to type II pneumocytes (**G**, AT2, ECM-Low vs ECM-High $p=0.0018$, ECM-High vs non-Tumor $p=3.1E-7$, ECM-Low vs Non-tumor $p=6.6E-5$, Mann-Whitney U test) and secretory club cells (**H**, ECM-Low vs ECM-High $p=0.00042$, ECM-High vs non-Tumor $p=1.8E-6$, ECM-Low vs Non-tumor $p=2.5E-8$, Mann-Whitney U test). **I** The ECM-High matreotype is significantly depleted of normal endothelial cells (ECM-Low vs ECM-High $p=0.034$, ECM-High vs non-Tumor $p=0.00038$, ECM-Low vs Non-tumor $p=0.013$, Mann-Whitney U test). **J–L** Immunological cell types in the three major SqCC matreotypes ECM-Low (blue), ECM-High (green), and non-tumor (yellow) corresponding to regulatory T cells (**J**, ECM-Low vs ECM-High $p=9.7E-10$, ECM-High vs non-Tumor $p=1.5E-7$, ECM-Low vs Non-tumor $p=0.0060$, Mann-Whitney U test), B cells (**K**, ECM-Low vs ECM-High $p=4.9E-6$, ECM-High vs non-Tumor $p=0.069$, ECM-Low vs Non-tumor $p=1$, Mann-Whitney U test), and particularly the follicular subset of B cells (**L**, ECM-Low vs ECM-High $p=0.0044$, ECM-High vs non-Tumor $p=3.7E-8$, ECM-Low vs Non-tumor $p=5.5E-5$, Mann-Whitney U test).



this matretypic. Furthermore, a higher MHC Class II expression score in the ECM-High matretypic (Additional file 1: Fig. S6E) suggests the potential for differences in immunological activation in these matretypes. While regulators of antitumoral immune surveillance are emerging, including matrix-mediated T cell exclusion

[99], these data warrant further functional and clinical investigations into the potential for matretypes to inform the prioritization of patients for immune checkpoint therapies.

Together, these data suggest that cellular heterogeneity in SqCC may contribute to differences in the ECM profile

Table 2 SqCC matreotype association with canonical SqCC molecular subtypes

Matreotype	Primitive	Classical	Secretory	Basal	Predominant canonical subtype
ECM-Low	12 (57.2%)	61 (64.9%)	7 (14.6%)	14 (23.2%)	Classical
ECM-High	9 (42.8%)	33 (35.1%)	41 (85.4%)	46 (76.7%)	Secretory/basal
<i>P-value</i>	0.167	5.18E−9	6.67E−9	6.93E−4	

of the tumors. The differences in these cellular ecosystems have potential implications for the tumor etiology and the efficacy of therapeutic interventions, and in particular immune checkpoint inhibitors.

Matrisomal features are associated with pro-invasive signaling

To further examine the distinct tumor biology associated with these matreotypes, we performed pathway enrichment analysis comparing the ECM-Low and ECM-High groups using the MSigDb Hallmark pathway database. As expected, ECM-related pathways were the most highly enriched in the ECM-High matreotype (Additional file 2: Table S1). Examination of non-ECM-associated pathways revealed multiple differentially expressed hallmark pathways, including enrichment of inflammatory, epithelial-to-mesenchymal transition, angiogenesis, and cell-cell adhesion pathways in the ECM-High matreotype (Fig. 5A). Comparatively, the ECM-Low matreotype was enriched in DNA repair pathways, including the G2/M checkpoint as well as ROS signaling (Fig. 5A).

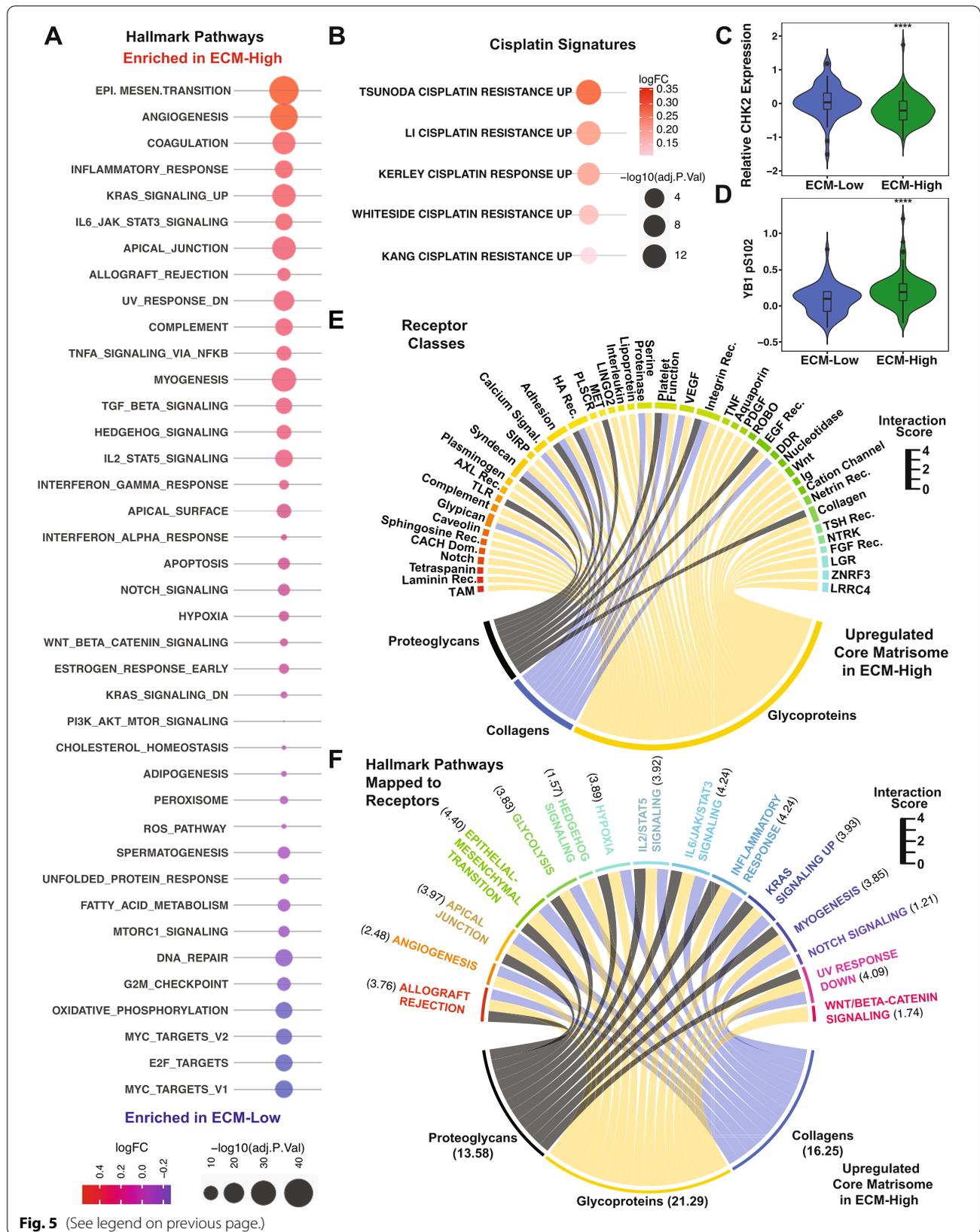
DNA damage pathways play a significant role in regulating resistance to conventional chemotherapy agents, including platinum-based agents which are a frontline therapy for SqCC [100]. Pathway analysis of ECM-High and ECM-Low matreotypes using the MSigDb C2 oncogenic database (Additional file 1: Fig. S7A, Additional file 2: Table S2), which includes signatures specific to

tumor biology and therapeutic responses, identified significant enrichment of multiple cisplatin resistance gene signatures in the ECM-High matreotype compared with the ECM-Low matreotype (Fig. 5B), suggesting that DNA damage checkpoint deficiencies in ECM-High tumors may contribute to poor treatment response in these patients. Analysis of reverse phase protein array proteomic TCGA data in these same tumors [42] indicated significantly reduced expression of the double-stranded break response protein CHK2 (Fig. 5C) and increased activation of YB1 (pS102, Fig. 5D) in ECM-High matreotype tumors compared with ECM-Low matreotype tumors at the protein level. CHK2 and YB1 are known mediators of platinum resistance [101–103] and may contribute to poor platinum response in ECM-High tumors.

ECM components directly interact with cell surface receptors to regulate the activity of numerous signaling pathways, including those involved in epithelial-to-mesenchymal transition and ECM production [35]. Ligand-receptor interaction analysis was performed to gain insight into the potential direct effects of these matreotypes on outside-in signaling. A database of curated ligand-receptor interactions [65] was mined for interactions between core matrisomal genes differentially expressed between ECM-High and ECM-Low SqCC tumors and their cognate receptors. An interaction strength was assigned to each core matrisomal

(See figure on next page.)

Fig. 5 ECM components contribute to signaling pathways associated with prognosis. **A** Top 25 up- and downregulated differentially enriched MSigDb hallmark pathways in the ECM-High matreotype compared with the ECM-Low matreotype. Circle size indicates the inverse of the $-\log_{10}(p\text{-value})$, color reflects the log fold change (logFC) of the ECM-High matreotype compared with the ECM-Low matreotype. **B** Pathway analysis of cisplatin-related signatures comparing ECM-High matreotype with the ECM-Low matreotype. **C** Abundance of CHK2 in the ECM-High matreotype (green) compared with the ECM-Low matreotype (blue). $p=1.5E-4$, Mann-Whitney U test. **D** Comparison of the phosphorylated YB1(S102) in the ECM-High matreotype (green) compared with the ECM-Low matreotype (blue). $p=1.1E-3$, Mann-Whitney U test. **E** Ligand-receptor interaction results of differentially expressed core matrisomal genes significantly upregulated in ECM-High compared with ECM-Low matreotypes and the receptor groups that they directly interact with. Sector width is the aggregated log fold change of the ligand-receptor interaction strength comparing the ECM-High to ECM-Low matreotype. The scale indicates the width of each link between the ECM category and its cognate receptors. The widths of each link are the maximum fold change of the ligand-receptor interaction scores comparing the ECM-High and ECM-Low groups for that ligand-receptor interaction. **F** Hallmark pathways regulated by ligand-receptor interactions that are significantly enriched in the ECM-High compared with the ECM-Low matreotypes. The scale indicates the width of each link between the ECM category and the hallmark pathway. The widths of each link are the maximum fold change of the ligand-receptor interaction scores comparing the ECM-High and ECM-Low groups for that hallmark pathway



ligand-receptor interaction based on their expression levels, to infer the potential for core matrisome-driven receptor activation. Receptors were then grouped into functional classes and mapped to signaling pathways by their MSigDb hallmark signatures. The ECM profile of the ECM-High matreotype was associated with significant enrichment of multiple receptor classes (Fig. 5E, Additional file 2: Table S3) involved in diverse signaling pathways. The most significantly enriched of these core matrisome-receptor pairs were fibrillar collagens (COL1A2, COL1A1, COL5A1, COL5A2) interacting with receptors FLT4/VEGFR3, integrins α 1, α 11, and β 1 as well as CD93 (Additional file 2: Table S3). Conversely, only interactions between the glycoprotein ZP3 and EGF and MERTK (TAM family) receptors were identified in the ECM-Low matreotype (Additional file 1: Fig. S7B), although the strength of these interactions was not significantly higher compared with the ECM-High matreotype. Importantly, ligand-receptor interactions that were significantly enriched in the ECM-High matreotype (Fig. 5E) could be mapped to significantly upregulated hallmark pathways in this matreotype (Fig. 5F) (including epithelial to mesenchymal transition, apical junction, and myogenesis (Fig. 5A)) suggesting that core matrisome components have the potential to contribute to the upregulation of these biological pathways within different SqCC matreotypes. Together, these data indicate the potential for the ECM profile of different SqCC to regulate cancer cell biology and the mechanisms that promote tumor progression.

The pro-fibrotic matrisome is associated with pro-invasive signaling in poor-prognosis SqCC

Integrins represented the largest class of receptors with the potential to be activated by the various matrix elements of the ECM-High matreotype. Integrins have been demonstrated to play a role in NSCLC tumorigenicity [104, 105] as well as activating epithelial to mesenchymal signaling to support cancer cell metastasis and in activating fibroblasts to a myofibroblast-like state seen in cancer-associated fibroblasts.

Collagens, proteoglycans, and glycoproteins that were enriched in the ECM-High matreotype were identified as ligands for a range of integrin receptors (Fig. 6A), with known functional roles in cell survival, migration, and invasion required for metastatic relapse and in fibroblast activation [106]. Similarly, integrin receptors themselves (Fig. 6B), as well as components of the integrin adhesome (Additional file 1: Fig. S8A), were also overexpressed in the ECM-High matreotype compared with the ECM-Low matreotype (Fig. 6B, C), suggesting the potential for further reinforcement of integrin-mediated signaling in these tumors.

In accordance with elevated integrin signaling in this matreotype, interrogation of TCGA RPPA proteomic data from these tumors identified significant phosphorylation of integrin phospho-adhesome component PKCdelta [107] (pS664, Additional file 1: Fig. S6B) as well as activation of the MEK/ERK signaling pathway through increased activation of MEK1 (pS217/221) and ERK1/2 (p202/204, Fig. 6D–G). Similarly, we observed increased expression of MEK/ERK downstream targets p21 (protein: Fig. 6G; transcript: Fig. 6K) and master EMT regulators of the ZEB (ZEB1, ZEB2), SNAIL (SNAI1, SNAI2), and TWIST (TWIST1, TWIST2, and SLUG (SNAI3)) transcription factor families (Fig. 6H–J and Additional file 1: Fig. S8C). It should be noted that other receptors, such as EGFR, can also activate MEK/ERK signaling and may also contribute to these proteomic phenotypes. Similarly, since these downstream effectors and transcription factors are also involved in integrin-dependent and integrin-independent fibroblast activation, higher levels of these factors may also reflect increased activated fibroblasts within ECM-High tumors. Interestingly, high expression of integrin/MEK/ERK target p21 has also been implicated in mediating cisplatin resistance in NSCLC and other cancers [108, 109] suggesting that elevated integrin signaling as a result of an ECM-High matreotype could underpin the chemotherapy resistance predicted for these tumors (Fig. 5B).

In addition, myosin IIA phosphorylation was significantly upregulated in the ECM-High matreotype (Fig. 6F). Myosin IIA activation supports cancer cell migration to promote metastasis in NSCLC [110, 111] and is also associated with cancer-associated fibroblast activation [112] warranting further mechanistic validation of these signaling events in supporting metastatic propensity and fibroblast activation in the ECM-High matreotype.

The matrisome contributes to fibrogenic signaling in SqCC

In addition to impacting cancer cell behavior, elevated integrin activation driven by matrisomal components of the ECM-High matreotype (Fig. 6) may also promote fibroblast to myofibroblast transdifferentiation [113–115] that is thought to contribute to the expression of pro-fibrotic ECM components in the ECM-High SqCC. In this way, integrin signaling amplified by the core matrisome may perpetuate increased pro-fibrotic ECM remodeling and subsequent pro-metastatic signaling.

The TGF β , PDGF, FGF, and VEGF signaling pathways are also well characterized fibrogenic pathways that can induce the transdifferentiation of fibroblasts towards a myofibroblast-like phenotype observed in NSCLC fibroblast subsets [116–118]. In order to understand the potential direct impact of these matreotypes

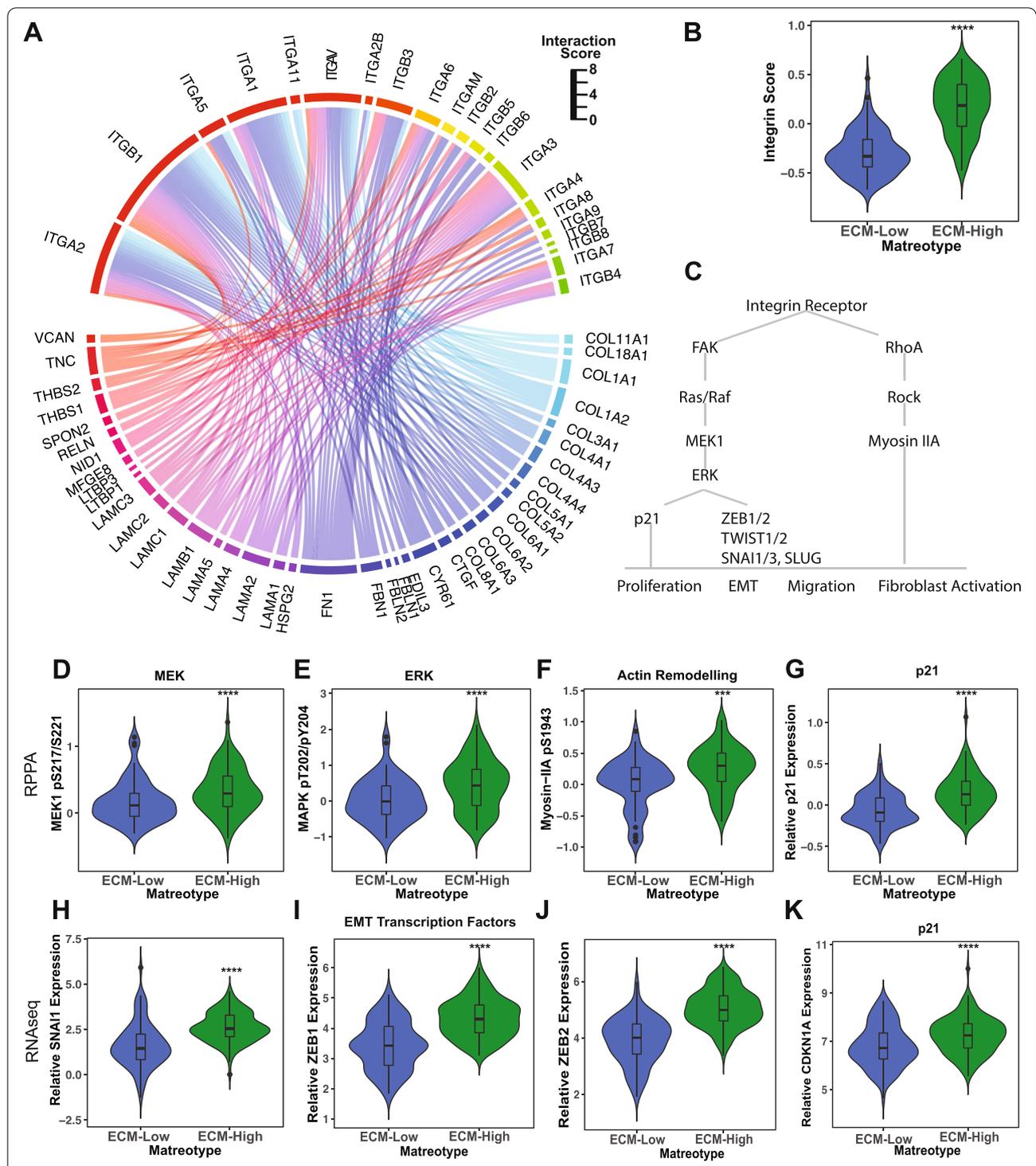


Fig. 6 ECM-driven integrin signaling is associated with EMT and fibroblast activation in the ECM-High matreotype. **A** Ligand-receptor interaction results of differentially expressed core matrisomal genes and their cognate integrin receptors upregulated in ECM-High compared with ECM-Low matreotypes. Sector width is the aggregated log fold change of the ligand-receptor interaction strength comparing the ECM-High to ECM-Low matreotype. **B** Integrins are upregulated in the ECM-High compared with the ECM-Low matreotypes as indicated by the integrin receptor score. $p=1.3E-23$, Mann-Whitney U test. **C** Schematic of integrin-mediated signaling pathways active in the ECM-High matreotype. **D–G** Comparison of integrin-dependent signaling in the ECM-High (green) and ECM-Low (blue) matreotypes as indicated by the RPPA abundance of MEK1 pS217/S221 (**D**, $p=1.6E-3$), ERK pT202/204 (**E**, MAPK pT202/204, $p=3.4E-4$), Myosin IIA pS1943 (**F**, $p=0.00027$), and p21 (**G**, $p=8.1E-8$), Mann-Whitney U test. **H–K** Expression of genes SNAI1 (**H**, $p=7.3E-12$), ZEB1 (**I**, $p=9.9E-15$), ZEB2 (**J**, $p=1.9E-21$), and P21/CDKN1A (**K**, $p=0.00014$) in RNAseq data in the ECM-High (green) and ECM-Low (blue) matreotypes; Mann-Whitney U test. *** $p<0.001$, **** $p<0.0001$

on pro-fibrogenic receptors in the tumor microenvironment, we performed a focused analysis of the interactions between core matrisomal genes that are significantly enriched in the ECM-High matreotype and the fibrogenic PDGF, FGF, and VEGF receptors. This analysis revealed that collagen I, fibronectin, and the milk fat globule EGF and factor V/VIII domain containing glycoprotein MFGE8, which are highly expressed in ECM-High tumors, interact with the PDGFRB and FLT4 receptors (Fig. 7A, B). High PDGFR pathway activity in the ECM-High matreotype is also reflected in significant enrichment of signatures for response to PDGFR inhibition by imatinib compared with the ECM-Low matreotype (Additional file 1: Fig. S9A, Additional file 2: Table S2).

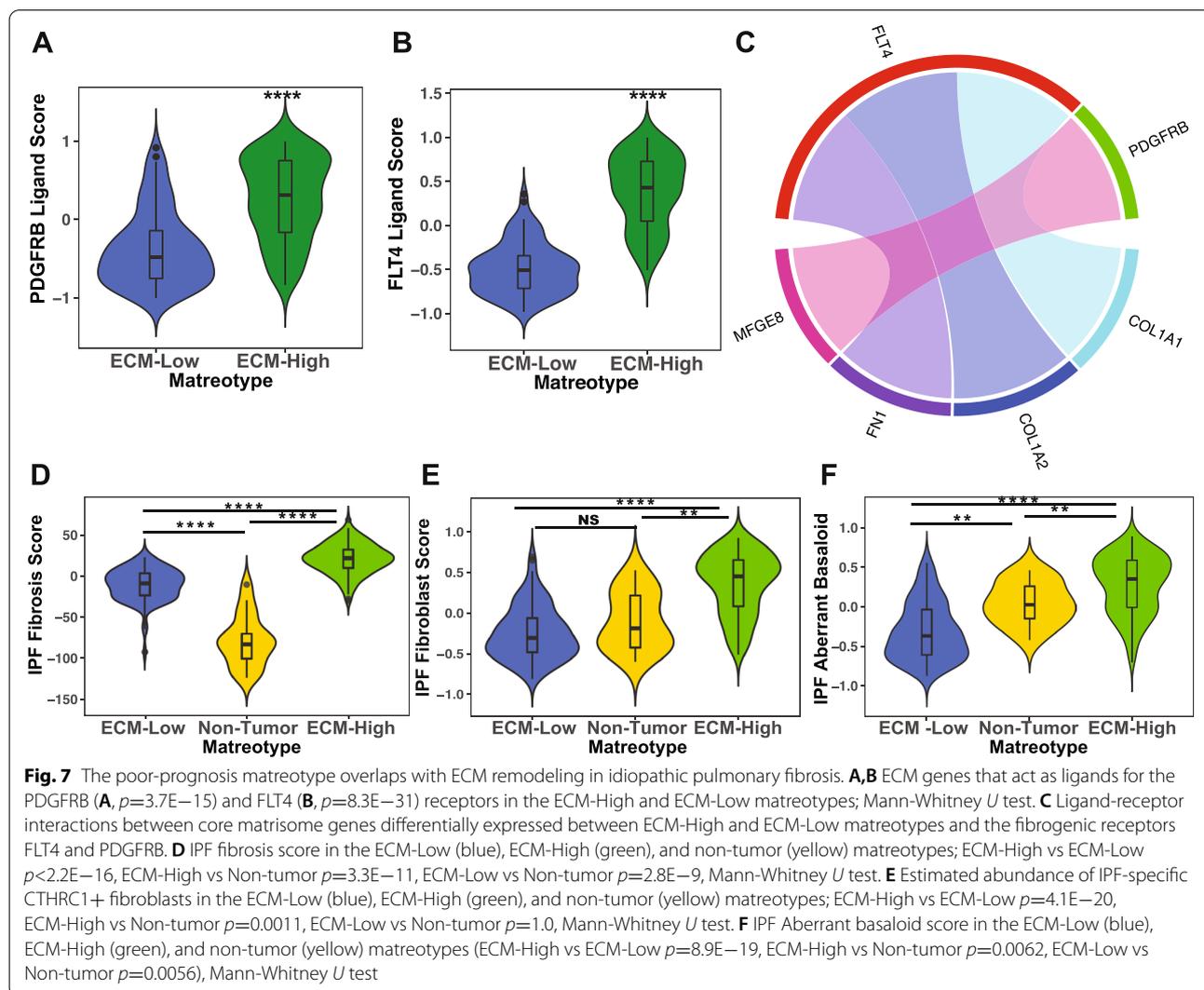
Therefore, in addition to integrin- and biomechanical-induced fibrogenic signaling, direct interaction of the core matrisome with PDGF and VEGF receptors may also activate fibrogenic signaling thereby amplifying the

deposition of pro-fibrotic ECM associated with poor prognosis in SqCC patients.

Prognostic matreotype overlap with idiopathic pulmonary fibrosis

Fibrogenic signaling through these tyrosine kinase receptors is known to play a role in the progression of idiopathic pulmonary fibrosis [119], a chronic, degenerative lung condition characterized by extensive ECM remodeling that typically presents at older age and is associated with increased risk of developing lung cancer [120].

Although the precise molecular drivers of IPF remain unclear, parallels have been drawn between the pathogenesis of IPF and lung cancer and ECM remodeling has been implicated in increased lung cancer risk in these patients [121]. We sought to understand to what extent the matrisomal profile of the poor-prognosis ECM-High matreotype overlaps with that of IPF. We applied



an IPF-derived fibrosis score developed from human IPF lung transcriptomic data [48] to SqCC. Our analysis showed that tumor tissue had a higher fibrosis score than non-tumor tissue and that this fibrosis score was highest in the ECM-High matreotype (Fig. 7D). This indicates that the ECM remodeling associated with poor prognosis in NSCLC overlaps to some extent with ECM remodeling that defines IPF.

Phenotypic reprogramming of fibroblasts towards a pro-fibrogenic state, with enhanced proliferative and invasive capacity, is believed to drive progressive lung fibrosis in IPF [122]. These IPF-specific fibroblasts have phenotypes that are distinct from normal lung fibroblasts and are characterized by high expression of fibrotic ECM genes such as COL1A1 and CTHRC1 [49, 123]. Deconvolution of bulk RNAseq data in tumors identified that the ECM-High matreotype was significantly enriched with the transcriptomic profile of the CTHRC1+ IPF-specific fibroblast population (Fig. 7E) thought to be important in IPF pathogenesis [49]. Similarly, the ECM-High matreotype was significantly enriched in IPF-specific aberrant basaloid cells identified exclusively in IPF lungs (Fig. 7F), which are thought to result from epithelial dysregulation in IPF and contribute to aberrant fibroblast phenotypes in this disease [50]. Together, this suggests that fibroblast and epithelial dysregulation featured in IPF pathogenesis may also operate in a subset of poor-prognosis SqCC, contributing to matrisomal dysregulation and tumor progression.

Overall, our analyses have identified ECM profiles of CIS lesions and carcinomas that are associated with progression and poor prognosis. Overlap in the ECM components that are represented in the matrix risk signature and are upregulated in the ECM-High matreotype suggest that early ECM remodeling programs initiated prior to the onset of, or very early on, in malignancy may establish an ECM profile that supports aggressive tumor behavior and that persists throughout tumor progression. In silico analysis raises the notion that an amplification loop of fibrotic ECM production, integrin activation, and pro-fibrogenic processes has the potential to promote chemotherapy resistance and metastatic recurrence in a subset of poor-prognosis SqCC. An improved understanding of the ECM landscape in SqCC may identify more effective therapeutic strategies that co-target the tumor-supporting components of the ECM.

Discussion

Treatment options and efficacy remain limited in the SqCC subtype. The lack of targetable mutations in these tumors has hindered progress in the development of more effective, personalized treatment approaches that have revolutionized patient outcome in subsets of lung

adenocarcinoma patients [4]. Furthermore, significant tumor heterogeneity [9] and the high mutational burden of these tumors [10] make it challenging to achieve sustained clinical responses by targeting tumor cells alone [11, 12]. Understanding how the tumor microenvironment contributes to tumor progression may reveal opportunities to target the tumor-supporting elements of this microenvironment in order to improve treatment efficacy. Furthermore, it is likely to assist in the identification of patients at higher risk of developing aggressive disease, or of relapsing. Through focused characterization of the tumor ECM, our analysis has revealed that subtype-specific ECM remodeling is associated with tumor initiation and progression in SqCC, uncovering potential opportunities to improve the diagnosis and treatment of subsets of SqCC patients with the poorest prognosis.

Components of the ECM form a highly interconnected and dynamic arrangement of macromolecules [35]. It is increasingly appreciated that the architecture of these components and their association with one another, not just their abundance, can induce heterogeneous effects on cell behavior [31]. Our matrix risk signature has defined key ECM genes that distinguish between tumor and non-tumor tissue in SqCC, as well as in lung adenocarcinoma and various cancer types. Our analysis indicates that these key ECM components are highly correlated with many other ECM components, and together with their known scaffolding functions with other ECM molecules, suggests that tumor- and aging-associated changes in the expression of these key matrix components likely result in profound changes in the organization of the lung ECM. These cumulative changes then feed into SqCC progression. We identified significant overlap in the SqCC and adenocarcinoma risk signatures for matrix components that regulate collagen fibril architecture including COL10A1, COL11A1, and CTHRC1 [124] suggesting that our risk signature represents changes in the organization of fibrillar collagens. COL10A1, COL11A1, and CTHRC1 were also identified in a matrix-specific NSCLC subtype-independent prognostic signature [17], supporting the involvement of these matrix components in lung tumorigenesis generally. EMILIN2, which is thought to confer elasticity to tissues and modulate receptor signaling, was identified as downregulated in both SqCC and adenocarcinomas. While its function in non-small cell lung tumors is not currently known, EMILIN2 has been associated with poor prognosis in gastric cancer where it modulates cancer cell apoptosis and tumor angiogenesis [125]. The glycoprotein osteopontin SPP1 was also identified as differentially expressed in both squamous and adenocarcinoma tumors. SPP1 has recently been identified as

a marker of macrophage polarization and implicated in mediating immune evasion [79] and immunotherapy response [126] consistent with immunological mechanisms regulating lung tumorigenesis [8]. However, our analysis also identified core matrisome components that were associated with SqCC tumors but not adenocarcinoma tumors. These squamous-specific components include COL7A1, which is known to play an important role in squamous tumorigenesis in the skin and esophagus [87, 88], and COL4A6, which we identified as correlating with the expression of squamous master regulator TP63. While an integrated picture of how these matrix components collectively regulate ECM composition and architecture in lung tumors generally, as well as in the specific SqCC and adenocarcinoma subtypes, is not yet clear, the applicability of this risk signature to non-lung primary tumors raises the notion that tumorigenesis may involve overlapping ECM remodeling processes across diverse tumor types.

The implementation of lung cancer screening has improved mortality through the early detection of overt tumors, as well as pre-invasive lesions, which are more prevalent in screening programs than positive tumor diagnoses [127, 128]. Pre-invasive lesions, which are dysplastic but are not yet invasive [129], spontaneously regress in approximately 30% of patients [7]. They are typically periodically monitored for accelerated growth and the acquisition of invasive characteristics before interventions are recommended. There are currently no markers for those premalignant lesions that are likely to progress to cancer, and as such the clinical management of these patients to avert the initiation of lung cancer remains challenging. Our analysis indicates that CIS lesions that progress to SqCC, which histologically are indistinguishable from those CIS lesions that will spontaneously regress, have acquired tumor-like ECM profiles. Identification of these features during monitoring may enable patients with high-risk premalignant lesions to be prioritized for surgical intervention to prevent the onset of invasive SqCC. Our analysis of fibrillar collagens by picrosirius red staining of tissue microarrays suggests that such an approach, which could integrate with existing pathological processing of tumor samples, could be used to prognosticate tumors by matreotype. It remains to be determined if key ECM markers from the matrix signatures could also be assessed at the protein level using immunohistochemistry for the prognostication of premalignant lesions (risk signature), and also at SqCC diagnosis (matreotyping as ECM-High or ECM-Low). Immunohistochemistry staining for specific ECM markers identified in our analysis may be more compatible with existing pathological analysis pipelines than transcriptomic approaches. This will be integral for rapid

implementation of ECM profiling for the prioritization of high-risk patients for more rigorous intervention or stromal co-targeting therapies.

Growing evidence points to immune surveillance as a major regulator of premalignant progression in SqCC [8]. With this in mind, the glycoprotein SPP1, which we identified as positively associated with SqCC risk, is produced by tumor-associated macrophages *in vitro* and induces PD-L1 expression in lung adenocarcinoma [79]. This suggests that the immunomodulatory activity of ECM components in our matrix risk signature may contribute to immune evasion mechanisms that enable premalignant progression. As has been demonstrated in other tumor types, the arrangement of multiple ECM components such as fibrillar collagens and their associated matrix molecules has the potential to also regulate immunological surveillance. For example, fibrillar collagen architecture, which is regulated by several ECM components in our risk signature (including COL11A1, CTHRC1, and COL10A1), is also associated with altered T cell-mediated immune surveillance [95, 130, 131]. Preliminary data also suggests that a COL11A1-expressing CAF subset may also promote T cell exclusion from the tumor microenvironment [132]. How these ECM components influence anti-tumor immune surveillance will be an important consideration in understanding and predicting the immune checkpoint therapy response.

Changes in collagen architecture also occur during lung aging [47, 92]. The enrichment of our matrix risk signature in older compared with younger lungs supports the notion that age-related ECM remodeling may prime lung tissue for lung tumor initiation and underpin the increased incidence of lung cancer with age. Age-related changes in dermal collagen architecture induced by HAPLN1 has been linked to reduced T cell motility and infiltration, and increased melanoma incidence with age [93, 95], raising the possibility that ECM remodeling represented by our matrix risk signature captures a similar process in SqCC. Whether ECM components that constitute the risk signature reflect or regulate immune surveillance to enable unhindered progression of dysplastic lesions warrants further investigation.

Several of the matrix components identified in the matrix risk signature are also significantly overexpressed in our prognostic ECM-High matreotype that constitutes a subgroup of SqCC patients with the worst prognosis. This, together with the stage independence of these matreotypes, suggests that ECM remodeling during the initiation of lung tumors may establish an ECM remodeling program that supports subsequent aggressive tumor growth and metastasis that persists throughout tumor progression. Our identification of matrisomally driven

integrin and fibrogenic signaling in this ECM-High matreotype suggests that the initiation of the pro-fibrotic ECM remodeling generates a positive feedback loop to promote further fibrotic ECM deposition and activation of these pro-metastatic signaling pathways.

The association of the ECM-enriched, ECM-High matreotype with worse prognosis is consistent with an observed association of increased peritumoral stroma (as defined by H&E pathology) with worse overall and recurrence-free survival, and significantly reduced expression of the EMT marker E-cadherin in SqCC [133]. Early establishment of these ECM remodeling programs is also consistent with observations that pro-fibrotic gene expression changes occur early in SqCC progression, coinciding with the acquisition of an invasive phenotype [134], and therefore may enable early cancer cell dissemination from the primary site. Furthermore, numerous studies support the association of pre-existing fibrotic ECM remodeling from chronic lung diseases with worse lung cancer survival [135, 136], potentially due to increased invasion in these tumors [135]. Our inference of matrisome-driven receptor activation suggests that the ECM itself has the potential to further amplify integrin signaling to support the metastatic spread of cancer cells and the activation of fibroblasts and warrants further functional studies to dissect the contribution of integrin signaling in these cell types to tumor biology.

Advanced stage SqCC and adenocarcinoma that lack targetable mutations and are candidates for immune checkpoint inhibitors are largely treated in a similar manner. However, our analysis indicates that prognostic ECM remodeling programs in SqCC are distinct from those in lung adenocarcinoma. While adenocarcinoma has been reported to be more fibrotic than SqCC [137], our analysis indicates that it is the upregulation of other specific core matrisomal components, not just fibrillar collagens, that are prognostic in SqCC. The distinct biology of these two NSCLC subtypes is supported by a recent proteogenomic study that identified the majority of SqCC clustered distinctly from adenocarcinoma subtypes [27]. That these tumors respond differently to the matrix environment may reflect the fact that SqCC and adenocarcinomas typically arise in different regions of the lung (and likely different cells of origin [138]), which are characterized by a different ECM composition [139]. Lung fibroblasts within these different compartments also display distinct phenotypes, and fibroblasts derived from SqCC tumors accumulate more than adenocarcinoma-associated fibroblasts via beta-1 integrin-dependent mechanisms [114], which may also accelerate matrisome-mediated pro-fibrogenic signaling in the ECM-High matreotype. Differential effects of the ECM on lung cancer histology have also been noted in the context of

LKB1-negative tumors, where LOX-dependent collagen crosslinking can drive the transdifferentiation of adenocarcinomas to SqCC [140], although the precise mechanism by which the ECM potentially modulates lung cancer histology remains unclear.

Our poor-prognosis ECM-High matreotype has an ECM profile which overlaps with IPF and appears enriched for an IPF-specific fibroblast phenotype, suggesting that matrix remodeling in this subset of SqCC with the worst prognosis may be driven by similar mechanisms as those underlying IPF. Enrichment of the IPF-specific aberrant basaloid cell signature in the ECM-High matreotype may also reflect the known epithelial dysregulation associated with SqCC and the transformation of basal cells, which are the likely cells of origin in SqCC [138]. Interestingly, this epithelial cell type has high expression of EMT markers and p21 [50], which were identified as downstream of elevated integrin signaling in our ECM-High matreotype. The physical proximity of aberrant basaloid cells covering myofibroblastic foci in IPF lungs [50] raises the possibility that crosstalk between neighboring IPF-like fibroblasts and transformed basal cells in SqCC contribute to the etiology of ECM-High matreotype tumors. A specific association of IPF with SqCC has been noted in the emergence of squamous metaplasia in fibrotic lung which can then develop into SqCC, as well as the higher incidence of SqCC than adenocarcinoma in IPF patients who develop NSCLC [135, 141–143]. In addition, lung cancer survival is poorer in IPF patients compared with the general population [142]. Our data indicates that a subset of poor-prognosis SqCC patients have hyperactivated fibrogenic signaling mimicking that seen in IPF. This also raises the notion that ECM remodeling, through aging or IPF mechanisms, may contribute to a field of cancerization that primes the lung tissue for accelerated tumorigenesis [144, 145]. These data also suggests that these patients may benefit from anti-stromal therapies targeting these pathways to disrupt the cycle of ECM deposition, fibrogenesis, and the corollary effects on tumor progression.

Anti-stromal therapies are showing considerable promise in the treatment of highly desmoplastic tumors such as pancreatic ductal adenocarcinoma [146], a solid tumor type in which our ECM-High matreotype is also prognostic. The anti-stromal IPF treatment nintedanib, which targets FGF, VEGF, and PDGF signaling, has been approved as a second-line therapy in lung adenocarcinoma as it has demonstrated greater clinical benefit in adenocarcinoma than in the SqCC histological subtype. However, nintedanib has shown some *in vitro* efficacy in SqCC models albeit lower than in adenocarcinoma models [147] and modest, but significant improvements in progression-free survival and disease control in human

SqCC [40, 148]. Our data suggests that correct patient selection would be critical since, unlike the ECM-Low SqCC, the ECM-High subset of SqCC would be the subset most likely to benefit from this treatment. Therefore, with further refinement, matrisomal profiling of SqCC may be used as a companion biomarker for anti-stromal efficacy, and matrisomal subtyping should be considered as an inclusion criterion in the design of future trials. Furthermore, our data suggests that ongoing clinical trials targeting pro-fibrogenic signaling in FGFR-amplified or -mutant SqCC, including those ongoing trials for AZD4547 (FGFR1,2,3; NCT01824901, NCT02154490), lucitanib (FGFR1,2; VEGFR1,2,3, CSF1; NCT012109016), BAY1163877 (NCT01976741), the pan-TKI dovitinib (FGFR + VEGFR, NCT01861197), as well as the use of the approved agent Nintedanib (VEGFR, FGF, PDGFR) [4] should consider using matrisomal subtyping, rather than FGF status alone as an inclusion criteria.

Our data predict that the core matrisome elements within FGFR-wildtype ECM-High matrisome tumors can signal through integrins, FLT4, and PDGFRB on fibroblasts to induce pro-fibrogenic signaling that promotes cancer cell growth, migration, and invasion. Therefore, FGF-wildtype ECM-High tumors may also benefit from these anti-stromal therapies. Furthermore, the association of our ECM-High matrisome with cisplatin resistance, potentially via integrin-mediated induction of cisplatin resistance mediator p21 [108, 109], also suggests that a co-targeting approach of anti-stromal therapies together with standard-of-care platinum doublet therapies may synergize to augment and potentially restore cisplatin sensitivity. Together, these findings indicate that matrisotyping SqCC tumors may be useful in identifying those patients with ECM-High tumors who are likely to benefit from stromal co-targeting therapies.

While this study has yielded clinically actionable insights from whole-matrisomal profiling of tumors, it has limitations. The penalized regression modeling approach used to identify the key core matrisomal components defining SqCC implemented parameters that minimized the model error while incorporating a reasonable number of matrisomal genes. However, in balancing model complexity and accuracy, additional matrisomal genes that were less predictive were not included in the final model, yet these may still have some biological or clinical importance in SqCC tumorigenesis. This is particularly the case for highly correlated core matrisomal components, where one ECM component included in the model may capture the predictive power of multiple co-linear ECM components. In addition, our bioinformatic approaches utilized separate tissue sources for the scRNAseq, where

SqCC tumor data is particularly limited, and the bulk transcriptomic/proteomic data (RNAseq, whole exome sequencing and RPPA proteomic), making it difficult to discern the precise cell of origin for ECM components. Furthermore, from the analysis of multi-omics data, it is not possible to discern the mechanistic contributions of the ECM component itself, from the exact cell type that produces it, when examining associations with patient outcome. This is in line with recent studies showing that core matrisomal gene expression signatures can robustly recapitulate specific cellular phenotypes [41]. Key ECM components in SqCC risk and prognosis identified in this study have been shown to be expressed by specific subtypes of stromal cells as well as cancer cells and so it is possible that clinical associations identified for matrisomal genes reflect the effects of these cell types more broadly. For example, cancer-associated fibroblasts have also been associated with lung cancer prognosis in adenocarcinoma [116, 149] and NSCLC generally [105, 117, 150], although their significance in SqCC prognosis remains unclear. This *in silico* analysis will require experimental interrogation to dissect the function of key ECM remodeling features. In particular, functional studies will be required to validate the mechanistic roles of individual matrix components in SqCC prognosis as distinct from effects of cellular subtypes that simply express these matrix components. For example, murine models in which cancer cells are orthotopically injected together with cancer-associated fibroblasts expressing high and low levels of key ECM genes will be fundamental first steps to establishing the functional importance of these ECM components in accelerating tumor progression and metastatic dissemination. Further *in vitro* studies using cell co-cultures in organotypic matrices will also be critical to dissect signaling mechanisms driving treatment response and metastasis.

Conclusions

This comprehensive analysis of the SqCC matrisomal landscape has defined key matrisomal components associated with an increased risk of developing lung cancer and of developing aggressive treatment-refractory disease. Through combining multi-omics datasets, we hypothesize that enriched elements of the core matrisome amplify fibrogenesis and activate intracellular signaling networks that support metastatic dissemination. Tumor ECM subtyping has revealed a subset of SqCC patients with the worst prognosis whose tumors are likely to respond well to treatment with existing repurposed anti-stromal therapies in combination with standard-of-care therapy. These findings highlight the importance of considering the ECM

profile of SqCC as part of a precision medicine framework to improve the outcome of SqCC patients.

Abbreviations

NSCLC: Non-small cell lung cancer; SqCC: Squamous NSCLC.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-022-01127-6>.

Additional file 1: Supplementary Materials and Methods. Figure S1.

The extracellular matrix is significantly dysregulated in tumor compared with non-tumor tissue in SqCC. **Figure S2.** ECM changes associated with increased lung cancer risk and premalignant progression. **Figure S3.** A) tSNE plot visualization of the expression scores for matrix risk signature genes with positive (B) and negative (C) odds ratios for different cell types (A) in SqCC scRNAseq data presented in Fig. 2D. D) ROC analysis of the minimum matrix risk signature distinguishing progressive from regressive premalignant lesions. E) Matrix risk score at age at diagnosis in the TCGA cohort. Blue line shows linear regression with standard error indicated by grey shading. $p=0.00073$, $r=-0.23$, Spearman's correlation. **Figure S4.** The ECM-High matreotype is associated with poor prognosis. A-B) Relative Cluster Stability Index (A) and p-values (B) for different cluster numbers confirm the presence of three major matreotypes in SqCC. C) The correlation plot for samples corresponding to the heatmap in Fig. 3A. D) Heatmap of marker genes for the ECM-High and ECM-Low matreotypes in the TCGA cohort. E-F) Matreotype association with survival in early stage (E, Stage I and II) patients (Ci) and late stage patients (F, Stage III and IV) patients in the TCGA LUSC cohort. G) Recurrence-free survival for ECM-High and ECM-Low matreotypes in the UHN cohort of early-stage tumors (log-rank $p=0.19$). H) Representative H&E and picrosirius red-stained tissue microarray cores corresponding to high (upper panel) and low (lower panel) picrosirius-red-stained tumors. Scale bar = 500 μm . I-J) Overall survival of patients according to picrosirius red staining for tumors across all stages (I, univariate cox-proportional hazards model HR = 1.76 [1.04-2.98], $p=0.035$; multivariate coxproportional hazards model HR = 1.87 [1.09-3.2], $p=0.023$) and stages I and IIA only (J, univariate cox-proportional hazards model HR = 2.4 [1.3-4.4], $p=0.0051$; multivariate cox-proportional hazards model HR = 2.4 [1.3-4.4], $p=0.0051$). Multivariate cox-proportional hazards models include stage as a clinical covariate. **Figure S5.** A) Mutational frequency plot showing no enrichment of driver mutations or FGFR amplification in each matreotype. B) Mutational frequency plot showing the top mutated genes differentially enriched in the ECM-High vs ECM-Low matreotype tumors. C) Disease-specific survival of adenocarcinoma samples assigned to SqCC matreotypes shows no significant association of SqCC matreotypes with prognosis in the adenocarcinoma subtype (log-rank $p=0.74$). D-E) The hazard ratios for Disease-specific survival (D) and Progression-Free Survival (E) of ECM-High pan-cancer tumors compared with ECM-Low matreotypes in multiple cancer types. Red dots indicate tumor types with significant hazard ratios for the ECM-High matreotype compared with the ECM-Low matreotype. Hazard ratios and confidence intervals are in Additional File 1: Table S8. **Figure S6.** SqCC Matreotypes have distinct immunological ecosystems. **Figure S7.** Extracellular Matrix components contribute to signaling pathways associated with prognosis. **Figure S8.** ECM-driven integrin signaling is associated with EMT and fibroblast activation in the ECM-High matreotype. **Figure S9.** The poor prognosis matreotype overlaps with ECM remodeling in Idiopathic Pulmonary Fibrosis. **Table S1.** Matrix Risk Score. **Table S2.** Pan-Cancer Risk Score Results. **Table S3.** Optimized Minimum Matrisomal Linear Model for Premalignant Progression. **Table S4.** Clinicodemographic Features of the TCGA Matreotypes. **Table S5.** Clinicodemographic Features of the NCI-MD cohort. **Table S6.** Clinicodemographic Features of the Tissue Microarray Cohort. **Table S7.** Cox Proportional Hazards Model Analysis of Picrosirius Red stained TMAs. **Table S8.** Pan-Cancer Prognostic Matreotype Results.

Additional file 2: Table S1. Differentially Expressed Hallmark Pathways in ECM-High vs ECM-Low Tumors. **Table S2.** Differentially Expressed MSigDb

C2 Pathways in ECM-High vs ECM-Low Tumors. **Table S3.** Ligand-Receptor Interactions in ECM-High vs ECM-Low Tumors.

Acknowledgements

The authors gratefully thank Drs. Bernhard Thienpont, Dietmar Lambrechts, and Bram Boeckx (KULeuven) for providing scRNAseq data of lung cancer tumors. This work was supported by Lung Foundation Australia Evangeline Lim Grant-in-Aid (ALP), Cancer Council NSW project (RG19-01 to MKC, RG19-09 to TRC), Cancer Institute NSW Career Development Fellowship (CDF171105 to TRC), National Health and Medical Research Council fellowship (1158590 to TRC), National Health and Medical Research Council project funding (1129766, 1140125, 2000937 to TRC), and Susan G Komen for the Cure (CCR17483294 to TRC) funding. The authors thank staff at the Garvan Histology Facility for expert assistance with tissue microarray sectioning and staining. The authors thank Drs Elyse Filipe and Walter Muskovic of the Garvan Institute for Medical Research for critical reading of the manuscript.

Authors' contributions

ALP: Conceptualization, Methodology, Investigation, Writing—Original Draft, Writing—Review and Editing, Visualization; EB: Resources, Data Curation; AZ: Resources; WAC: Resources, Data Curation; MKC: Resources, Data Curation; BMR: Resources; CCH: Resources, Project Administration; TRC: Conceptualization, Writing—Review and Editing, Supervision, Project Administration, Funding Acquisition. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Lung Foundation Australia Evangeline Lim Grant-in-Aid (ALP), Cancer Council NSW project (RG19-01 to MKC, RG19-09 to TRC), Cancer Institute NSW Career Development Fellowship (CDF171105 to TRC), National Health and Medical Research Council project and fellowship funding (1129766, 1140125, 2000937, and 1158590 to TRC), and Susan G Komen for the Cure (CCR17483294 to TRC) funding.

Availability of data and materials

The NCI-MD cohort RNAseq data analyzed in this study are accessible at the NCBI GEO website under the accession number GSE201221 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE201221>) [71]. Publicly available TCGA LUSC and LUAD RNAseq, WES and CNV data are available from Broad GDAC Firehose (<https://gdac.broadinstitute.org>). Publicly available Pan-Cancer RNAseq and survival data is available from Genomic Data Commons repository (<https://gdc.cancer.gov/about-data/publications/pancanatlas>). Publicly available TCGA RPPA data is available from the TCGA Portal (<https://tcppa.ortol.org/tcpa/download.html>). Publicly available data from the UHN cohort is available in the Gene Expression Omnibus repository, GSE50081 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50081>). Publicly available RNAseq of squamous carcinoma in situ is available in the Gene Expression Omnibus repository, GSE108124 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108124>). scRNAseq data are available from <https://lambrechtslab.sites.vib.be/en/data-access>. RNAseq data of the aging lung is available in the Gene Expression Omnibus repository, GSE165192 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165192>).

Declarations

Ethics approval and consent to participate

Publicly available data was obtained as described (Table 1) and analyzed in accordance with the respective guidelines of each platform. NSCLC tumor and non-tumor biopsies from the NCI-MD cohort were collected according to approved procedures as described previously [52] (National Cancer Institute, USA, IRB OH98-C-N027). Written informed consent was received prior to participation in these studies. This research conformed to the principles of the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

Brid M. Ryan had no competing interests while involved with this research and is now a paid employee of Mina Therapeutics. The remaining authors declare that they have no competing interests.

Author details

¹Matrix and Metastasis Lab, Cancer Ecosystems Program, Garvan Institute of Medical Research and The Kinghorn Cancer Centre, 384 Victoria St, Darlinghurst, NSW 2052, Australia. ²School of Clinical Medicine, UNSW Sydney, Sydney 2052, Australia. ³Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA. ⁴Present address: MiNA Therapeutics, London, UK. ⁵Department of Tissue Pathology and Diagnostic Oncology, NSW Health Pathology, Royal Prince Alfred Hospital, Camperdown, NSW 2050, Australia. ⁶Sydney Medical School, University of Sydney, Sydney, NSW 2050, Australia. ⁷Discipline of Pathology, School of Medicine, Western Sydney University, Liverpool, NSW 2170, Australia. ⁸Woolcock Institute of Medical Research, Sydney, NSW 2037, Australia. ⁹Microbiome Research Centre, School of Clinical Medicine, UNSW Sydney, Sydney 2052, Australia. ¹⁰Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia.

Received: 21 April 2022 Accepted: 14 October 2022

Published online: 21 November 2022

References

- John T, Cooper WA, Wright G, Siva S, Solomon B, Marshall HM, et al. Lung cancer in Australia. *J Thorac Oncol*. 2020;15(12):1809–14.
- Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, et al. SEER Cancer Statistics Review (1975–2012). 2014.
- AIHW. Cancer data in Australia 2021. Available from: <https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia> [cited 26 Jan 2022].
- Thomas A, Liu SV, Subramaniam DS, Giaccone G. Refining the treatment of NSCLC according to histological and molecular subtypes. *Nat Rev Clin Oncol*. 2015;12(9):511–26.
- Jang KM, Lee KS, Shim YM, Han D, Kim H, Kwon OJ, et al. The rates and CT patterns of locoregional recurrence after resection surgery of lung cancer: correlation with histopathology and tumor staging. *J Thorac Imaging*. 2003;18(4):225–30.
- National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395–409.
- Teixeira VH, Pipinikas CP, Pennycook A, Lee-Six H, Chandrasekharan D, Beane J, et al. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. *Nat Med*. 2019;25(3):517–25.
- Pennycook A, Teixeira VH, Abduljabbar K, Raza SEA, Lund T, Akarca AU, et al. Immune surveillance in clinical regression of preinvasive squamous cell lung cancer. *Cancer Discov*. 2020;10(10):1489–99.
- Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med*. 2017;376(22):2109–21.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Banki AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415–21.
- Middleton G, Fletcher P, Popat S, Savage J, Summers Y, Greystoke A, et al. The National Lung Matrix Trial of personalized therapy in lung cancer. *Nature*. 2020;583(7818):807–12.
- Redman MW, Papadimitrakopoulou VA, Minichiello K, Hirsch FR, Mack PC, Schwartz LH, et al. Biomarker-driven therapies for previously treated squamous non-small-cell lung cancer (Lung-MAP SWOG S1400): a biomarker-driven master protocol. *Lancet Oncol*. 2020;21(12):1589–601.
- Hou P, Kapoor A, Zhang Q, Li J, Wu C-J, Li J, et al. Tumor microenvironment remodeling enables bypass of oncogenic KRAS dependency in pancreatic cancer. *Cancer Discov*. 2020;10:1058–77.
- Gocheva V, Naba A, Bhubkar A, Guardia T, Miller KM, Li CM-C, et al. Quantitative proteomics identify Tenascin-C as a promoter of lung cancer progression and contributor to a signature prognostic of patient survival. *Proc Natl Acad Sci U S A*. 2017;114(28):E5625–34.
- Qi L, Li T, Shi G, Wang J, Li X, Zhang S, et al. An individualized gene expression signature for prediction of lung adenocarcinoma metastases. *Mol Oncol*. 2017;11(11):1630–45.
- Wu J, Li L, Zhang H, Zhao Y, Zhang H, Wu S, et al. A risk model developed based on tumor microenvironment predicts overall survival and associates with tumor immunity of patients with lung adenocarcinoma. *Oncogene*. 2021;40(26):4413–24.
- Lim SB, Tan SJ, Lim W-T, Lim CT. An extracellular matrix-related prognostic and predictive indicator for early-stage non-small cell lung cancer. *Nat Commun*. 2017;8(1):1734.
- Zuo S, Wei M, Zhang H, Chen A, Wu J, Wei J, et al. A robust six-gene prognostic signature for prediction of both disease-free and overall survival in non-small cell lung cancer. *J Transl Med*. 2019;17(1):152.
- Boutros PC, Lau SK, Pintilie M, Liu N, Shepherd FA, Der SD, et al. Prognostic gene signatures for non-small-cell lung cancer. *Proc Natl Acad Sci U S A*. 2009;106(8):2824–8.
- Der SD, Sykes J, Pintilie M, Zhu C-Q, Strumpf D, Liu N, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J Thorac Oncol*. 2014;9(1):59–64.
- Raponi M, Zhang Y, Yu J, Chen G, Lee G, Taylor JMG, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res*. 2006;66(15):7466–72.
- Larsen JE, Pavey SJ, Passmore LH, Bowman R, Clarke BE, Hayward NK, et al. Expression profiling defines a recurrence signature in lung squamous cell carcinoma. *Carcinogenesis*. 2007;28(3):760–6.
- Aramburu A, Zudaire I, Pajares MJ, Agorreta J, Orta A, Lozano MD, et al. Combined clinical and genomic signatures for the prognosis of early stage non-small cell lung cancer based on gene copy number alterations. *BMC Genomics*. 2015;16:752.
- Wilkerson MD, Yin X, Hoadley KA, Liu Y, Hayward MC, Cabanski CR, et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin Cancer Res*. 2010;16(19):4864–75.
- Stewart PA, Welsh EA, Slebos RJC, Fang B, Izumi V, Chambers M, et al. Proteogenomic landscape of squamous cell lung cancer. *Nat Commun*. 2019;10(1):3578.
- Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet*. 2016;48(6):607–16.
- Lehtio J, Arslan T, Siavelis I, Pan Y, Socciarelli F, Berkovska O, et al. Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune evasion mechanisms. *Nat Cancer*. 2021;2(11):1224–42.
- Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J Clin Oncol*. 2006;24(31):5079–90.
- Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics*. 2012;11(4):M111.014647.
- Ragelle H, Naba A, Larson BL, Zhou F, Prijji M, Whittaker CA, et al. Comprehensive proteomic characterization of stem cell-derived extracellular matrices. *Biomaterials*. 2017;128:147–59.
- Cox TR. The matrix in cancer. *Nat Rev Cancer*. 2021;21(4):217–38.
- Naba A, Pearce OMT, Del Rosario A, Ma D, Ding H, Rajeeve V, et al. Characterization of the extracellular matrix of normal and diseased tissues using proteomics. *J Proteome Res*. 2017;16(8):3083–91.
- Naba A, Clauser KR, Lamar JM, Carr SA, Hynes RO. Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. *Elife*. 2014;3:e01308.
- Hebert JD, Myers SA, Naba A, Abbruzzese G, Lamar JM, Carr SA, et al. Proteomic profiling of the ECM of xenograft breast cancer metastases in different organs reveals distinct metastatic niches. *Cancer Res*. 2020;80(7):1475–85.
- Parker AL, Cox TR. The role of the ECM in lung cancer dormancy and outgrowth. *Front Oncol*. 2020;10:1766.
- Ratajczak-Wielgomas K, Kmiecik A, Grzegorzka J, Piotrowska A, Gomułkiewicz A, Partynska A, et al. Prognostic significance of stromal periostin expression in non-small cell lung cancer. *Int J Mol Sci*. 2020;21(19):7025.

37. Hong L-Z, Wei X-W, Chen J-F, Shi Y. Overexpression of periostin predicts poor prognosis in non-small cell lung cancer. *Oncol Lett.* 2013;6(6):1595–603.
38. Weng T-Y, Wang C-Y, Hung Y-H, Chen W-C, Chen Y-L, Lai M-D. Differential expression pattern of THBS1 and THBS2 in lung cancer: clinical outcome and a systematic-analysis of microarray databases. *PLoS One.* 2016;11(8):e0161007.
39. Hanna NH, Kaiser R, Sullivan RN, Aren OR, Ahn M-J, Tiangco B, et al. Nintedanib plus pemetrexed versus placebo plus pemetrexed in patients with relapsed or refractory, advanced non-small cell lung cancer (LUME-Lung 2): A randomized, double-blind, phase III trial. *Lung Cancer.* 2016;102:65–73.
40. Forster M, Hackshaw A, De Pas T, Cobo M, Garrido P, Summers Y, et al. A phase I study of nintedanib combined with cisplatin/gemcitabine as first-line therapy for advanced squamous non-small cell lung cancer (LUME-Lung 3). *Lung Cancer.* 2018;120:27–33.
41. Sacher F, Feregrino C, Tschopp P, Ewald CY. Extracellular matrix gene expression signatures as cell type and cell state identifiers. *Matrix Biol Plus.* 2021;10:100069.
42. Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, et al. TCPA: a resource for cancer functional proteomics data. *Nat Methods.* 2013;10(11):1046–7.
43. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012;489(7417):519–25.
44. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014;511(7511):543–50.
45. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell.* 2018;173(2):400–16 e11.
46. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med.* 2018;24(8):1277–89.
47. Angelidis I, Simon LM, Fernandez IE, Strunz M, Mayr CH, Greiffo FR, et al. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nat Commun.* 2019;10(1):963.
48. McDonough JE, Ahangari F, Li Q, Jain S, Verleden SE, Herazo-Maya J, et al. Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight.* 2019;4(22):e131597.
49. Tsukui T, Sun K-H, Wetter JB, Wilson-Kanamori JR, Hazelwood LA, Henderson NC, et al. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat Commun.* 2020;11(1):1920.
50. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv.* 2020;6(28):eaba1983.
51. Faruki H, Mayhew GM, Serody JS, Hayes DN, Perou CM, Lai-Goldman M. Lung adenocarcinoma and squamous cell carcinoma gene expression subtypes demonstrate significant differences in tumor immune landscape. *J Thorac Oncol.* 2017;12(6):943–53.
52. Mathé EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, Bowman ED, et al. Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Res.* 2014;74(12):3259–70.
53. Parker AL, Toulabi L, Oike T, Kanke Y, Patel D, Tada T, et al. Creatine riboside is a cancer cell-derived metabolite associated with arginine auxotrophy. *J Clin Invest.* 2022;132(14):e157410.
54. Selinger CI, Cooper WA, Al-Sohaily S, Mladenova DN, Pangon L, Kennedy CW, et al. Loss of special AT-rich binding protein 1 expression is a marker of poor survival in lung cancer. *J Thorac Oncol.* 2011;6(7):1179–89.
55. Chaisaingmongkol J, Budhu A, Dang H, Rabibhadana S, Pupacdi B, Kwon SM, et al. Common molecular subtypes among asian hepatocellular carcinoma and cholangiocarcinoma. *Cancer Cell.* 2017;32(1):57–70. e3.
56. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
57. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012;28(6):882–3.
58. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 2012;40(10):4288–97.
59. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
60. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847–9.
61. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
62. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1–22.
63. John CR, Watson D, Russ D, Goldmann K, Ehrenstein M, Pitzalis C, et al. M3C: Monte Carlo reference-based consensus clustering. *Sci Rep.* 2020;10(1):1816.
64. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
65. Ramilowski JA, Goldberg T, Harshbarger J, Kloppmann E, Lizio M, Sata-gopam VP, et al. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun.* 2015;6:7866.
66. Wu SZ, Roden DL, Wang C, Holliday H, Harvey K, Cazet AS, et al. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J.* 2020;39(19):e104063.
67. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. *Bioinformatics.* 2014;30(19):2811–2.
68. Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etlloglu E, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* 2020;30(9):745–62.
69. Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.* 2019;47(W1):W212–24.
70. Vennin C, Chin VT, Warren SC, Lucas MC, Herrmann D, Magenau A, et al. Transient tissue priming via ROCK inhibition uncouples pancreatic cancer progression, sensitivity to chemotherapy, and metastasis. *Sci Transl Med.* 2017;9(384):eaai8504.
71. Parker AL, Khan M, Harris CC. Association of gene expression with creatine riboside, a prognostic metabolite of metabolome imbalance in the NCI-MD Non-Small Cell Lung Cancer Cohort. *GSE201221*, NCBI Gene Expression Omnibus. 2022. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE201221>. Accessed July 2018.
72. Sund M, Maeshima Y, Kalluri R. Bifunctional promoter of type IV collagen COL4A5 and COL4A6 genes regulates the expression of alpha5 and alpha6 chains in a distinct cell-specific fashion. *Biochem J.* 2005;387(Pt 3):755–61.
73. Han L, Xu J, Grigg E, Slack M, Chaturvedi P, Jiang R, et al. Osr1 functions downstream of Hedgehog pathway to regulate foregut development. *Dev Biol.* 2017;427(1):72–83.
74. Hargadon KM, Györfy B, Strong EW. The prognostic significance of FOXC2 gene expression in cancer: a comprehensive analysis of RNA-seq data from the cancer genome atlas. *Cancer Gene Ther.* 2021;254–255:58–64.
75. Wang T, Zheng L, Wang Q, Hu Y-W. Emerging roles and mechanisms of FOXC2 in cancer. *Clin Chim Acta.* 2018;479:84–93.
76. Hu Y-T, Li B-F, Zhang P-J, Wu D, Li Y-Y, Li Z-W, et al. Dlx2 exhibits a tumor-promoting function in hepatocellular carcinoma cell lines via regulating Shh-Gli1 signaling. *World J Gastroenterol.* 2019;25(8):923–40.
77. Brena RM, Morrison C, Liyanarachchi S, Jarjoura D, Davuluri RV, Otterson GA, et al. Aberrant DNA methylation of OLIG1, a novel prognostic factor in non-small cell lung cancer. *PLoS Med.* 2007;4(3):e108.
78. Tammemägi MC, Ten Haaf K, Toumazis I, Kong CY, Han SS, Jeon J, et al. Development and validation of a multivariable lung cancer risk prediction model that includes low-dose computed tomography screening results: a secondary analysis of data from the national lung screening trial. *JAMA Netw Open.* 2019;2(3):e190204.
79. Zhang Y, Du W, Chen Z, Xiang C. Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. *Exp Cell Res.* 2017;359(2):449–57.

80. Chiou J, Chang Y-C, Tsai H-F, Lin Y-F, Huang M-S, Yang C-J, et al. Follistatin-like protein 1 inhibits lung cancer metastasis by preventing proteolytic activation of osteopontin. *Cancer Res.* 2019;79(24):6113–25.
81. Hu Z, Lin D, Yuan J, Xiao T, Zhang H, Sun W, et al. Overexpression of osteopontin is associated with more aggressive phenotypes in human non-small cell lung cancer. *Clin Cancer Res.* 2005;11(13):4646–52.
82. Kang CG, Han HJ, Lee H-J, Kim S-H, Lee E-O. Rho-associated kinase signaling is required for osteopontin-induced cell invasion through inactivating cofilin in human non-small cell lung cancer cell lines. *Bioorg Med Chem Lett.* 2015;25(9):1956–60.
83. Boldrini L, Donati V, Dell'Omodarme M, Prati MC, Faviana P, Camacci T, et al. Prognostic significance of osteopontin expression in early-stage non-small-cell lung cancer. *Br J Cancer.* 2005;93(4):453–7.
84. Shen L, Yang M, Lin Q, Zhang Z, Zhu B, Miao C. COL11A1 is overexpressed in recurrent non-small cell lung cancer and promotes cell proliferation, migration, invasion and drug resistance. *Oncol Rep.* 2016;36(2):877–85.
85. Wu Y-H, Huang Y-F, Chang T-H, Chen C-C, Wu P-Y, Huang S-C, et al. COL11A1 activates cancer-associated fibroblasts by modulating TGF- β 3 through the NF- κ B/IGFBP2 axis in ovarian cancer cells. *Oncogene.* 2021;40(26):4503–19.
86. Wu YH, Chang TH, Huang YF, Huang HD, Chou CY. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene.* 2014;33(26):3432–40.
87. Tu H, Li J, Lin L, Wang L. COL11A1 was involved in cell proliferation, apoptosis and migration in non-small cell lung cancer cells. *J Investig Surg.* 2021;34(6):664–9.
88. Kita Y, Mimori K, Tanaka F, Matsumoto T, Haraguchi N, Ishikawa K, et al. Clinical significance of LAMB3 and COL7A1 mRNA in esophageal squamous cell carcinoma. *Eur J Surg Oncol.* 2009;35(1):52–8.
89. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A comprehensive pan-cancer molecular study of gynecologic and breast cancers. *Cancer Cell.* 2018;33(4):690–705 e9.
90. Tammemägi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, et al. Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. *PLoS Med.* 2014;11(12):e1001764.
91. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7–30.
92. Lee S, Islam MN, Boostanpour K, Aran D, Jin G, Christenson S, et al. Molecular programs of fibrotic change in aging human lung. *Nat Commun.* 2021;12(1):6309.
93. Ecker BL, Kaur A, Douglass SM, Webster MR, Almeida FV, Marino GE, et al. Age-related changes in HAPLN1 increase lymphatic permeability and affect routes of melanoma metastasis. *Cancer Discov.* 2019;9(1):82–95.
94. Kaur A, Webster MR, Marchbank K, Behera R, Ndoye A, Kugel CH, et al. sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance. *Nature.* 2016;532(7598):250–4.
95. Kaur A, Ecker BL, Douglass SM, Kugel CH, Webster MR, Almeida FV, et al. Remodeling of the collagen matrix in aging skin promotes melanoma metastasis and affects immune cell motility. *Cancer Discov.* 2019;9(1):64–81.
96. Rittié L. Method for picosirius red-polarization detection of collagen fibers in tissue sections. *Methods Mol Biol.* 2017;1627:395–407.
97. Piersma B, Hayward MK, Weaver VM. Fibrosis and cancer: a strained relationship. *Biochim Biophys Acta Rev Cancer.* 2020;1873(2):188356.
98. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* 2019;37(7):773–82.
99. Salmon H, Franciszkiewicz K, Damotte D, Dieu-Nosjean M-C, Validire P, Trautmann A, et al. Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J Clin Invest.* 2012;122(3):899–910.
100. Zhou J, Kang Y, Chen L, Wang H, Liu J, Zeng S, et al. The drug-resistance mechanisms of five platinum-based antitumor agents. *Front Pharmacol.* 2020;11:343.
101. Mo D, Fang H, Niu K, Liu J, Wu M, Li S, et al. Human helicase RECQL4 drives cisplatin resistance in gastric cancer by activating an AKT-YB1-MDR1 signaling pathway. *Cancer Res.* 2016;76(10):3057–66.
102. Xu J, Hu Z. Y-box-binding protein 1 promotes tumor progression and inhibits cisplatin chemosensitivity in esophageal squamous cell carcinoma. *Biomed Pharmacother.* 2016;79:17–22.
103. Pilié PG, Tang C, Mills GB, Yap TA. State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat Rev Clin Oncol.* 2019;16(2):81–104.
104. Navab R, Strumpf D, To C, Pasko E, Kim KS, Park CJ, et al. Integrin α 11 β regulates cancer stromal stiffness and promotes tumorigenicity and metastasis in non-small cell lung cancer. *Oncogene.* 2016;35(15):1899–908.
105. Navab R, Strumpf D, Bandarchi B, Zhu C-Q, Pintilie M, Ramnarine VR, 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):7160–5.
106. Cooper J, Giancotti FG. Integrin signaling in cancer: mechanotransduction, stemness, epithelial plasticity, and therapeutic resistance. *Cancer Cell.* 2019;35(3):347–67.
107. Robertson J, Jacquemet G, Byron A, Jones MC, Warwood S, Selley JN, et al. Defining the phospho-adhesome through the phosphoproteomic analysis of integrin signalling. *Nat Commun.* 2015;6:6265.
108. Zamagni A, Pasini A, Pirini F, Ravaoli S, Giordano E, Tesei A, et al. CDKN1A upregulation and cisplatin-pemetrexed resistance in non-small cell lung cancer cells. *Int J Oncol.* 2020;56(6):1574–84.
109. Xia X, Ma Q, Li X, Ji T, Chen P, Xu H, et al. Cytoplasmic p21 is a potential predictor for cisplatin sensitivity in ovarian cancer. *BMC Cancer.* 2011;11:399.
110. Sackmann E. How actin/myosin crosstalks guide the adhesion, locomotion and polarization of cells. *Biochim Biophys Acta.* 2015;1853(11 Pt B):3132–42.
111. Rodriguez-Hernandez I, Cantelli G, Bruce F, Sanz-Moreno V. Rho, ROCK and actomyosin contractility in metastasis as drug targets. *F1000Res.* 2016;5:783.
112. Sanz-Moreno V, Gaggioli C, Yeo M, Albregues J, Wallberg F, Virois A, et al. ROCK and JAK1 signaling cooperate to control actomyosin contractility in tumor cells and stroma. *Cancer Cell.* 2011;20(2):229–45.
113. Iwai M, Tulafu M, Togo S, Kawaji H, Kadoya K, Namba Y, et al. Cancer-associated fibroblast migration in non-small cell lung cancers is modulated by increased integrin α 11 expression. *Mol Oncol.* 2021;15(5):1507–27.
114. Puig M, Lugo R, Gabasa M, Giménez A, Velásquez A, Galgoczy R, et al. Matrix stiffening and β 1 integrin drive subtype-specific fibroblast accumulation in lung cancer. *Mol Cancer Res.* 2015;13(1):161–73.
115. Zhu C-Q, Popova SN, Brown ERS, Barsyte-Lovejoy D, Navab R, Shih W, et al. Integrin α 11 regulates IGF2 expression in fibroblasts to enhance tumorigenicity of human non-small-cell lung cancer cells. *Proc Natl Acad Sci U S A.* 2007;104(28):11754–9.
116. Hegab AE, Ozaki M, Kameyama N, Gao J, Kagawa S, Yasuda H, et al. Effect of FGF/FGFR pathway blocking on lung adenocarcinoma and its cancer-associated fibroblasts. *J Pathol.* 2019;249(2):193–205.
117. Chen Y, Zou L, Zhang Y, Chen Y, Xing P, Yang W, et al. Transforming growth factor- β 1 and α -smooth muscle actin in stromal fibroblasts are associated with a poor prognosis in patients with clinical stage I-IIIa nonsmall cell lung cancer after curative resection. *Tumour Biol.* 2014;35(7):6707–13.
118. Bremnes RM, Dønnem T, Al-Saad S, Al-Shibli K, Andersen S, Sirera R, et al. The role of tumor stroma in cancer progression and prognosis: emphasis on carcinoma-associated fibroblasts and non-small cell lung cancer. *J Thorac Oncol.* 2011;6(1):209–17.
119. Wollin L, Wex E, Pautsch A, Schnapp G, Hostettler KE, Stowasser S, et al. Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. *Eur Respir J.* 2015;45(5):1434–45.
120. Yoo H, Jeong B-H, Chung MJ, Lee KS, Kwon OJ, Chung MP. Risk factors and clinical characteristics of lung cancer in idiopathic pulmonary fibrosis: a retrospective cohort study. *BMC Pulm Med.* 2019;19(1):149.
121. Ballester B, Milara J, Cortijo J. Idiopathic pulmonary fibrosis and lung cancer: mechanisms and molecular targets. *Int J Mol Sci.* 2019;20(3):593.
122. Wu H, Yu Y, Huang H, Hu Y, Fu S, Wang Z, et al. Progressive pulmonary fibrosis is caused by elevated mechanical tension on alveolar stem cells. *Cell.* 2020;180(1):107–21 e17.

123. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv.* 2020;6:eaba1983.
124. Sun M, Luo EY, Adams SM, Adams T, Ye Y, Shetye SS, et al. Collagen XI regulates the acquisition of collagen fibril structure, organization and functional properties in tendon. *Matrix Biol.* 2020;94:77–94.
125. Andreuzzi E, Fejza A, Capuano A, Poletto E, Pivetta E, Doliana R, et al. Deregulated expression of Elastin Microfibril Interfacier 2 (EMILIN2) in gastric cancer affects tumor growth and angiogenesis. *Matrix Biol Plus.* 2020;6–7:100029.
126. Leader AM, Grout JA, Maier BB, Nabet BY, Park MD, Tabachnikova A, et al. Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. *Cancer Cell.* 2021;39(12):1594–609 e12.
127. Patz EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial. *Lancet Oncol.* 2016;17(5):590–9.
128. Diederich S, Wormanns D, Semik M, Thomas M, Lenzen H, Roos N, et al. Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. *Radiology.* 2002;222(3):773–81.
129. Nicholson AG, Perry LJ, Cury PM, Jackson P, McCormick CM, Corrin B, et al. Reproducibility of the WHO/IASLC grading system for pre-invasive squamous lesions of the bronchus: a study of inter-observer and intra-observer variation. *Histopathology.* 2001;38(3):202–8.
130. Pruitt HC, Lewis D, Ciccaglione M, Connor S, Smith Q, Hickey JW, et al. Collagen fiber structure guides 3D motility of cytotoxic T lymphocytes. *Matrix Biol.* 2020;85–86:147–59.
131. Sun X, Wu B, Chiang HC, Deng H, Zhang X, Xiong W, et al. Tumour DDR1 promotes collagen fibre alignment to instigate immune exclusion. *Nature.* 2021;599(7886):673–8.
132. Grout JA, Sirven P, Leader AM, Maskey S, Hector E, Puisieux I, et al. Spatial positioning and matrix programs of cancer-associated fibroblasts promote T cell exclusion in human lung tumors. *Cancer Discov.* 2022;12:1–20.
133. Takahashi Y, Ishii G, Taira T, Fujii S, Yanagi S, Hishida T, et al. Fibrous stroma is associated with poorer prognosis in lung squamous cell carcinoma patients. *J Thorac Oncol.* 2011;6(9):1460–7.
134. Lonergan KM, Chari R, Coe BP, Wilson IM, Tsao M-S, Ng RT, et al. Transcriptome profiles of carcinoma-in-situ and invasive non-small cell lung cancer as revealed by SAGE. *PLoS One.* 2010;5(2):e9162.
135. Maeda R, Funasaki A, Motono N, Sekimura A, Usuda K, Uramoto H. Combined pulmonary fibrosis and emphysema predicts recurrence following surgery in patients with stage I non-small cell lung cancer. *Med Oncol.* 2018;35(3):31.
136. Lee T, Park JY, Lee HY, Cho Y-J, Yoon HI, Lee JH, et al. Lung cancer in patients with idiopathic pulmonary fibrosis: clinical characteristics and impact on survival. *Respir Med.* 2014;108(10):1549–55.
137. Alcaraz J, Carrasco JL, Millares L, Luis I-C, Fernández-Porras FJ, Martínez-Romero A, et al. Stromal markers of activated tumor associated fibroblasts predict poor survival and are associated with necrosis in non-small cell lung cancer. *Lung Cancer.* 2019;135:151–60.
138. Ferone G, Song J-Y, Sutherland KD, Bhaskaran R, Monkhorst K, Lambooi J-P, et al. SOX2 is the determining oncogenic switch in promoting lung squamous cell carcinoma from different cells of origin. *Cancer Cell.* 2016;30(4):519–32.
139. Burgstaller G, Oehrlé B, Gerckens M, White ES, Schiller HB, Eickelberg O. The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. *Eur Respir J.* 2017;50(1):1601805.
140. Han X, Li F, Fang Z, Gao Y, Li F, Fang R, et al. Transdifferentiation of lung adenocarcinoma in mice with Lkb1 deficiency to squamous cell carcinoma. *Nat Commun.* 2014;5:3261.
141. Gao L, Xie S, Liu H, Liu P, Xiong Y, Da J, et al. Lung cancer in patients with combined pulmonary fibrosis and emphysema revisited with the 2015 World Health Organization classification of lung tumors. *Clin Respir J.* 2018;12(2):652–8.
142. Whittaker Brown S-A, Padilla M, Mhango G, Taioli E, Powell C, Wisnivesky J. Outcomes of older patients with pulmonary fibrosis and non-small cell lung cancer. *Ann Am Thorac Soc.* 2019;16(8):1034–40.
143. Nezka H, Igor P, Izidor K. Idiopathic pulmonary fibrosis in patients with early-stage non-small-cell lung cancer after surgical resection. *Radiol Oncol.* 2019;53(3):357–61.
144. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med.* 2007;13(3):361–6.
145. Kadara H, Sivakumar S, Jakubek Y, San Lucas FA, Lang W, McDowell T, et al. Driver mutations in normal airway epithelium elucidate spatiotemporal resolution of lung cancer. *Am J Respir Crit Care Med.* 2019;200(6):742–50.
146. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer.* 2020;20(3):174–86.
147. Ikemori R, Gabasa M, Duch P, Vizoso M, Bragado P, Arshakyan M, 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):276–90.
148. Reck M, Kaiser R, Mellemegaard A, Douillard J-Y, Orlov S, Krzakowski M, et al. Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial. *Lancet Oncol.* 2014;15(2):143–55.
149. Schliekelman MJ, Creighton CJ, Baird BN, Chen Y, Banerjee P, Bota-Rabasedas N, et al. Thy-1+ cancer-associated fibroblasts adversely impact lung cancer prognosis. *Sci Rep.* 2017;7(1):6478.
150. Kilvaer TK, Khanekhenari MR, Hellevik T, Al-Saad S, Paulsen E-E, Bremnes RM, et al. Cancer associated fibroblasts in stage I-IIIa NSCLC: prognostic impact and their correlations with tumor molecular markers. *PLoS One.* 2015;10(8):e0134965.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

