

## ORIGINAL RESEARCH

# A five-gene signature is a prognostic biomarker in pan-cancer and related with immunologically associated extracellular matrix

Chunlai Yu<sup>1</sup> | Mingliang You<sup>4</sup> | Peizhen Zhang<sup>3</sup> | Sheng Zhang<sup>1</sup> | Yuzhu Yin<sup>3</sup> | Xiao Zhang<sup>1,2</sup> 

<sup>1</sup>Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou, China

<sup>2</sup>CAS Key Laboratory of Regenerative Biology, Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, and Guangzhou Medical University, Guangzhou, Guangdong, China

<sup>3</sup>Department of Obstetrics and Gynecology, Third Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China

<sup>4</sup>Key Laboratory of Clinical Cancer Pharmacology and Toxicology Research of Zhejiang Province, Hangzhou Cancer Institute, Affiliated Hangzhou Cancer Hospital, Zhejiang University School of Medicine, Hangzhou, China

## Correspondence

Xiao Zhang, Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), 510005 Guangzhou, China.  
Email: zhang\_xiao@gibh.ac.cn

Yuzhu Yin, Department of Obstetrics and Gynecology, Third Affiliated Hospital, Sun Yat-Sen University, Guangzhou 510630, China.  
Email: yinyuzhu@mail.sysu.edu.cn

## Funding information

This work was funded by the Scientific Instrumentation Development Program of Chinese Academy of Sciences (ZDZBGCH2018005) and the Key Research & Development Program of Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory) (2019GZR1104060).

## Abstract

The tumor microenvironment (TME) is related to extracellular matrix (ECM) dynamics and has a broad fundamental and mechanistic role in tumorigenesis and cancer progression. We hypothesized that ECM regulators might play an essential role in pan-cancer attribution by causing a generic effect through its regulation of the dynamics of ECM alteration. By analyzing data from TCGA using GSEA and univariate Cox regression analysis, we found that ECM regulator genes were significantly enriched and contributed to mortality in various cancer types. Notably, UMAP analysis revealed that ECM regulator genes dominated the differences between tumor and adjacent normal tissues based on 59 or 31 pan-survival-related ECM gene sets. Subsequently, a five-gene signature consisting of the predominant ECM regulators *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* was identified. We found that this five-gene signature was pro-mortality in 18 types of cancer in TCGA, and validated eleven other cancer types in TCGA and seven types in the TARGET and CoMMpass databases using overall survival analysis. KEGG pathway enrichment and Pearson correlation analysis indicated that these five component genes that were correlated with specific ECM proteins involved in tumorigenesis from the ECM receptor interaction gene set. Additionally, the fitted results of a linear model were applied to strengthen the discovery, demonstrating that the five genes were correlated with immune infiltration score and especially associated with typically immunologically “cold” tumors. We thus conclude that the *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* signature showed a close association with a pan-cancer effect on prognosis and is related to ECM proteins in the TME which corresponding with immunologically “cold” cancer types.

## KEY WORDS

ECM deposition, ECM regulator, immune cell infiltration, immunologically cold, pan-cancer signature

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The extracellular matrix (ECM) is a complex of fibrous, crosslinked proteins that are biochemically and biophysically crucial in the regulation of cell proliferation, survival, differentiation, and migration.<sup>1</sup> In analyzing ECM structure and function, a complete “parts list” of all the proteins existing in any given matrix or contributing to matrices under different conditions are termed the “matrisome”.<sup>2</sup> In mammals, the main components of the matrisome consist of almost 43 collagen subunits, 35 proteoglycans, and approximately 200 glycoproteins. In addition, there are about 176 ECM-affiliated proteins, 250 ECM regulators, and 352 secreted factors defined as matrisome-associated proteins associated within the core matrisome.<sup>1,3,4</sup>

Extensive research has shown that increased ECM deposition and crosslinking indicates the development and progression of cancer.<sup>5,6</sup> The process for controlling ECM dynamics, called ECM remodeling, is precisely regulated during development and is primarily accomplished by the expression of ECM enzymes and activities at multiple levels.<sup>5,7</sup> Of note, the main contributors to the activities of ECM remodeling enzymes are ECM regulators. Dysregulation or mutations in ECM regulators can result in tissue damage and even death.<sup>8,9</sup> For example, ECM overproduction or reduced ECM turnover resulting from ECM regulator dysfunction are prominent in fibrosis,<sup>10</sup> which is related to the increased deposition of collagen I, II, III, V, and IX during tumor formation.<sup>11-13</sup> It is noteworthy that ECM dynamic alterations can cause insufficient numbers of effector immune cells that have impaired ability to recognize tumor antigens or limited avidity regarding the infiltration of the tumor stroma.<sup>14</sup> Recently, it was reported that ECM regulators related to ECM remodeling formed an unfavorable microenvironment to prevent immune cell penetration or serve as a barrier to anti-cancer agents,<sup>5,15,16</sup> which is called immunologic destruction of advanced cold tumors.

As described above, the regulation of ECM alteration in the tumor microenvironment (TME) shares a common molecular mechanism in tumorigenesis and cancer progression, which can be regarded as pan-cancer attribution. In this respect, sustained growth and invasion of tumor tissue can be illustrated as a highly unstable escape from the primary tumor facilitated by ECM remodeling and regulated by ECM regulators, which lead to tumor tissue acquiring cellular aberrations and allowing them to travel and colonize distant organs.<sup>17,18</sup> Thus, we hypothesized that ECM regulators might play an essential role in pan-cancer attribution through a generic effect on the dynamics of ECM alteration.

In this study, we conducted a comprehensive analysis of all known matrisome genes, and identified that ECM regulators served as the primary contributor correlating with overall survival and accounted for much of the difference in

matrisome expression between tumor and adjacent normal tissues. Our research aims to provide a combined gene signature by using an in-house developed pipeline, and to utilize the full spectrum of ECM-related genes for studying the prognostic biomarker in pan-cancer and its related genes.

## 2 | MATERIALS AND METHODS

### 2.1 | Data acquisition and difference analysis

The transcriptome profiles with HTSeq-count format, mutation, methylation, and corresponding clinical information for 33 cancer types collected in The Cancer Genome Atlas (TCGA) database were downloaded from the UCSC Xena browser (<https://xenabrowser.net/>). Low expressing genes were filtered with a threshold mean value of  $\log_2(\text{count} + 1) < 1$ . Differentially expressed genes (DEGs) with  $|\log_2(\text{fold-change})| > 1$  and  $p < 0.05$  between the tumor and the normal adjacent tissues in 21 types of cancer with no fewer than three adjacent normal tissues were identified using the DESeq2 R package.<sup>19</sup> The expression levels of signature genes were compared between primary tumors and adjacent normal tissues using Student's *t*-test. The human matrisome gene list was obtained from the Matrisome Project (<http://matrisomeproject.mit.edu>). Other validated data were downloaded from the Tumor Alterations Relevant for Genomics-driven Therapy (TARGET) and Clinical Outcomes in Multiple Myeloma to Personal Assessment of Genetic Profile (CoMMpass) studies using the UCSC Xena browser (<https://xenabrowser.net/>); the Chinese Glioma Genome Atlas (CGGA, <http://www.cgga.org.cn/index.jsp>), and three cohorts (GSE17536, GSE71187, GSE78229) from the GENE EXPRESSION OMNIBUS (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) database. Immunohistochemistry (IHC) information and images were downloaded from the Human Protein Atlas (HPA, <https://www.proteinatlas.org>). Proteomic data for BRAC, READ, and OV cancer were downloaded from the Clinical Proteomic Tumor Analysis Consortium (CPTAC, <https://proteomics.cancer.gov>). Information on all the datasets used are listed in Table S1.

### 2.2 | Gene set enrichment analysis

Gene set enrichment analysis (GSEA) based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) was carried out using the *ClusterProfiler* R package.<sup>20</sup> GSEA based on ECM categories was performed using the *fgsea* R package.<sup>21</sup> The results were visualized using the *ggplot2* R package. Statistical significance was considered when  $p < 0.05$  or adjusted  $p < 0.05$ .

## 2.3 | Survival analysis

All differentially expressed ECM (DE-ECM) genes underwent univariate Cox regression analysis to screen survival-related ECM (SR-ECM) genes with a PH test:  $p > 0.05$ ; log rank test  $p < 0.05$ , Wald test  $p < 0.05$ , hazard ratio (HR) =  $\exp((\text{Expression of gene} + 1) \times \text{Coefficient}) / \exp(\text{Expression of gene} \times \text{Coefficient}) = \exp(\text{Coefficient})$ . Multivariate Cox regression analysis was used to construct a risk signature based on multiple genes with C\_index calculated by the *survival* R package. The ROC curve plots and AUC values were obtained to validate the prognostic value of the signature using the *survivalROC* R package. The risk score for each type of cancer was calculated using the following formula: risk score = (Expression of gene<sub>1</sub> × Coefficient of gene<sub>1</sub>) + (Expression of gene<sub>2</sub> × Coefficient of gene<sub>2</sub>) + ... + (Expression of gene<sub>i</sub> × Coefficient of gene<sub>i</sub>),<sup>22</sup> where  $i$  represents the gene index. Then, the risk score was scaled by the formula: (risk score – mean (risk score))/sd(risk score). Kaplan-Meier survival analysis was used to evaluate the prognostic value of types of cancers; log rank test  $p < 0.05$  was considered statistically significant. The HR of a signature in each type of cancer was calculated as  $\exp(\text{Coefficient of gene}_1 + \text{Coefficient of gene}_2 + \dots + \text{Coefficient of gene}_i)$ , where  $i$  represents the gene index.

## 2.4 | Investigation of pan survival-related ECM genes

The pan-survival-related ECM (pan-SR-ECM) genes, which can distinguish the tumor and the adjacent normal tissues by classification of 18 types of cancer, were defined by Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) analysis combined with Principal Component Analysis (PCA) using the *umap* R package. The value of transcripts per million (TPM) or count per million (CPM) of the selected pan-SR-ECM genes were used to perform PCA, and the valuable PC components were subjected to UMAP analysis.<sup>23</sup> The correlation of these pan-SR-ECM genes was analyzed using hierarchical clustering combined with the correlation coefficients in all primary tumor tissues from the 18 types of cancer.

## 2.5 | Correlation of infiltrating immune cells with the gene signature

The immune infiltration score and the abundance of 24 immune cell types from tumor tissues were evaluated using ImmuCellAI software,<sup>24</sup> and single- and multi-correlation coefficients with the five ECM regulator genes were calculated based on the fitted results of the linear model,

multi-correlation coefficient calculated by the formula:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k, \quad R = \frac{\sum (y - \bar{y}) (\hat{y} - \bar{\hat{y}})}{\sqrt{\sum (y - \bar{y})^2 (\hat{y} - \bar{\hat{y}})^2}}$$

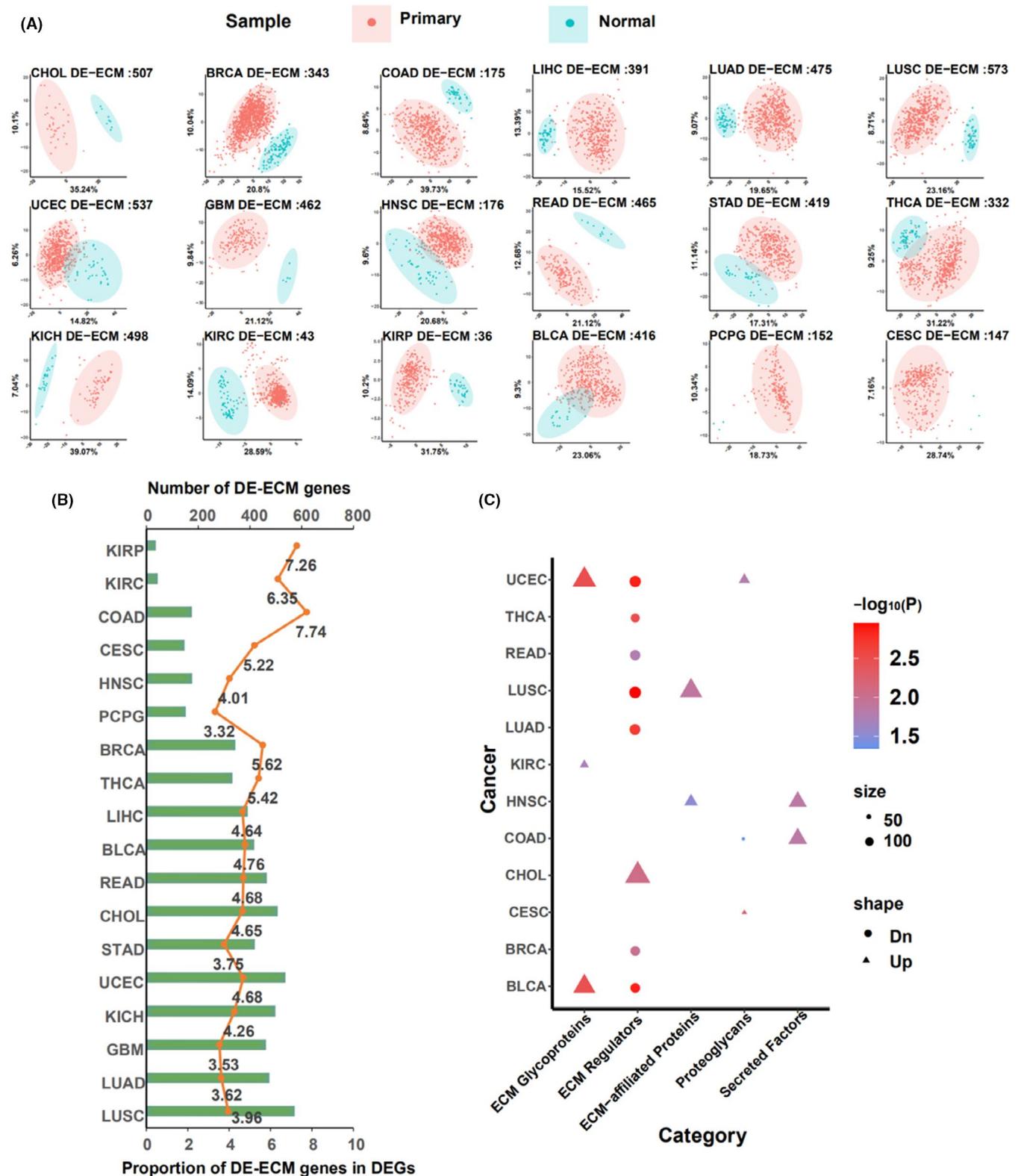
## 2.6 | Statistical analysis

Methylation levels of the five signature ECM regulator genes were compared between primary tumor and adjacent normal tissues by Student's  $t$ -test. Mutations in these genes were compared between wild-type and mutated-type primary tumor tissues by Student's  $t$ -test. If not specified above, a two-sided Student's  $t$ -test was performed and  $p < 0.05$  was considered to be statistically significant. A flowchart for the whole analysis process is shown in Figure S1.

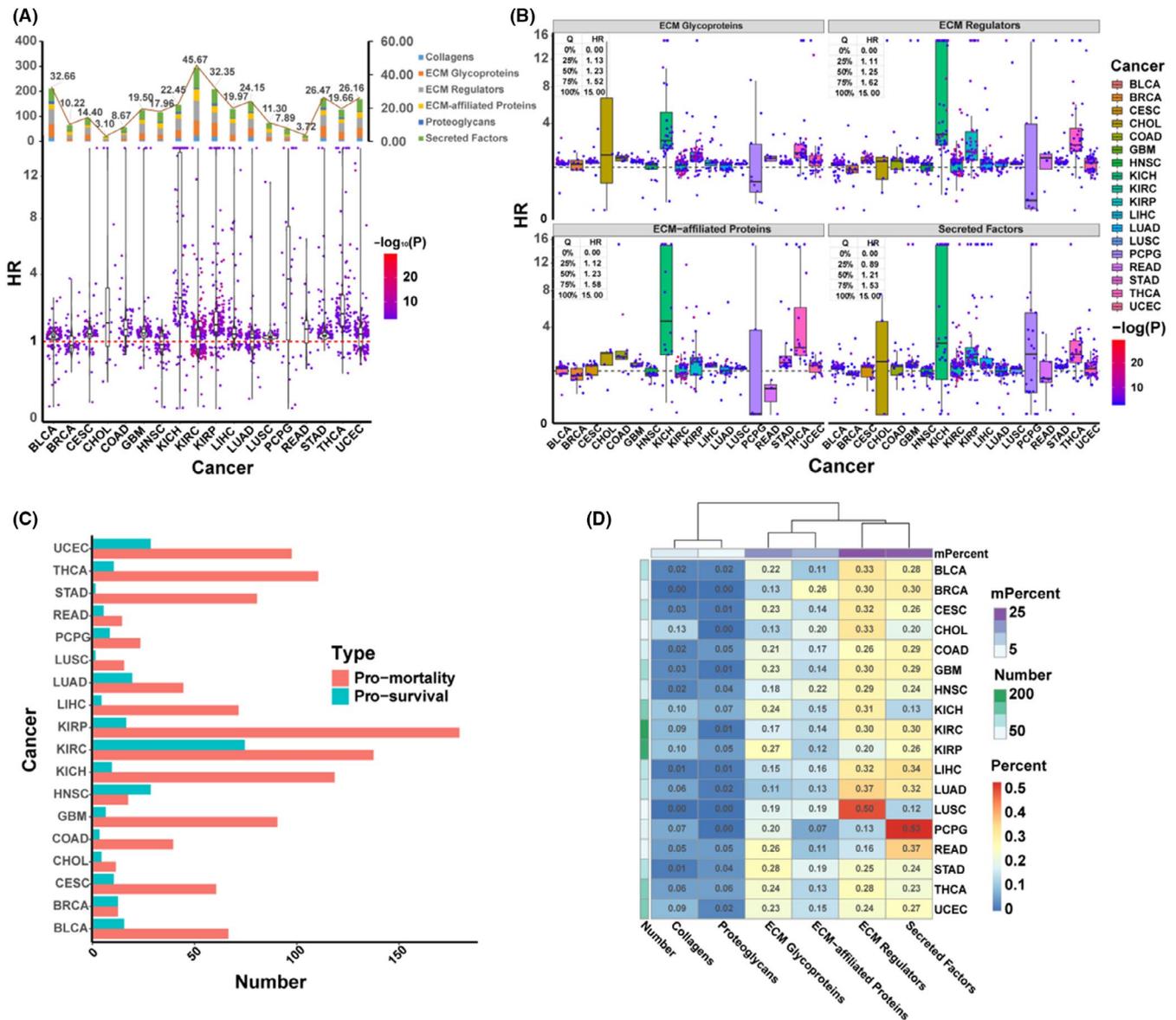
# 3 | RESULTS

## 3.1 | ECM regulator genes were significantly enriched in the matrisome across various cancer types

To investigate the influence of ECM gene expression in cancer, we determined differentially expressed genes (DEGs) between tumor and adjacent normal tissues in the 21 cancer types with no fewer than three adjacent normal samples in TCGA using the *DESeq2* R package. Results obtained from PCA based on differentially expressed ECM genes (DE-ECMs) showed that, 18 of 21 cancers types were selected as the tumor and normal adjacent tissues could be distinguished clearly. DEG analysis and the sample number of the selected cancer types demonstrated that despite only 3.53%–7.74% (mean, 4.86%) of total DEGs being DE-ECMs, only 37 and 44 DE-ECM genes in KIRP and KIRC, respectively, distinguished between tumor and adjacent normal tissues (Figure S2A, B). Meanwhile, KEGG and GO enrichment analysis of expressed ECM (Figure S3) revealed that the ECM genes of different cancers were involved in similar pathways, such as upregulation of cell proliferation, differentiation, and downregulation of immune cell recruitment. Subsequently, in order to identify the predominant category of ECM genes, GSEA analysis was carried out using all ECM and DE-ECM genes based on six matrisome categories (Figure 1C, Figure S2B). The results showed that ECM regulators are significantly enriched in more cancer types (ECMs GSEA: 15 cancer types, DE-ECMs GSEA: 8cancer types) than the other matrisome subtypes. Hence, according to GSEA analysis, ECM regulators may have a more dominant effect between tumor tissue and adjacent parts in various cancer types.



**FIGURE 1** Analysis differentially expressed ECM genes across cancer types. (A) 18 types of cancer were selected by PCA based on their differentially expressed ECM (DE-ECM) genes. Each plot presented a sample (red, primary tumor tissue or Additional - New Primary; blue, adjacent normal tissue). Shadows represent the 95% confidence threshold interval. (B) Distribution of DE-ECM genes in each cancer type. The top y-axis shows the number of DE-ECM genes in each type and the bottom y-axis shows the proportion of DE-ECM genes in DEGs. (C) Enrichment of DE-ECM genes using the *fgsea* R package based on ECM categories. The color presents the p value; results with  $p < 0.05$  are shown. Dn/Up stands for ES < 0/ES > 0. ES: enrichment score



**FIGURE 2** Survival-related ECM (SR-ECM) genes of DE-ECM genes in 18 cancer types. (A) Distribution of SR-ECM genes. SR-ECM genes were screened using univariate Cox regression analysis under the control: PH test.  $p > 0.05$  & Log rank test.  $p < 0.05$  & Wald test.  $p < 0.05$ . Each plot presents a gene, genes with HR >1 are beneficial to survival, genes with HR <1 are unfavorable to survival. (B) Distribution of the four main categories of SR-ECM genes. 16 cancer types showed a median HR >1 for ECM regulators. (C) Pro-mortality and pro-survival SR-ECM genes counts in various cancers. Pro-mortality: HR >1.2; pro-survival: HR <0.8. (D) Distribution of pro-mortality SR-ECM genes categorized in ECM

### 3.2 | ECM regulator genes with a pronounced pan-cancer mortality association

Further analysis was performed to determine if a relationship existed between DE-ECM genes and overall survival (OS). In the 18 cancer types, univariate Cox regression analysis was performed using all DE-ECM genes, and a total of 651 survival-related ECM (SR-ECM) genes were found to have a statistically significant correlation. There were 16 cancer types with a median hazard ratio (HR) value >1, which implied that DE-ECM gene expression played a significant role in the correlation with mortality (Figure 2A).

The predominant groups were secreted factors (mean: 34 genes per cancer), ECM regulators (mean: 33 genes per cancer), ECM glycoproteins (mean: 28 genes per cancer), and ECM-affiliated proteins (mean: 18 genes per cancer). In particular, ECM regulators had HR >1 in 16 of 18 cancer types and the median and upper quartile HR values were greater than other groups, which implied that expression of ECM regulators played a more critical role in mortality (Figure 2B). In addition, SR-ECM genes were divided into pro-mortality (HR >1.2) or pro-survival (HR <0.8) groups in these 18 cancer types (Figure 2C). Pro-mortality genes appeared at significantly higher frequency than the



between the tumor and the adjacent normal tissues in 18 cancer types, was defined by UPMA analysis combined with PCA. We identified 59 SR-ECM genes within at least six cancer types (Figure 3A, Figure S4), which we named pan-SR-ECM genes. As expected, ECM regulators accounted for the highest proportion of these pan-SR-ECM genes (27.12%, 16/59). These data demonstrated that organs could be specified using only these 59 pan-SR-ECM genes. Specific lung (LUSC and LUAD), kidney (KICH, KIRC, and KIRP), liver (LIHC and CHOL), and intestine (COAD and READ) formed characteristic clusters. However, brain resident tumors, such as GBM and PCPG, showed a significant separation, which is consistent with an organ-specific ECM microenvironment. As primary contributors to the difference between tumor and adjacent normal tissues, the gene expression correlation coefficients of these 59 pan-SR-ECM genes were further analyzed using hierarchical clustering in all primary tissues (Figure 3B). The findings showed that the 16 ECM regulator genes identified by this analysis mainly came from five gene families: a disintegrin and metalloproteinase (ADAMs)/related ADAMs with thrombospondin motifs (ADAMTS), “classical” matrix metalloproteinases (MMPs), serine proteinase inhibitors (SERPINs), procollagen-lysine, 2-oxoglutarate 5-dioxygenases (PLODs), and prolyl 4-hydroxylases (P4Hs). Genes from cluster 4 (13 genes) were strongly positively correlated with each other in all primary tissues. ECM regulator genes *P4HA3* and *ADAM12* showed a high expression correlation in cluster 4, representing a cluster of other pan-SR-ECM genes, including four out of six pan-SR-ECM collagen genes, *COL11A1*, *COL6A2*, *COL5A2*, *COL5A1*, and *COL1A2*. *MMP1* from cluster 3 was also strikingly and positively related to *ANXA2* and *S100A19*. In cluster 5, ECM regulators *P4HA1*, *PLOD1*, *SERPINE1*, and *TLL1* were clustered together and were positively correlated with each other. Collectively, using different approaches, we identified five ECM regulator families which showed great significance combined with the most pronounced pan-cancer effect over various cancer types.

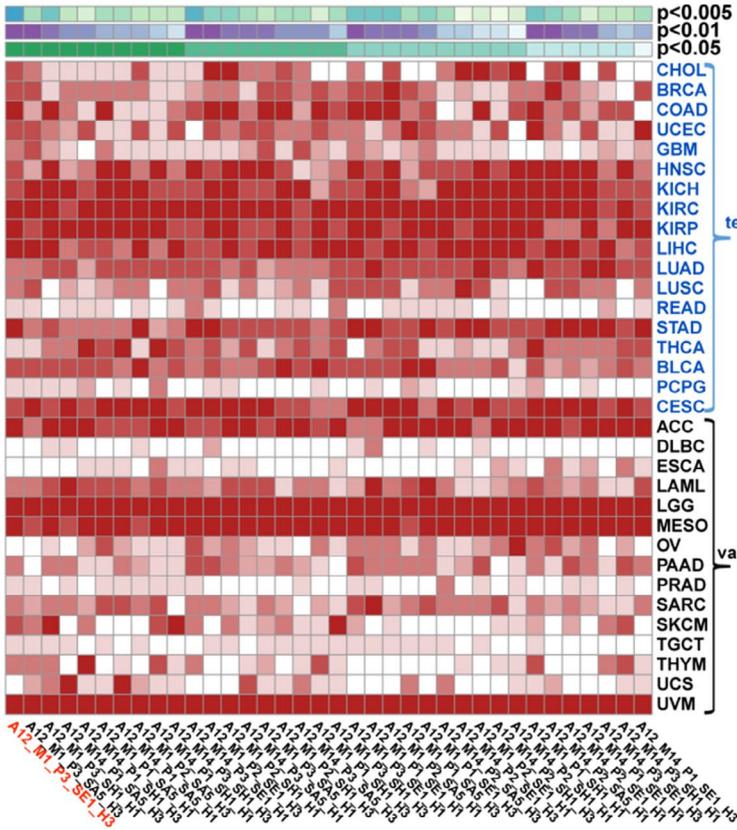
### 3.4 | ADAM12, MMP1, SERPINE1, PLOD3, and P4HA3 as a five-gene signature associated with poor PAN-cancer survival

Since the aim of this study was to identify a representative and multiple gene signature describing pan-biological functional and pan-survival ability, we selected one gene each from the five gene families, which consisted of 11 total pan-SR-ECM regulator genes that were survival-related to at least seven types of cancer. Collectively, this produced 36 types of signatures. The coefficient of *PLOD3* made a greater contribution than the other ten ECM regulator genes using the least absolute shrinkage and selection operator (LASSO) Cox regression model in the 18 cancer types

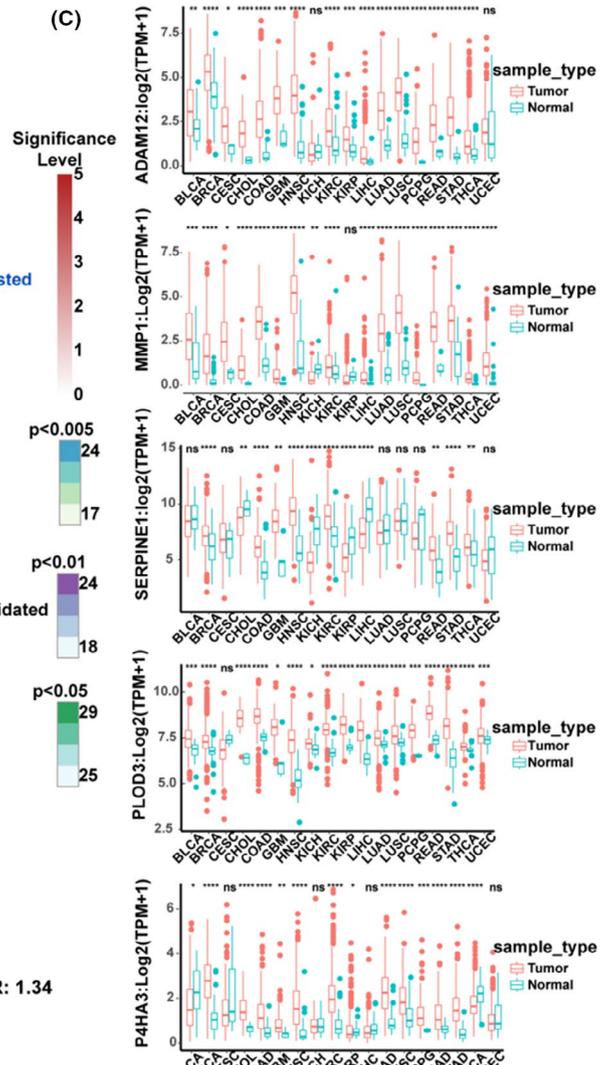
(Figure S5A). Based on the previous analysis, we speculated that a potentially promising five-gene signature would be composed of *P4HA3* and *ADAM12* from cluster 4, *MMP1* from cluster 3, *PLOD3* from cluster 1, and *SERPINE1* from cluster 5. In the 18 cancer types tested, and another 15 validated cancer types, multivariate Cox regression analysis was performed to obtain the scaled risk score of these 36 possible signatures, following which the samples were divided into high and low groups according to the scaled risk score to conduct Kaplan-Meier pan-cancer survival analysis (Figure 4A, Table S2). As expected, all signatures were prominently survival-related in pan-cancer, and the five-gene signature consisting of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* was survival-related in the most types of cancer under different  $p$  threshold values ( $p < 0.05$ , 29 types of cancer;  $p < 0.01$ , 24 types;  $p < 0.005$ , 24 types; mean C\_index = 0.65, mean AUC value = 0.68). Moreover, in order to carry out the evaluation process, the five-gene signature was validated in seven different cancer categories using the independent databases TARGET and CoMMpass (Figure 4B). This result agreed that this five-gene signature was broadly associated with poor survival in various cancer types, which was further consolidated under the number of cancer types in multiple databases for verification purposes (Figure S5B). The survival analysis results based on these cancer types were also examined using the ROC curve method (Figure S5C-D). Moreover, survival analysis based on the signature genes was verified using CPTAC clinical proteomics datasets, which showed that the gene products were related to survival, as expected (Figure S6).

Further analysis of differences in the gene expression levels of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* between tumor and adjacent normal tissues in 18 cancer types from TCGA were examined (Figure 4C). The results showed that the expression levels of these five regulators were significantly different across most types of cancer. In detail, both *ADAM12* and *MMP1* expression were upregulated in 16 cancer types ( $p < 0.05$ ), *PLOD3* was upregulated in 17 cancer types ( $p < 0.05$ ). As a comparison, *SERPINE1* was upregulated in 8 out of 12 cancer types ( $p < 0.05$ ) and *P4HA3* was 12 of 14 cancer types ( $p < 0.05$ ). This finding indicates considerable upregulation in these five genes in cancers, which is consistent with the expression pattern of most ECM regulators in most of the tumor tissues. The findings were further verified using IHC staining slides from the Human Protein Atlas. As a result, the higher protein expression levels of *ADAM12*, *PLOD3*, *P4HA3*, and *SERPINE1* were detected in multiple types of cancer compared with the normal tissue specimens, which was consistent with their mRNA expression patterns (Figure S7). The protein expression levels of *PLOD3* and *P4HA3* were high in most cancer tissues. *MMP1* was not listed in the database, and *ADAM12* showed limited detection. Interestingly, *SERPINE1* demonstrated minimal detection

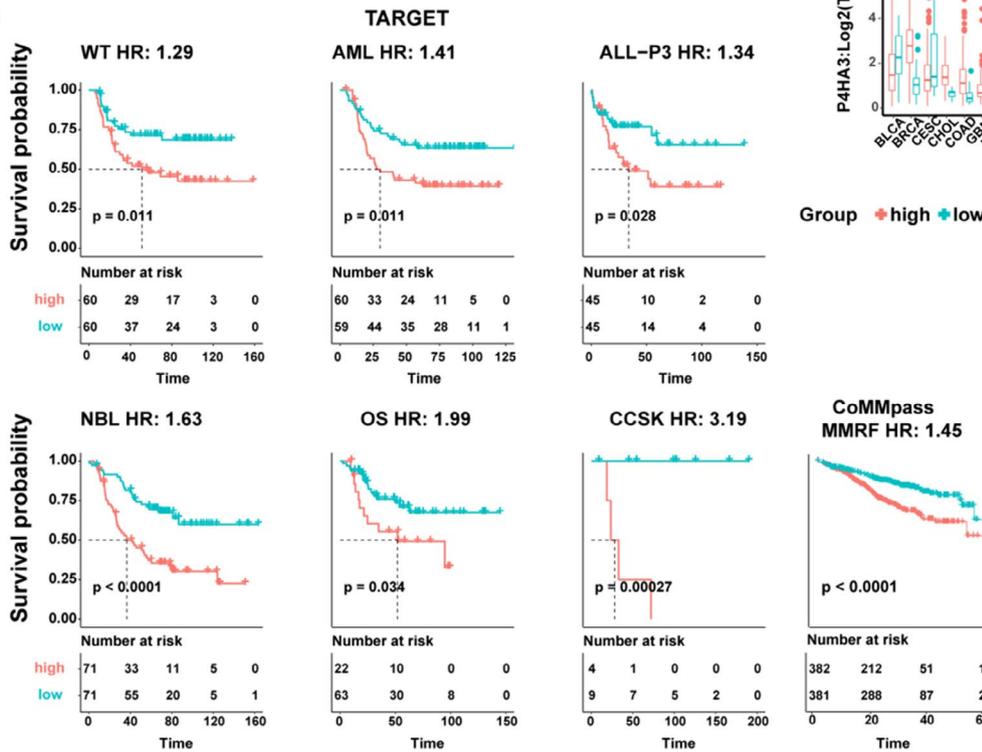
(A)



(C)



(B)



**FIGURE 4** Analysis of the five regulator genes *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* in pan-cancer. (A) Results of different signatures consisting of 11 pan-SR-ECM regulators from the ADAM, MMP, SERPIN, PLOD, and P4H gene families using Kaplan-Meier survival analysis in various cancer types from TCGA. The 18 cancer types of cancer tested with another left 15 types use for validation. A: ADAM, M: MMP, S: SERPIN, P: PLOD, H: P4HA. Statistical significance is depicted as: 0:  $p \geq 0.05$ , 1:  $p < 0.05$ , 2:  $p < 0.01$ , 3:  $p < 0.005$ , 4:  $p < 0.001$ , 5:  $p < 0.0001$ . (B) Kaplan-Meier plot curves of the *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3* and *P4HA3* five-gene signature validated in other cancer types from other databases. The p value was calculated using the log-rank test. HR was calculated using the Cox proportional-hazards model based on scaled risk score. (C) Expression levels of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* in the tumor and adjacent normal tissues across 18 cancer types. TPM: transcripts per million mapped reads. Statistical analyses were performed using the unpaired *t*-test (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ )

level in most tumor tissues, but as suggested in the database, it is generally positive in the tumor-related stroma population.

To determine whether the expression level was the primary factor associated with clinical relevance, gene mutation and methylation status were investigated. In the 18 cancer types from TCGA, we calculated tissues with mutations of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3*, and they had limited mutations, and their mutation sites had few significant effects on gene expression or function gain across the 18 cancer types (Table S3, Figure S8A–D). In contrast, multiple methylation sites demonstrated a substantial impact on the expression of these five genes (Figure S9, Table S4), especially *PLOD3*, *MMP1*, and *ADAM12*, most of which were downregulated in pan-cancer tumor tissues compared with adjacent normal tissues. These data demonstrate that the expression levels of these five genes have universal effects on many cancers.

### 3.5 | Expression of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* in an ECM-receptor interaction gene set correlated with immune cell infiltration

To understand the mechanistic aspect of the five-gene signature, KEGG pathway enrichment analysis between the high- and low-scaled risk score groups was performed in 29 cancer types. The number of significantly enriched pathways (adjusted  $p < 0.05$ ) with a mean value of the normalized enrichment score (mNES) was counted (Figure 5A, Figure S10A). The common proliferation- and apoptosis-related PI3K-Akt and MAPK signaling pathways were listed in 15 and 16 of the 29 cancer types, respectively. Gene sets directly related to ECM included ECM-receptor interaction and focal adhesion were especially enriched in most cancers with the highest mNES scores. The immune response-related gene sets cytokine-cytokine receptor interaction and TNF signaling pathway were also in the top three clusters.

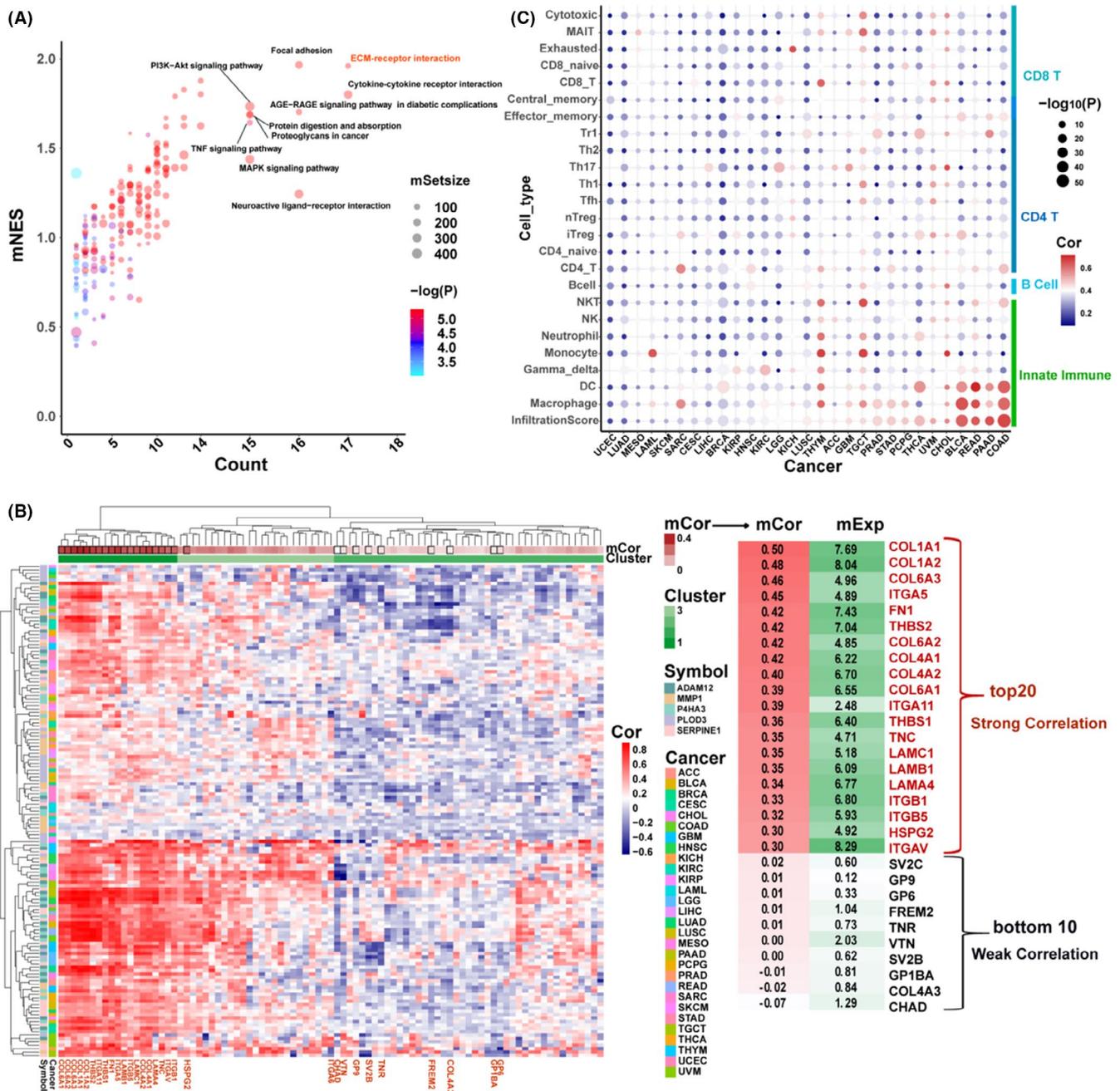
The ECM-receptor interaction gene set was further analyzed for correlation with *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* in the 29 cancer types. The heatmap shows that the ECM-receptor interaction genes were grouped into three clusters, most of which had consistent expression

levels (Figure 5B and Figure S10B). We found that cluster1 contained 19 genes that were highly correlated with expression levels of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* and had higher expression levels as a whole. The top 20 genes with the strongest or top 10 genes with the weakest correlation are presented in Figure 5B. In the top 10 most substantially correlated genes, the collagen family occupied 7 of 10 genes, and *COL1A1* was at the top of the list. In comparison, the weakest related genes showed a much more diversified ECM-related receptor profile (Figure 5B). This finding might imply the function of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* and the molecular properties of the tumor tissue.

Next, we used ImmuCellAI to evaluate the immune infiltration score and the abundance of 24 immune cells from tumor tissues, and then calculated the multiple and single correlation coefficients of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* with mRNA expression and immune infiltration score (Figure 5C, Figure S10C–H). In general, the mRNA expression levels of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* were correlated with immune infiltration score and macrophages and dendritic cells (DCs) in most types of cancer. Specifically, the top three most pronounced cancer types were typical immunologically “cold,” such as COAD, PAAD, and READ. Interestingly, the data illustrated that COAD, PAAD, and READ were highly related to macrophage infiltration with less pronounced CD8+ T cell count; this finding was corroborated using TARGET and CoMMpass (Figure S5B). This observation suggests that *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* are related to the immunological status of the ECM environment in various cancer types.

## 4 | DISCUSSION

The pan-cancer prognostic signature usually involves five clusters: cell cycle phase, immune response, cell adhesion, metabolic process, and gene expression regulation.<sup>25</sup> As reported elsewhere, tumorigenesis and progression are correlated with increased ECM deposition and crosslinking, consistent with all pan-cancer effects listed. Several studies have provided evidence that the ECM contributes to cancer



**FIGURE 5** (A) Distribution of the significantly enriched KEGG pathways with  $mNES > 0$ ; x axis: the KEGG pathway count with  $p$ . adjust  $< 0.05$ , y axis: mNES, medium value of normalized enrichment score. Each plot is a KEGG pathway, coloring based on minimum  $p$ .adjust. (B) The correlation of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* with the ECM-receptor interaction gene set. (C) The multiple correlation coefficients of *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3* with infiltration score and the abundance of types of immune cells. The list was ordered by the correlation coefficient with infiltration score. Results with  $p < 0.05$  are shown

pathogenesis, where ECM regulators are involved in (a) stimulating intracellular signaling that promotes invasion and proliferation; (b) promoting and creating a favorable micro-environmental niche for metastasis; (c) interfering with the communication between cancer and immune cells; and (d) forming an unfavorable microenvironment to prevent immune cell penetration or act as a barrier to anti-cancer agents.<sup>5,15,16</sup> These results imply that ECM dynamic regulation is related to a much broader spectrum of ECM composition. Here,

we discovered that ECM regulator genes were significantly enriched in various cancer types, which also contributed to mortality and differences in 18 different cancer types. The most predominant ECM regulators, *ADAM12*, *MMP1*, *SERPINE1*, *PLOD3*, and *P4HA3*, have been classified as a five-gene signature that produces a broad effect over 29 types of cancer. This effect is related to ECM receptor interaction and other tumor proliferation gene sets, indicating its association with impaired immune infiltration score. This regulation

of ECM dynamics is also consistent with other pan-cancer hallmarks, such as telomerase-associated and epithelial-to-mesenchymal transition (EMT) signatures, which implicate a fundamental effect of regulating cell fate changes in tumorigenesis and tumor progression, such as cell-cell contact, cellular polarity, metabolic processes, and gene expression regulation.<sup>26-30</sup> Our results demonstrated that this five-gene signature is highly representative of alterations to core ECM genes, which is broadly linked to tumorigenesis, and hence can be a definitive collection of prognostic biomarkers.

This five-gene signature can represent a broad spectrum of ECM proteins in tumorigenesis and be used in a diagnostic manner. The ECM provides the functional composition to construct the tumor microenvironment (TME) and facilitates and plays critical roles during multiple stages of tumorigenesis. Increased deposition and crosslinking of ECM proteins contributes to fundamental biochemical and biophysical conditions to facilitate cancer cell proliferation, migration, and invasion.<sup>31-34</sup> Therefore, as the primary contributor to ECM dynamics, ECM regulators are involved in controlling ECM protein deposition and crosslinking. This essential TME effect has a mechanistic advantage in discovering pan-cancer prognostic biomarker markers. Indeed, we found in this study that UMAP analysis combined with PCA showed that these 59 pan-selected genes ( $\geq 6$  cancer types) enabled us to distinguish tumor tissues from adjacent normal tissues at tissue-specific resolution (Figure 3A, Figure S4C). This is consistent with the more stringent criteria that only 31 pan-selected genes ( $\geq 7$  cancer types) genes with 11 ECM regulator genes could differentiate tumor and adjacent normal tissues across 18 cancer types (Figure S4D).

ECM degradation plays an essential role in cancer cell proliferation, migration, invasion, and angiogenesis. *ADAM12* is a member of the disintegrin and metalloproteinase family and contains both a metalloproteinase and a disintegrin domain.<sup>35</sup> It is involved in tumor progression, tumor cell escape, and long-distance transformation in both orthotopic and transgenic models, which cause tumor cell proliferation, migration, and invasion<sup>36-38</sup> and induces tumor cell resistance to apoptosis.<sup>39</sup> *ADAM12* was also reported as a novel regulator in tumor angiogenesis via STAT3 signaling and as a prognostic marker for various cancers,<sup>40</sup> including breast and prostate cancers.<sup>41</sup> Similarly, MMPs are proteinase enzymes that proteolytically digest different protein substrates, including other proteinases, growth factors, cell-surface receptors, and many components of the ECM. *MMP1* is upregulated in various cancers and has been reported to be associated with tumor invasion and metastasis,<sup>42</sup> which is significantly negatively correlated with cancer survival.<sup>43</sup> In contrast, overexpression of *MMP1* has been reported in different cancer types, making a path in the tumor-associated microenvironment and contributing to cancer cell invasion.<sup>44,45</sup> These findings agreed with our observation that the expression levels of these two

proteinase regulators' function related to various collagen genes in a pan-cancer manner (Figure 3B, Figure 5B).

Furthermore, our observations agree with the literature that ECM composition is regulated by a category of ECM modification enzymes, including P4Hs and PLODs. In addition, components of collagen deposition during wound healing were observed in hypoxia in tumor tissue.<sup>46-48</sup> This regulation and modification were processed via enzymes such as P4H, PLOD, and LOX.<sup>49-51</sup> Noteworthy expression of the P4H family is significantly upregulated in breast cancer, whereas downregulation of P4HA causes inhibition of mammary tumor growth and metastasis to the lungs, which agrees with decreased P4HA activity depresses cancer cell arrangement along collagen fibers.<sup>52-54</sup> Similarly, PLOD family expression is also associated with a risk of mortality in breast cancer patients, which indicates metastasis to lymph nodes and lungs.<sup>55</sup> This finding suggests that increasing fibril collagen formation increases tumor elasticity. On the other hand, the role of *SERPINE1* (*PAI-1*) is less clear. *In vitro*, *SERPINE1* prevents excessive pericellular degradation of ECM proteins necessary for cell adhesion and migration. It functions in modulated cell migration by both promoting cell detachment from core ECM proteins and by preventing excessive pericellular ECM degradation.<sup>56</sup> In contrast, *in vivo* studies of *SERPINE1* in metastasis have reported conflicting results, as some experiments suggested a pro-metastatic effect<sup>57-63</sup> while others suggested either an inhibitory role<sup>64-66</sup> or an absence of effect on metastasis.<sup>67,68</sup>

Interestingly, we demonstrated that this five-gene signature showed a significant correlation in cancer progression in a broad spectrum of cancer types (Figure 4A–B). This finding was verified in multiple databases at the gene expression level and CPTAC clinical proteomics datasets and was further confirmed in IHC slides from the Human Protein Atlas (Figure S6 and S8). Although the cross-omics comparison cannot directly recapitulate gene expression levels, these findings suggest that the five-gene signature can produce different kinds of combinatory arrangements. This maybe because the five-gene signature might have a cooperative mechanism, which might not be related to co-expression but provides an additive function in tumorigenesis.

The five-gene signature-related ECM proteins are associated with the TME. Enhanced expression of this mixed signature was related to poor survival and was associated with pan-cancer mortality. This can be explained by the fact that these signature regulator genes might enhance deposition and crosslinking, which alters the biochemical and biophysical conditions of the TME, which leads to cell proliferation, migration, and invasion.<sup>31-34</sup> KEGG enrichment analysis of the five-gene signature suggests underlying mechanisms in the relationship between gene clusters (Figure 5A). The gene sets with the broadest effect over 17 different cancer types

enriched in ECM-receptor interaction and cytokine-cytokine receptor interaction gene sets. The second tier, in 16 types of cancer, was enriched in focal adhesion, which is consistent with previous reports that cancer-derived ECM proteins (fibronectin, collagen, and laminin) are reported to protect cancer cells from chemotherapy-induced apoptosis via activation of the PI3K/AKT pathway.<sup>64,65</sup>

The ECM-receptor interaction gene cluster has a broad effect, which is highly relevant to this five-gene signature in the tumor microenvironment. Interestingly, the top 20 genes from this cluster contained a TME-specific ECM receptor from the tumor stroma (Figure 5B). This is consistent with investigations of the matrisome in colon, lung, and breast cancer tissues,<sup>69,70</sup> suggesting that abnormal collagen deposition in tumor stroma leads to cancer progression, especially increased collagen VI depositions that stimulate cancer cell proliferation.<sup>48,71,72</sup> A bulk of the tumor ECM is only expressed by cancer cells, namely *LAMA4* and *LAMB1*.<sup>60,69,70</sup> Nevertheless, *THBS1*, *FNI*, *TNC*, *COL1*, and *COL4* were expressed in both stroma and cancer cells.<sup>63</sup> Recently, it was reported that stromal cells in micro-dissected human breast tumors expressed the most pronounced level of collagen crosslinking enzymes. This suggests that targeting these enzymes may have therapeutic potential.<sup>73</sup> Moreover, other evidence suggests that genes in this cluster, such as the integrin gene *ITGAV*, were associated with higher progression and spread of various cancers via perineural invasion.<sup>74</sup> In contrast, the ECM receptor *FRME2* demonstrated a weak correlation with the five-gene signature. This observation implies that these regulators are unfavorable to conventional embryo development before birth as the FRAS/FREM complex before birth. This observation indicates that the five-gene signature-associated ECM contributes to the TME. Collectively, these findings can explain the underlying mechanism by which tumor modality correlates with the expression of the five-gene signature.

The five-gene signature implied immunogenicity in cancer for immune checkpoint-related treatment. As reported elsewhere, the TME contains the cellular components of the tumor stroma, which include fibroblasts, endothelial cells, adipocytes, and immune cells. Cancer-associated fibroblasts produce and regulate ECM remodeling in the TME, suggesting that cancer cells might play a role in ECM deposition. Moreover, researchers have utilized nanoparticles to apply MMP2 enzyme activity to reduce ECM deposition in the tumor microenvironment, which results in the enhancement of immune infiltration.<sup>75-77</sup> This molecular operation confirmed that modulating tumor-related ECM deposition and crosslinking could convert “cold” tumors into “hot” tumors via the potentiation of immune infiltration through less ridged ECM. Consequently, overcoming the deficiency of natural T cell responses to tumor cells, as well as T cell immunity resistance within the TME, are two critical

challenges to any anti-tumor T cell-related therapy. More evidence has showed that in anti-PD-1/PD-L1 monotherapy, the response rates which are typically around 25% and can be as low as 5% in immune-deserted (or immunologically “cold”) tumors, such as uveal melanoma, where T cells are almost completely inattentive to the tumor.<sup>78-80</sup> The evaluation of the five-gene signature-related immune infiltration score ranked 29 types of cancer according to the correlation. The most significantly correlated were READ (colon adenocarcinoma) and COAD (rectum adenocarcinoma), which are typically immunologically “cold” tumors, and KICH (kidney chromophobe) which was listed as a weak “cold” tumor.<sup>81</sup> Hence, the five-gene signature might provide a new insight for understanding T cell infiltration in terms of ECM deposition and crosslinking.

### ETHICS STATEMENT

All data of this study were public and required no ethical approval.

### ACKNOWLEDGMENTS

Thanks Dr. Yufei Shui for the contribution to revised the manuscript.

### CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

### AUTHOR CONTRIBUTIONS

SZ, XZ and YCL conceived the project, YCL, SZ and YML designed the research. YCL performed analysis. YCL and XZ co-wrote the final text. SZ, YZY, PZZ and, provided technical assistance. XZ and YZY revised and approved the final manuscript.

### DATA AVAILABILITY STATEMENT

All data generated and analyzed during this work are included in the method section.

### ORCID

Xiao Zhang  <https://orcid.org/0000-0002-3894-8513>

### REFERENCES

1. Naba A, Clauser KR, Hoersch S, et al. 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.
2. Hynes RO, Naba A. Overview of the matrisome—an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol*. 2012;4(1):a004903.
3. Theocharis AD, Skandalis SS, Gialeli C, et al. Extracellular matrix structure. *Adv Drug Deliv Rev*. 2016;97:4-27.
4. Theocharis AD, Manou D, Karamanos NK. The extracellular matrix as a multitasking player in disease. *FEBS J*. 2019;286(15):2830-2869.

5. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol*. 2014;15(12):786-801.
6. Xiong G-F, Xu R. Function of cancer cell-derived extracellular matrix in tumor progression. *J Cancer Metastasis Treatment*. 2016;2:357-364.
7. Lu P, Takai K, Weaver VM, et al. Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harbor Perspectives Biol*. 2011;3(12):a005058.
8. Novoseletskaya ES, Grigorieva OA, Efimenko AY, et al. Extracellular matrix in the regulation of stem cell differentiation. *Biochemistry (Mosc)*. 2019;84(3):232-240.
9. Brown Y, Hua S, Tanwar PS. Extracellular matrix-mediated regulation of cancer stem cells and chemoresistance. *Int J Biochem Cell Biol*. 2019;109:90-104.
10. Frantz C, Stewart KM, Weaver VM. The extracellular matrix at a glance. *J Cell Sci*. 2010;123(Pt 24):4195-4200.
11. Zhu L, Slee GR, Domenico DR, et al. B-cell lymphoma of ampulla of Vater: observation for six years. *Pancreas*. 1995;10(2):208-210.
12. Kauppila S, Stenbäck F, Risteli J, et al. Aberrant type I and type III collagen gene expression in human breast cancer in vivo. *J Pathol*. 1998;186(3):262-268.
13. Huijbers IJ, Irvani M, Popov S, et al. A role for fibrillar collagen deposition and the collagen internalization receptor endo180 in glioma invasion. *PLoS One*. 2010;5(3):e9808.
14. Yuzhalin AE, Lim SY, Kutikhin AG, et al. Dynamic matrisome: ECM remodeling factors licensing cancer progression and metastasis. *Biochim Biophys Acta Rev Cancer*. 2018;1870(2):207-228.
15. Rainero E. Extracellular matrix endocytosis in controlling matrix turnover and beyond: emerging roles in cancer. *Biochem Soc Trans*. 2016;44(5):1347-1354.
16. Zhu J, Xiong G, Trinkle C, et al. Integrated extracellular matrix signaling in mammary gland development and breast cancer progression. *Histol Histopathol*. 2014;29(9):1083-1092.
17. Chambers AF, Werb Z. Invasion and metastasis—recent advances and future challenges. *J Mol Med (Berl)*. 2015;93(4):361-368.
18. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. *Cell*. 2017;168(4):670-691.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
20. Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic*. 2012;16(5):284-287.
21. Sergushichev AA. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*. 2016:060012.
22. Chen H-Y, Yu S-L, Chen C-H, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med*. 2007;356(1):11-20.
23. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMANat Biotechnol, 2018.
24. Miao Y-R, Zhang Q, Lei Q, et al. ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv Sci (Weinh)*. 2020;7(7):1902880.
25. Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med*. 2015;21(8):938-945.
26. Luo Z, Wang W, Li F, et al. Pan-cancer analysis identifies telomerase-associated signatures and cancer subtypes. *Mol Cancer*. 2019;18(1):106.
27. Mak MP, Tong P, Diao L, et al. A patient-derived, pan-cancer emt signature identifies global molecular alterations and immune target enrichment following epithelial-to-mesenchymal transition. *Clin Cancer Res*. 2016;22(3):609-620.
28. Barger C, Branick C, Chee L, et al. Pan-cancer analyses reveal genomic features of FOXM1 overexpression in cancer. *Cancers*. 2019;11(2):251.
29. Yuzhalin AE, Urbonas T, Silva MA, et al. A core matrisome gene signature predicts cancer outcome. *Br J Cancer*. 2018;118(3):435-440.
30. Chakravarthy A, Khan L, Bensler NP, et al. TGF- $\beta$ -associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. *Nature Commun*. 2018;9(1):4692.
31. Bix M, Liao N-S, Zijlstra M, et al. Rejection of class I MHC-deficient haemopoietic cells by irradiated MHC-matched mice. *Nature*. 1991;349(6307):329-331.
32. Domínguez-Soto A, Sierra-Filardi E, Puig-Kröger A, et al. Dendritic cell-specific ICAM-3-grabbing nonintegrin expression on M2-polarized and tumor-associated macrophages is macrophage-CSF dependent and enhanced by tumor-derived IL-6 and IL-10. *J Immunol*. 2011;186(4):2192-2200.
33. Donadon M, Hudspeth K, Cimino M, et al. Increased infiltration of natural killer and T cells in colorectal liver metastases improves patient overall survival. *J Gastrointest Surg*. 2017;21(8):1226-1236.
34. Donskov F. Immunomonitoring and prognostic relevance of neutrophils in clinical trials. *Semin Cancer Biol*. 2013;23(3):200-207.
35. Kveiborg M, Albrechtsen R, Couchman JR, et al. Cellular roles of ADAM12 in health and disease. *Int J Biochem Cell Biol*. 2008;40(9):1685-1702.
36. Roy R, Rodig S, Bielenberg D, et al. ADAM12 transmembrane and secreted isoforms promote breast tumor growth: a distinct role for ADAM12-S protein in tumor metastasis. *J Biol Chem*. 2011;286(23):20758-20768.
37. Rao VH, Kandel A, Lynch D, et al. A positive feedback loop between HER2 and ADAM12 in human head and neck cancer cells increases migration and invasion. *Oncogene*. 2012;31(23):2888-2898.
38. Georges S, Chesneau J, Hervouet S, et al. A disintegrin and metalloproteinase 12 produced by tumour cells accelerates osteosarcoma tumour progression and associated osteolysis. *Eur J Cancer*. 2013;49(9):2253-2263.
39. Kveiborg M, Fröhlich C, Albrechtsen R, et al. A role for ADAM12 in breast tumor progression and stromal cell apoptosis. *Cancer Res*. 2005;65(11):4754-4761.
40. Roy R, Dagher A, Butterfield C, et al. ADAM12 is a novel regulator of tumor angiogenesis via STAT3 signaling. *Mol Cancer Res*. 2017;15(11):1608-1622.
41. Shao S, Li Z, Gao W, et al. ADAM-12 as a diagnostic marker for the proliferation, migration and invasion in patients with small cell lung cancer. *PLoS One*. 2014;9(1):e85936.
42. Chambers AF, Matrisian LM. Changing views of the role of matrix metalloproteinases in metastasis. *J Natl Cancer Inst*. 1997;89(17):1260-1270.
43. Murray GI, Duncan ME, O'Neil P, et al. Matrix metalloproteinase-1 is associated with poor prognosis in colorectal cancer. *Nat Med*. 1996;2(4):461-462.

44. Pritchard S, Nicolson M, Lloret C, et al. Expression of matrix metalloproteinases 1, 2, 9 and their tissue inhibitors in stage II non-small cell lung cancer: implications for MMP inhibition therapy. *Oncol Rep.* 2001;8(2):421-424.
45. Schütz A, Schneidenbach D, Aust G, et al. Differential expression and activity status of MMP-1, MMP-2 and MMP-9 in tumor and stromal cells of squamous cell carcinomas of the lung. *Tumour Biol.* 2002;23(3):179-184.
46. Dieu-Nosjean M-C, Antoine M, Danel C, et al. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. *J Clin Oncol.* 2008;26(27):4410-4417.
47. Di Giacomo AM, Ascierto PA, Pilla L, et al. Ipilimumab and fotemustine in patients with advanced melanoma (NIBIT-M1): an open-label, single-arm phase 2 trial. *Lancet Oncol.* 2012;13(9):879-886.
48. Fridlender ZG, Sun J, Kim S, et al. Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN. *Cancer Cell.* 2009;16(3):183-194.
49. DeNardo DG, Brennan DJ, Rexhepaj E, et al. Leukocyte complexity predicts breast cancer survival and functionally regulates response to chemotherapy. *Cancer Discov.* 2011;1(1):54-67.
50. Chen Q, Ross AC. All-trans-retinoic acid and the glycolipid  $\alpha$ -galactosylceramide combined reduce breast tumor growth and lung metastasis in a 4T1 murine breast tumor model. *Nutr Cancer.* 2012;64(8):1219-1227.
51. Chiba S, Ikushima H, Ueki H, et al. Recognition of tumor cells by Dectin-1 orchestrates innate immune cells for anti-tumor responses. *Elife.* 2014;3:e04177.
52. Chowell D, Morris LGT, Grigg CM, et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science.* 2018;359(6375):582-587.
53. Clemente CG, Mihm MC, Bufalino R, et al. Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. *Cancer.* 1996;77(7):1303-1310.
54. Dilmeç F, Ozgonul A, Uzunkoy A, et al. Investigation of CTLA-4 and CD28 gene polymorphisms in a group of Turkish patients with colorectal cancer. *Int J Immunogenet.* 2008;35(4-5):317-321.
55. Chong TW, Goh FY, Sim MY, et al. CD1d expression in renal cell carcinoma is associated with higher relapse rates, poorer cancer-specific and overall survival. *J Clin Pathol.* 2015;68(3):200-205.
56. Fridman WH, Pagès F, Sautès-Fridman C, et al. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer.* 2012;12(4):298-306.
57. Frisch M, Biggar RJ, Engels EA, et al. Association of cancer with AIDS-related immunosuppression in adults. *JAMA.* 2001;285(13):1736-1745.
58. Gabrilovich DI, Chen HL, Girgis KR, et al. Production of vascular endothelial growth factor by human tumors inhibits the functional maturation of dendritic cells. *Nat Med.* 1996;2(10):1096-1103.
59. Gebremeskel S, Clattenburg DR, Slauenwhite D, et al. Natural killer T cell activation overcomes immunosuppression to enhance clearance of postsurgical breast cancer metastasis in mice. *Oncoimmunology.* 2015;4(3):e995562.
60. Gasser S, Orsulic S, Brown EJ, et al. The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature.* 2005;436(7054):1186-1190.
61. Frydrychowicz M, Boruckowski M, Kolecka-Bednarczyk A, et al. The dual role of treg in cancer. *Scand J Immunol.* 2017;86(6):436-443.
62. Furudate S, Fujimura T, Kambayashi Y, et al. Sequential therapy with nivolumab followed by ipilimumab induces complete response in metastatic melanoma of the lung but with severe hepatotoxicities. *Case Rep Oncol.* 2016;9(3):644-649.
63. Germano G, Lamba S, Rospo G, et al. Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. *Nature.* 2017;552(7683):116-120.
64. Gong L, Cumpian AM, Caetano MS, et al. Promoting effect of neutrophils on lung tumorigenesis is mediated by CXCR2 and neutrophil elastase. *Mol Cancer.* 2013;12(1):154.
65. Goldberg SB, Gettinger SN, Mahajan A, et al. Pembrolizumab for patients with melanoma or non-small-cell lung cancer and untreated brain metastases: early analysis of a non-randomised, open-label, phase 2 trial. *Lancet Oncol.* 2016;17(7):976-983.
66. Thomas SS, Makar KW, Li L, et al. Tissue-specific patterns of gene expression in the epithelium and stroma of normal colon in healthy individuals in an aspirin intervention trial. *BMC Med Genet.* 2015;16:18.
67. Gopinathan A, Morton JP, Jodrell DI, et al. GEMMs as preclinical models for testing pancreatic cancer therapies. *Dis Model Mech.* 2015;8(10):1185-1200.
68. Gonzalez H, Robles I, Werb Z. Innate and acquired immune surveillance in the postdissemination phase of metastasis. *FEBS J.* 2018;285(4):654-664.
69. Byrne AT, Alférez DG, Amant F, et al. Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nat Rev Cancer.* 2017;17(4):254-268.
70. Campbell DJ. Control of regulatory T cell migration, function, and homeostasis. *J Immunol.* 2015;195(6):2507-2513.
71. Francisco LM, Salinas VH, Brown KE, et al. PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. *J Exp Med.* 2009;206(13):3015-3029.
72. Finisguerra V, Di Conza G, Di Matteo M, et al. MET is required for the recruitment of anti-tumoural neutrophils. *Nature.* 2015;522(7556):349-353.
73. Maller O, Drain AP, Barrett AS, et al. Tumour-associated macrophages drive stromal cell-dependent collagen crosslinking and stiffening to promote breast cancer aggression. *Nat Mater.* 2021;20(4):548-559.
74. Cheuk IW, Siu MT, Ho JC, et al. ITGAV targeting as a therapeutic approach for treatment of metastatic breast cancer. *Am J Cancer Res.* 2020;10(1):211-223.
75. Veiseh O, Gunn JW, Kievit FM, et al. Inhibition of tumor-cell invasion with chlorotoxin-bound superparamagnetic nanoparticles. *Small.* 2009;5(2):256-264.
76. Zhu L, Wang T, Perche F, et al. Enhanced anticancer activity of nanopreparation containing an MMP2-sensitive PEG-drug conjugate and cell-penetrating moiety. *Proc Natl Acad Sci USA.* 2013;110(42):17047-17052.
77. Luo L, Lv M, Zhuang X, et al. Irradiation increases the immunogenicity of lung cancer cells and irradiation-based tumor cell vaccine elicits tumor-specific T cell responses in vivo. *Onco Targets Ther.* 2019;12:3805-3815.
78. Bedard PL, Hansen AR, Ratain MJ, et al. Tumour heterogeneity in the clinic. *Nature.* 2013;501(7467):355-364.
79. Zimmer L, Vaubel J, Mohr P, et al. Phase II DeCOG-study of ipilimumab in pretreated and treatment-naïve patients with metastatic uveal melanoma. *PLoS One.* 2015;10(3):e0118564.
80. Zappasodi R, Merghoub T, Wolchok JD. Emerging concepts for immune checkpoint blockade-based combination therapies. *Cancer Cell.* 2018;33(4):581-598.

81. Li Y, Kang K, Krahn JM, et al. A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data. *BMC Genom.* 2017;18(1):508.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Yu C, You M, Zhang P, et al. A five-gene signature is a prognostic biomarker in pan-cancer and related with immunologically associated extracellular matrix. *Cancer Med.* 2021;10:4629–4643. <https://doi.org/10.1002/cam4.3986>