

ARTICLE

DOI: 10.1038/s41467-018-06654-8

OPEN

TGF- β -associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure

Ankur Chakravarthy^{1,2}, Lubaba Khan^{3,4,5}, Nathan Peter Bensler^{3,4,5},
Pinaki Bose^{3,4,5,6} & Daniel D. De Carvalho ^{1,2}

The extracellular matrix (ECM) is a key determinant of cancer progression and prognosis. Here we report findings from one of the largest pan-cancer analyses of ECM gene dysregulation in cancer. We define a distinct set of ECM genes upregulated in cancer (C-ECM) and linked to worse prognosis. We found that the C-ECM transcriptional programme dysregulation is correlated with the activation of TGF- β signalling in cancer-associated fibroblasts and is linked to immunosuppression in otherwise immunologically active tumours. Cancers that activate this programme carry distinct genomic profiles, such as *BRAF*, *SMAD4* and *TP53* mutations and *MYC* amplification. Finally, we show that this signature is a predictor of the failure of PD-1 blockade and outperforms previously-proposed biomarkers. Thus, our findings identify a distinct transcriptional pattern of ECM genes in operation across cancers that may be potentially targeted, pending preclinical validation, using TGF- β blockade to enhance responses to immune-checkpoint blockade.

¹Princess Margaret Cancer Centre, University Health Network, Toronto M5G 1L7 ON, Canada. ²Department of Medical Biophysics, University of Toronto, Toronto M5G 1L7 ON, Canada. ³Ohlson Research Initiative, Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary T2N 4N1 AB, Canada. ⁴Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary T2N 4N1 AB, Canada. ⁵Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary T2N 4N1 AB, Canada. ⁶Department of Surgery, Cumming School of Medicine, University of Calgary, Calgary T2N 4N1 AB, Canada. These authors contributed equally: Ankur Chakravarthy, Lubaba Khan. Correspondence and requests for materials should be addressed to P.B. (email: pbose@ucalgary.ca) or to D.D.D.C. (email: daniel.decarvalho@uhnresearch.ca)

The ability to disseminate, invade and successfully colonise other tissues is a critical hallmark of cancer that involves remodelling of the extracellular matrix (ECM) laid down by fibroblasts¹. Moreover, cancer-associated fibroblasts (CAFs) produce key growth factors and cytokines as components of the ECM that fuel tumour growth, metastasis and chemoresistance and immune response^{2–4}. Further, ECM changes also predict prognosis in pancreatic⁵ and colorectal cancers^{6,7}.

Here we examine the pan-cancer landscape of ECM gene dysregulation and find that a subset of ECM genes is dysregulated specifically in cancer and is enriched among transcriptional changes that distinguish normal from malignant tissue. We further show that the high expression of this subset of genes is adversely prognostic in pan-cancer analyses. Then, using deconvolution and analyses of transcriptional profiles from dissociated tumour fractions, we show that these genes are modulated in CAFs.

Subsequently, based on multiplatform analysis of The Cancer Genome Atlas (TCGA) data, we correlated these profiles to transforming growth factor (TGF)- β signalling in the tumour microenvironment and show that this transcriptional programme is enriched in immunologically active cancers, suggesting a possible role in immune evasion/adaptation. Finally, we demonstrate that this transcriptional programme predicts responses to immune checkpoint blockade better than mutation burden⁸, cytolytic activity (CYT)⁹, TGF- β expression alone, a CAF signature¹⁰ or a T cell-inflamed signature¹¹. We have thus identified a novel signature of immune evasion that is a potential target for pharmacological modulation and may facilitate effective patient stratification in precision immunotherapy, pending preclinical validation.

Results

Definition of a pan-cancer ECM dysregulation profile. Initially, to study ECM gene dysregulation across cancers, we defined a transcriptional signature to distinguish malignant ($n = 8043$) and normal tissues ($n = 704$) accounting for tumour type ($n = 15$) from TCGA and tested for enrichment of an ECM-associated gene-set we curated based on gene ontology terms (Supplementary Table 1, Supplementary Figure 1A). We were motivated to define such a pan-cancer signature given the wide variability of ECM gene transcription per se across tissue types (Supplementary Figure 1B).

This identified 58 out of the 249 ECM genes represented in the RNA-seq data set to be cancer associated (from hereon referred to as cancer-associated ECM genes (C-ECM genes)) (Supplementary Table 2), comprising 30 out of the 522 upregulated genes, and 28 out of the 644 downregulated genes, representing significant enrichment among both upregulated (odds ratio (OR) = 3.51, $p < 3.9e-8$, Fisher's Exact Test, two-sided) and downregulated (OR = 2.57, $p = 3e-5$, Fisher's Exact Test) genes in malignant tissues (Fig. 1a). These C-ECM genes showed generally high correlations and clustering the Pearson correlation matrix of intergene correlations across TCGA cancers segregated them into generally distinct blocks of upregulated and downregulated genes (Supplementary Figure 1C, two sided), suggesting co-regulation.

Notably, 48 out of the 58 of these genes were also implicated in a previous proteomics-based approach to define a cancer matrisome¹² and we further validated our signature at the proteomic level by examining transcript–protein correlations using matched BRCA samples from CPTAC¹³, wherein mostly positive correlations were observed for 37 out of the 49 C-ECM genes covered by both mass spectrometry and RNA-seq (Supplementary Figure 1D). Analysis using the CPTAC ovarian cancer data set also yielded similar correlations, with the caveat

that only 24 C-ECM genes were represented in the mass spectrometric data set (Supplementary Figure 1E)¹⁴.

Upon summarisation using ssGSEA (single sample Gene Set Enrichment Analysis) scores^{15,16}, these C-ECM genes show broad variation across tumour types (Fig. 1b, Supplementary Figure 1B, F). We then performed a Cox regression based on quartile-thresholded C-ECM scores with American Joint Committee on Cancer stage and tumour type as strata to examine the prognostic impact of this dysregulation; upregulated C-ECM genes were significantly associated with poor prognosis (Fig. 1c, d, hazard ratio (HR) = 1.73, $p < 6.3e-7$ for top versus bottom quartile, Cox regression, two sided) while downregulated C-ECM genes were not (Supplementary Figure 1G), suggesting that the variation we observed in C-ECM gene transcription is clinically relevant. We also tested for the independent prognostic capability of C-ECM genes by repeating this regression with non-ECM upregulated and downregulated gene score quartiles included as covariates and found that C-ECM-up scores were still prognostic (HR = 1.41, $p = 0.003$ for top versus bottom quartile, full model coefficients in Supplementary Table 3, Cox regression, two sided).

We also performed survival analyses where the quartiles of up- and-down scores were combined into single categories. This showed that cancers with the highest quartile of the up-score (Q4) and the lowest quartile of the down-score (Q1) had markedly increased risk of death (Supplementary Figure 1H, HR = 3.18, $p < 8.5e-5$, Cox regression, two-sided).

C-ECM dysregulation is associated with the presence of CAFs.

Given the previously identified role of distinct stromal cells in determining the composition and behaviour of the ECM¹⁷, we then attempted to infer the potential cell types associated with C-ECM transcriptional variation. Using a range of computational approaches, we examined whether changes in cellular composition, along with cell-type-specific transcriptional changes, could be associated with C-ECM gene dysregulation and found multiple indicators that C-ECM gene dysregulation originated in CAFs.

First, we sought to estimate whether there was a cancer-specific contribution to the C-ECM signature by examining correlations between C-ECM scores and tumour purity. This was motivated by the reasoning that if the C-ECM signature originated in the tumour epithelial compartment it would be positively associated with tumour cellularity.

We found tumour purity estimated using ABSOLUTE, which jointly estimates cancer cell fractions of variants, purity and ploidy using copy-number data and whole-exome data¹⁸, was inversely correlated for both C-ECM-up and -down scores (Fig. 2a, S2A). This finding was additionally supported by inverse correlations between C-ECM scores based on the median variant allele frequency (VAF) ($\text{Rho} = -0.17$, for the up-score, and $\text{Rho} = -0.36$, down-score, $p < 2.2e-16$, Spearman's correlation) and purity estimates independently computed using allele-specific copy number profiles using ASCAT¹⁹, downloaded from COSMIC²⁰ ($\text{Rho} = -0.28$, up-score, and -0.3 , down-score, $p < 2.2e-16$, Spearman's correlation). The lower correlation of the median VAF with ECM scores is expected given the confounding influence of ploidy on the relationship between cancer cell fraction and VAF²¹.

The breakdown of correlations between C-ECM scores and purity estimates for different methods by tumour type have been presented in Supplementary Figure 2B, and collectively, these findings hinted at a stromal origin for C-ECM transcriptional changes.

We then leveraged multiple transcriptomic data sets that were separated into epithelial and stromal components by microdissection or cell sorting at various resolutions from ovarian

cancer, head and neck cancer and colorectal cancer, as well as orthogonal pan-cancer DNA methylation-based deconvolution approaches in order to further evaluate the hypothesis of stromal origins and to identify specific cell types associated with the C-ECM dysregulation. First, projecting the C-ECM signature onto microdissected ovarian cancer stroma, matched epithelium and their normal counterparts²² (GSE40595) resulted in clustering by sample type with strong stromal expression (Fig. 2b). Additionally, probes differentially expressed between cancer epithelium and stroma were significantly enriched for both C-ECM-up and -down genes while differentially expressed probes between cancer and normal stroma were enriched in C-ECM-up genes alone (Fig. 2c). Altogether, these results suggest that our C-ECM-up signature comprises genes upregulated in the cancer stroma versus the normal stroma, while our C-ECM-down signature comprises genes upregulated in stroma (normal and cancer) versus epithelial tissue. In contrast, non-ECM genes in the pan-cancer cancer-versus-normal signature displayed weaker enrichment or epithelial associations in some cases (Supplementary Table 4).

Subsequently, DNA methylation-based deconvolution analysis using MethylCIBERSORT²³ implicated CAFs, CD8 T cells and CD14 monocytes as directly correlated with C-ECM signature

scores pan-cancer (Fig. 2d) in addition to confirming inverse associations with tumour purity (Supplementary Figure 2B). Importantly, C-ECM-up genes (ssGSEA scores) showed a positive correlation to the inferred CAF frequency in most TCGA cancer types (Supplementary Figure 2C). We also validated these DNA methylation-based inferences of cellular association using transcript levels of known marker genes for CYT (geometric mean of *GZMA* and *PRFI*), CD8 T cells (*CD8A* expression), CAFs (*ACTA2* expression) and monocytes (*CD14* expression) where-upon we noticed strong, consistent agreement (Supplementary Figure 2D).

Further validation of stromal and CAF association was then performed in three colorectal cancer microarray data sets (GSE39397 and GSE35144)^{6,7}. In these data sets, C-ECM genes were strongly expressed in the stromal compartment and, when cell types were resolved further, mostly localised to fibroblast fractions (Supplementary Figure 2E). We also found that upon comparison of xenograft expression profiles (which retain just human tumour epithelium) with their matched primary tumour counterparts, the expression of C-ECM genes was completely ablated upon the loss of human tumour stroma (Supplementary Figure 2E).

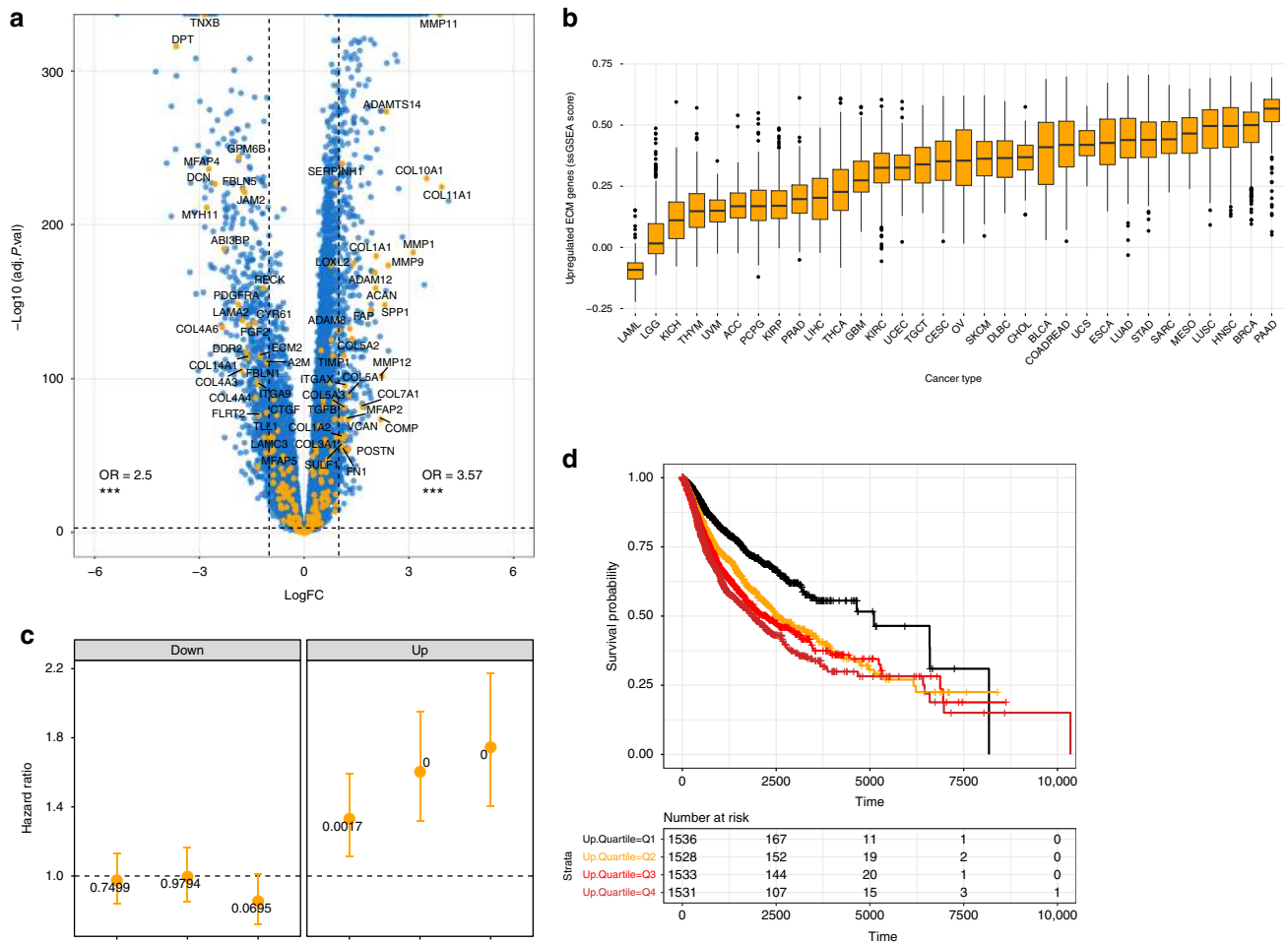


Fig. 1 ECM genes are significantly associated with tumorigenesis and prognosis. **a** Volcano plot showing fold changes for genes differentially expressed between cancer and normal samples. ECM genes are enriched in upregulated and downregulated genes. **b** Boxplots of C-ECM-up enrichment scores show variation across tumour types ($n = 9716$, see S1F for C-ECM-down genes). **c** Plot of Cox model coefficients by quartile for C-ECM-up and -down scores pan-cancer, error bars indicate 95% confidence intervals ($n = 6128$). **d** Unadjusted Kaplan-Meier curves showing survival by C-ECM-up-quartile., asterisks indicate statistical significance. $***p < 0.001$. On the volcano plot, y axis = $-\log_{10}$ fold change, x axis = test statistic/fold change/Spearman’s Rho. On volcano plots, all enrichment statistics are from Fisher’s Exact Tests

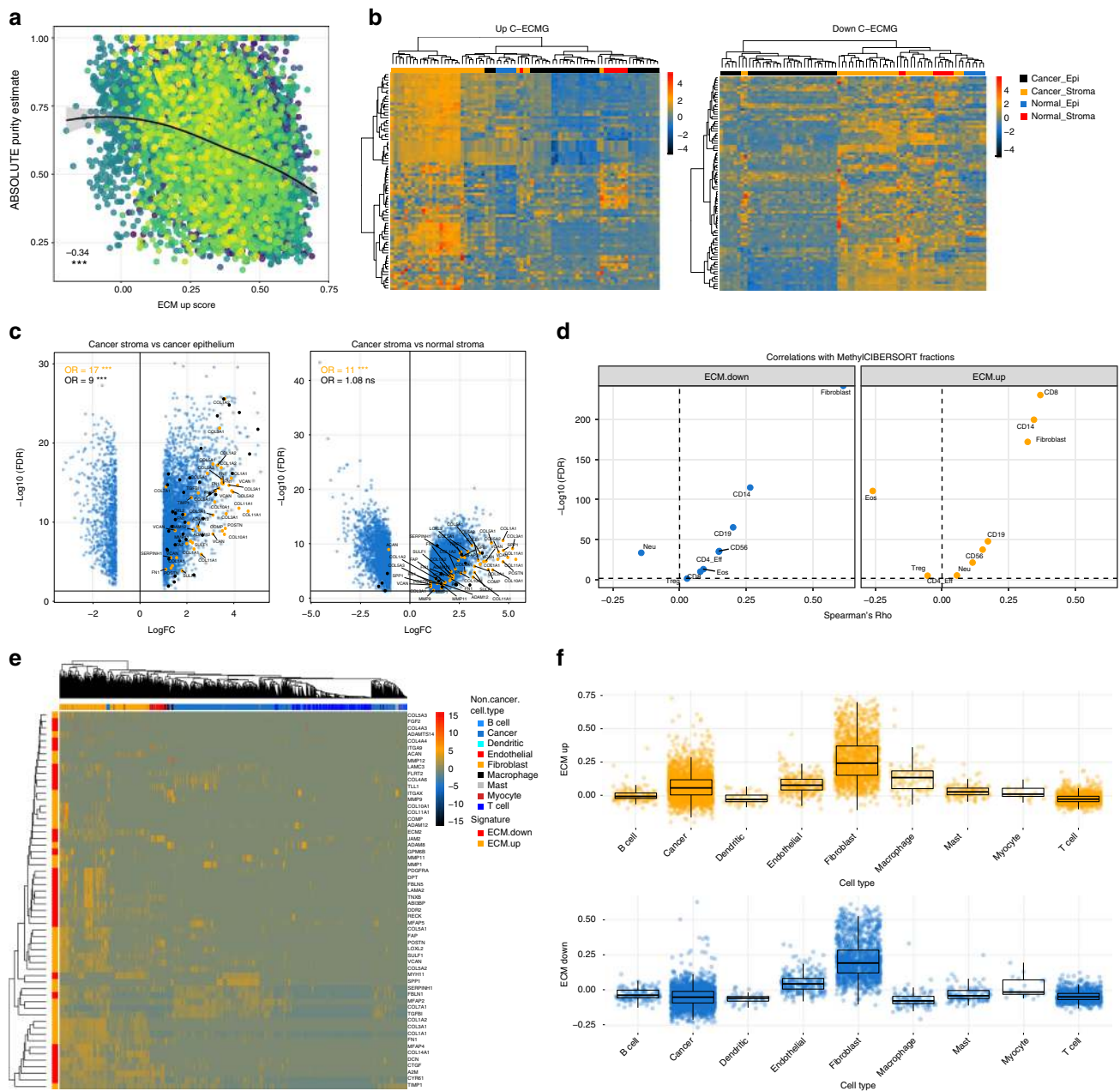


Fig. 2 C-ECM transcription is associated with stroma, especially CAFs. **a** ABSOLUTE purity estimates are inversely correlated with C-ECM-up score, suggesting stromal origin; colours represent cancer types, number shows Spearman's Rho ($n = 8128$). **b** Heatmaps of C-ECM-up and -down signatures projected onto microdissected epithelium and stroma from ovarian cancers. Rows show expression z-scores, samples are in columns. Annotation bars indicate tissue type. **c** Volcano plots show C-ECM genes (upregulated in orange, downregulated in black) differentially expressed between cancer stroma and epithelium and between cancer and normal stroma. **d** Volcano plots showing Spearman's correlations between MethylCIBERSORT cell-type fractions and C-ECM scores. **e** Heatmap of C-ECM genes in single-cell HNSCC RNA-seq data ($n = 5902$). **f** CAFs show the highest expression of C-ECM genes relative to other cell types in single-cell HNSCC data. Numbers on scatterplots indicate Spearman's Rho, asterisks indicate statistical significance. $***p < 0.001$. On all volcano plots, y axis = $-\log_{10}$ fold change, x axis = test statistic/fold change/Spearman's Rho. On volcano plots, all enrichment statistics are from Fisher's Exact Tests

Notably, the CAF specificity of C-ECM transcription was further indicated by the absence of expression in the leukocyte compartment in these data sets, suggesting that while monocyte enrichment and CD8 infiltration co-occur with C-ECM upregulation, they are not the source of C-ECM transcriptional changes (Supplementary Figure 2E).

Finally, as an ultimate test of a CAF origin, we examined a data set of single-cell transcriptomes from head and neck squamous cell carcinoma (HNSCC; GSE103322), which we selected owing

to the large collection of CAFs ($n = 1440$) profiled in the study²⁴. We found markedly higher expression of C-ECM genes in CAFs, which clustered together when the signature was projected onto the data set (Fig. 2e). In this data set, C-ECM-up and -down ssGSEA scores were significantly elevated in CAFs compared to other cell types (Fig. 2f), and this was also independently verified in an additional colorectal cancer single-cell RNA-seq data set, which contained a small number of CAFs (GSE81861, Supplementary Figure 2F)²⁵. Finally, we also performed statistical

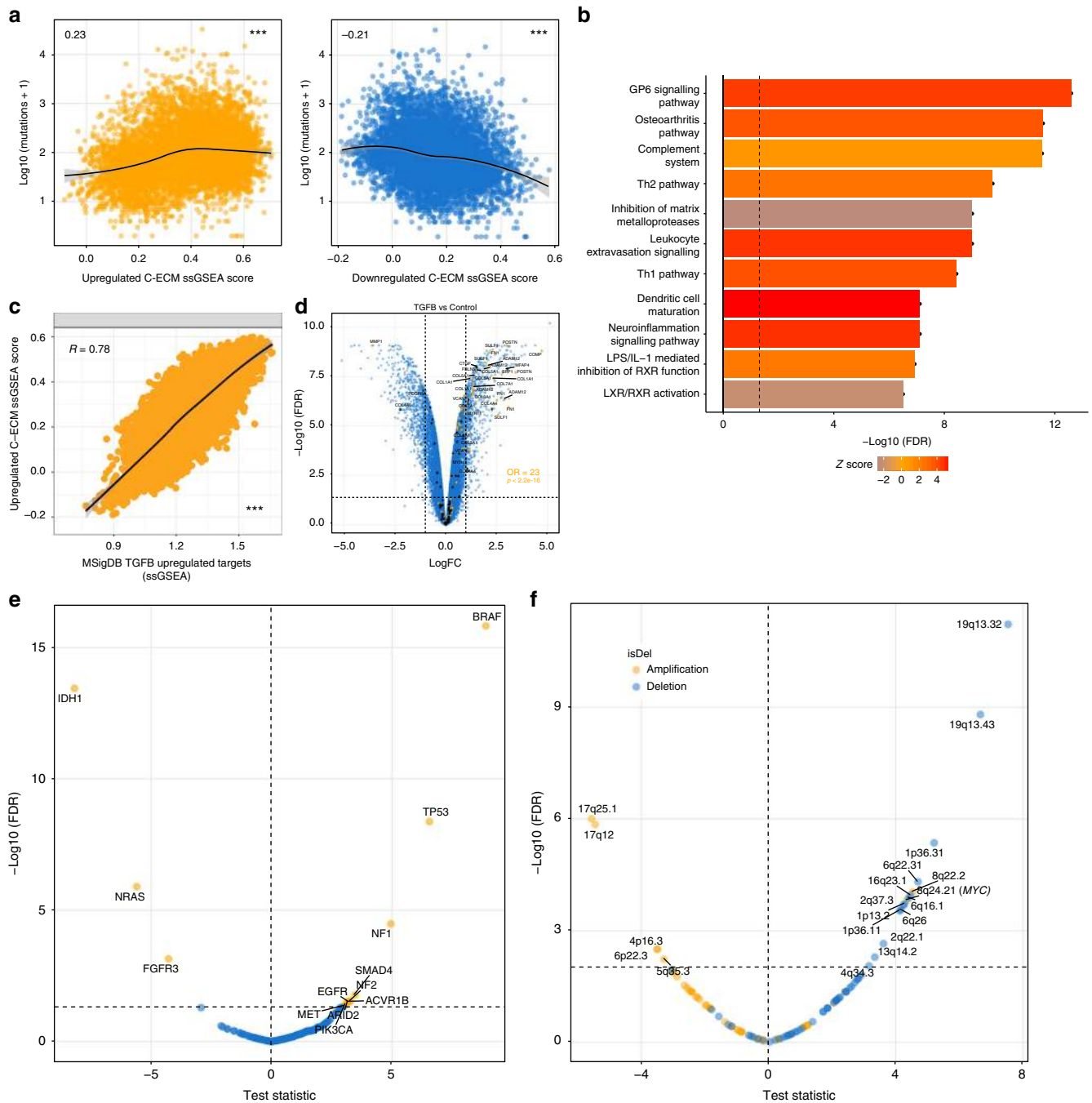


Fig. 3 E-ECM scores are associated with immunologically hot tumours and TGF- β . **a** C-ECM-up and down scores are significantly associated with mutational burden across cancer types but in opposite directions. **b** Canonical pathway analysis shows activation of inflammatory/adaptive-immune pathways. **c** Correlation between C-ECM-up scores and enrichment for an aggregated list of upregulated non-C-ECM TGF-beta target genes across tissues and species from MSigDB. **d** Volcano plot showing enrichment for C-ECM genes in TGF- β -induced transcriptional changes in immortalised normal fibroblasts. Linear model *t*-statistics for candidate **e** mutational and **f** copy-number alterations associated with ECM-up ssGSEA scores, adjusted for tumour type, on volcano plots. Numbers on scatterplots indicate Spearman’s Rho, asterisks indicate statistical significance. ****p* < 0.001. On all volcano plots, y axis = $-\log_{10}$ fold change, x axis = test statistic/fold change/Spearman’s Rho. On volcano plots, all enrichment statistics are from Fisher’s Exact Tests

testing for CAF association for each C-ECM gene in the HNSCC scRNA-seq data set and found that 45 genes were significantly overexpressed (mean difference >0, false discovery rate (FDR) < 0.05, Wilcoxon’s Rank Sum Test) and 10 genes were significantly underexpressed in CAFs compared to other cell types (Supplementary Figure 2G).

Collectively, these lines of evidence suggest that C-ECM profiles are generated mainly through the modulation of

transcriptional profiles in CAFs as a general feature of malignancy.

C-ECM dysregulation is correlated with immunological activity. Given that C-ECM scores correlate with CD8 T cells and CYT (Fig. 2d and Supplementary Figure 2D, and the fact that C-ECM-up scores are adversely prognostic despite the positive prognostic

impact of CYT⁹, we postulated that the C-ECM-up score may be enriched in immunologically 'hot' tumours and may reflect an adaptive mechanism of immune evasion. Our subsequent analyses uncovered robust evidence for this association using multiple orthogonal approaches.

Initially, we discovered that C-ECM-up score was positively correlated with mutational burden ($Rho = 0.23$, $p < 2.2e-16$) while the down-signature was negatively correlated ($Rho = -0.21$, $p < 2.2e-16$) (Fig. 3a). Similarly, C-ECM scores and Class I neoantigen burden also showed concordant associations ($Rho = 0.21$ and -0.21 , $p < 2.2e-16$, Supplementary Figure 3A, Spearman's correlation).

Then we tested for associations between C-ECM scores and microsatellite instability, an immunotherapy response biomarker per se²⁶, and found significant associations (Supplementary Figure 3B).

Additionally, we assessed macrophage polarisation using CIBERSORT²⁷ and found that the ECM-up signature was associated with a greater fraction of M1 (immunoactive) relative to M2 (immunosuppressive) macrophages (Supplementary Figure 3C). Finally, we found that multiple immune checkpoints, including *IDO1*, *B7-H3* and *PD-L2*, were overexpressed in samples in the top quartile of the C-ECM-up score distribution relative to bottom quartile cancers after adjusting for tumour type (2 fold change (FC), $FDR < 0.01$, Supplementary Data 1), indicating the upregulation of adaptive resistance mechanisms to immune-cell-mediated destruction (Supplementary Figure 3D). Moreover, these themes were broadly reinforced by ingenuity pathway analysis (IPA) canonical pathway analysis (Supplementary Data 2), which identified enrichment for inflammatory processes and adaptive immune responses enriched in samples in the top quartile of the C-ECM-up score (Fig. 3b).

C-ECM dysregulation is linked to TGF- β activation. Next, since our data suggest that the C-ECM-up signature associated with the presence of CAFs and not with normal stroma (Fig. 2b, c), we endeavoured to find putative drivers responsible for this dysregulation. To do this, we divided samples into quartiles based on the C-ECM-up score and then performed linear modelling using limma-trend with cancer type as a covariate.

IPA causal network analysis, after restriction to candidate regulators which by themselves differentially expressed between C-ECM-up top and bottom quartiles, identified TGF- β as one of the most activated regulators (Supplementary Figure 3E) and upstream regulatory analysis further identified multiple *SMAD* transcription factors, *API* complex members that associate with *SMADs*²⁸ and *SMARCA4*²⁹ (Supplementary Figure 3F, Supplementary Data 3), all critical for TGF- β transcriptional responses as activated in c-ECM-up-high cancers.

Moreover, we also performed orthogonal analyses using TCGA reverse phase protein array (RPPA) data ($n = 4278$) and identified 13 differentially abundant peptides between upper and lower quartiles of the ECM-up score ($FC > 1.3$, $FDR < 0.01$, Supplementary Figure 3G). Notably, these included increased levels of fibronectin and PAI1, both ECM components, and a vast majority of these peptides have known prior associations in the literature with TGF- β (see Supplementary Table 5 for each reference), reinforcing the inference of activated TGF- β signalling.

Importantly, in our RNA-seq analyses, TGF- β is significantly overexpressed in upper quartile C-ECM-up cancers, along with multiple mediators of ECM deposition such as FGF family members (*FGF1*, *FGF18*), BMPs (*BMP1* and *BMP8A*) and the local sequestrators of TGF- β , *FBP1* and *LTBP1*. Further, we compiled a collection of TGF- β upregulated target genes across

various tissues and across species from MSigDB (C2 collection) while excluding C-ECM genes and found a very strong correlation between C-ECM-up scores and this set ($R = 0.78$, $p < 2.2e-16$, Fig. 3c), suggesting TGF- β activation, and not only presence, is correlated with the C-ECM-up score. As a direct empirical test of the hypothesis that C-ECM-up genes are induced by TGF- β activation in fibroblasts, we compared the expression profiles of TGF- β -treated immortalised ovarian fibroblasts (GSE40266)²² versus untreated controls.

This revealed marked enrichment for C-ECM-up genes among differentially expressed genes (DEGs) (Fig. 3d, $OR = 23$, $p < 2.2e-16$, Fisher's Exact Test), further strengthening the notion that C-ECM-up gene dysregulation is associated with TGF- β signalling in CAFs. However, C-ECM-down genes were not significantly enriched among downregulated genes following TGF- β stimulation, suggesting that C-ECM-up genes represent TGF- β activation in CAFs, while C-ECM-down genes represent more normal fibroblasts.

TGF- β is known to exert a multitude of effects in the tumour microenvironment; it is capable of stimulating fibrosis, inducing epithelial-mesenchymal transition and driving metastasis^{30,31}. In contrast, it can also induce tumour-suppressive cytoskeleton³². We therefore tested whether distinct genomic adaptations disproportionately occurred in cancers with high C-ECM-up scores and whether these had known roles in TGF- β signalling.

We initially defined a set of candidate driver genes with evidence of positive selection in cancer and carried out linear modelling regressing the presence of nonsilent mutations against C-ECM-up score with cancer type as a covariate to identify genomic events that permitted adaptation to activation of the TGF- β pathway³³. This uncovered multiple notable candidates, including *TP53*, *SMAD4*, *BRAF*, *ACVR1B* and *NFI2* (Fig. 3e).

We also repeated the analysis using a compendium of 111 GISTIC³⁴ copy-number peaks defined across TCGA and found 18 out of the 111 candidate peaks to be associated with C-ECM-up score, most notably *MYC* amplification at 8q24.3 (Fig. 3f). Most of these genomic events have been directly associated with TGF- β pathway signalling (Supplementary Table 6 lists the relevant literature).

C-ECM dysregulation predicts failure of PD-1 blockade.

Finally, we tested whether C-ECM dysregulation is an immune-evasion mechanism in the context of PD-1/PD-L1 blockade, where immunologically 'hot' tumours are associated with responses³⁵. We did not include CTLA4-blockade data because of the small number of patients and qualitative differences in the biology of PD-1 and CTLA4 blockade³⁶. Eligible data sets were required to contain RNA-seq data and matched mutation/class I major histocompatibility complex (MHC) counts from whole-exome data from pre-treatment biopsies.

In two out of three cohorts of PD-1-blockade-treated patients (two melanoma, one bladder)³⁷⁻³⁹, the C-ECM-up score was significantly higher in progressors (Fig. 4a, $p < 0.05$, Wilcoxon's Rank Sum Test). This was also true in pooled logistic regression accounting for cancer type, CYT, mutational load, a T cell-inflamed signature¹¹, cohort, antibody and prior anti-CTLA4 treatment (Fig. 4b).

Next, comparing prediction performance using logistic regression with 0.632 bootstrapping⁴⁰ showed that models with C-ECM ssGSEA scores significantly outperformed those involving CYT, a T cell-inflamed signature and mutation load alone (Fig. 4c, S4A). This is notable given that the latter three factors have been proposed as biomarkers for patient stratification for immunotherapy. Moreover, the aggregate score is comparable to a random forest fit with individual C-ECM genes, suggesting that

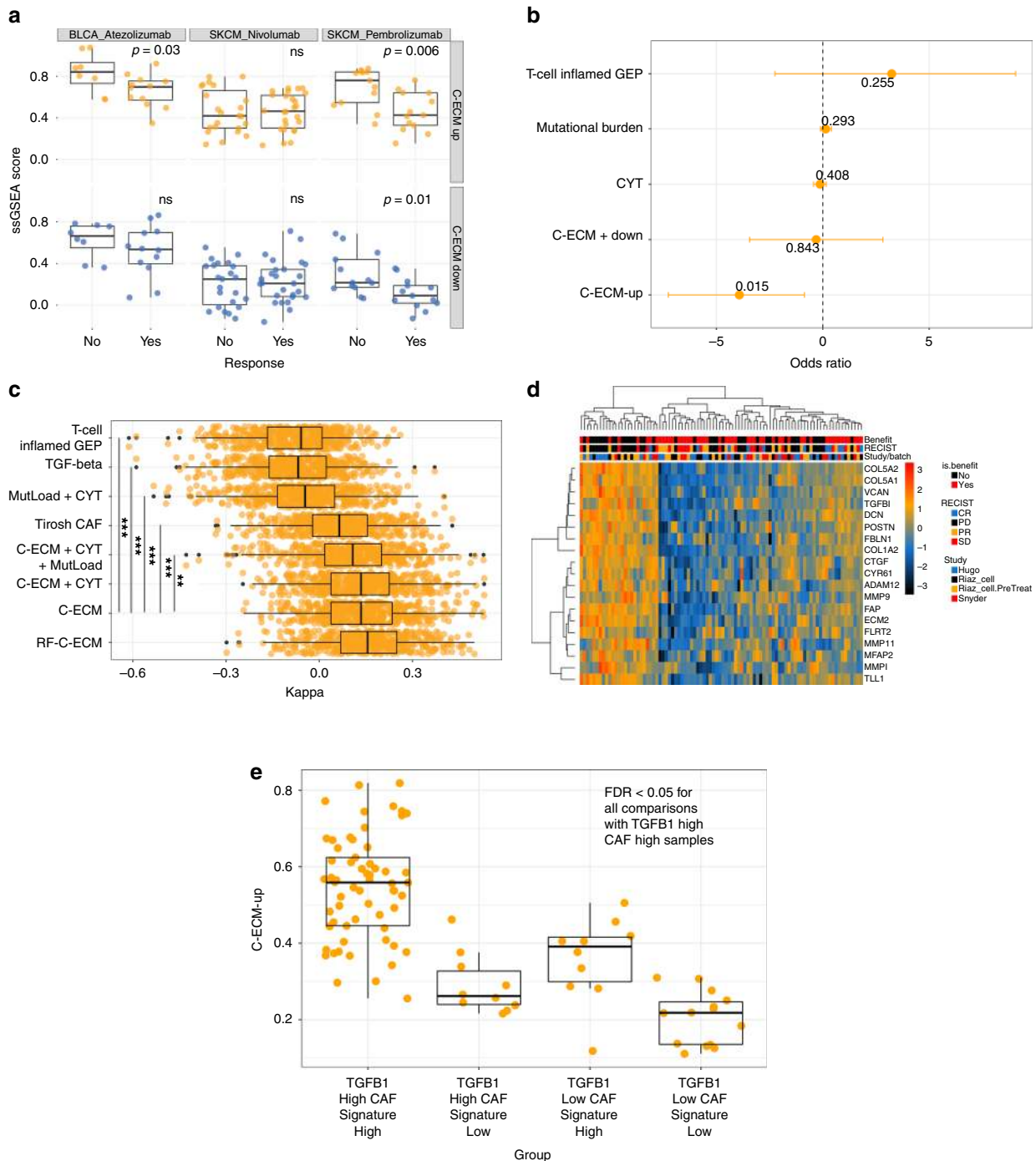


Fig. 4 C-ECM scores predict failure of PD-1 blockade. **a** Boxplots showing distributions of C-ECM ssGSEA scores across multiple data sets of pretreatment biopsies from patients treated with PD-1 blockade. Non-responders = progressive disease (PD). *p* Values are from Wilcoxon’s Rank Sum Test. **b** Coefficients from pooled logistic regression analysis evaluating various predictors on PD-1-blockade response, error bars indicate 95% confidence intervals. **c** Boxplots of Cohen’s Kappa from 0.632 bootstrapping (500 resamples), showing that ECM-based models outperform other candidate biomarkers. Asterisks show *q*-values. **d** Heatmap showing C-ECM genes differentially expressed between ICB responders and nonresponders after controlling for study-specific variation. Rows show z-scores of expression (log2 CPM) and columns show samples; annotation ribbons display clinical information. **e** Boxplot showing C-ECM-up score distributions based on discretised categories of *TGFB1* expression and the Tiresh CAF signature. All comparisons with the *TGFB1* high, CAF high group were significant at FDR < 0.05 (Wilcoxon’s Rank Sum Test)

these genes may be largely co-regulated. Finally, restricted hypothesis testing using limma-trend found 19 C-ECM genes overexpressed at FDR < 0.1 (Fig. 4d) between responders and nonresponders, defining a practical signature for clinical application (Supplementary Figure 4B)

C-ECM profiles independently predict PD-1-blockade failure. Our results were consistent with three models that we went on to further resolve. In one, *TGFB1/TGFB2* expression is associated with immunotherapy failure and there is no direct link with CAFs and the C-ECM programme. In the second, the presence of CAFs

alone is sufficient and TGF- β -induced C-ECM gene expression does not add predictive value. In the final model, the C-ECM programme is activated by TGF- β in CAFs and is the causal determinant of immunotherapy failure.

To differentiate between these models, we applied the pooled logistic regression approach described earlier, as well as the 0.632 bootstrapping approach, to estimate model performance in an unbiased manner.

In pooled analyses, *TGFB1* and *TGFB2* expression by themselves did not predict immunotherapy failure (Supplementary Figure 4C, middle plot). A CAF signature derived from single-cell analyses of melanomas¹⁰ was used to estimate CAF abundance and this score was significantly associated (coefficient = -2.61, $p = 0.03$) (Supplementary Figure 4C, bottom plot), suggesting that CAF abundance plays a role.

However, this effect was stronger for the C-ECM-up score in the C-ECM-based model (coefficient = -4.03, $p = 0.01$), and when C-ECM scores, the CAF signature score and *TGFB1/2* expression are all included in a joint model, the CAF score loses significance while the C-ECM signature shows even stronger associations (coefficient = -8.47, $p = 0.007$) (Supplementary Figure 4C, top plot), consistent with the predictive value of CAFs being mediated through an association with the C-ECM-up signature. In accordance with these observations, in our machine learning evaluation, *TGFB1/TGFB2*-based models and the CAF-signature-based models are significantly outperformed by C-ECM-based models (Fig. 4c). Collectively, these results suggest that the C-ECM-up programme per se, and not simply CAF abundance or TGF- β activation in the tumour microenvironment in general, is associated with PD-1 blockade failure.

As a final test of the hypothesis that both TGF- β and CAFs are required for induction of the C-ECM signature, we binned samples from our immunotherapy data into low and high groups (first quartile versus all other quartiles) for *TGFB1* expression and the Tirosh CAF ssGSEA score. We predicted that CAF-high *TGFB*-high cancers would display the highest C-ECM-up score, that CAF-low *TGFB*-low cancers would display the lowest enrichment, while the other two combinations would display significantly lower expression than CAF-high *TGFB*-high cancers because both factors are necessary but insufficient alone. Statistical comparisons (pairwise Mann-Whitney tests, FDR correction) confirmed our prediction (Fig. 4e).

Discussion

In this study, we have identified a pan-cancer role for ECM dysregulation. Using a range of approaches, we show that ECM dysregulation is associated with the presence of CAFs and the activity of TGF- β , among other regulators. Our model fits a scenario where the C-ECM-down signature represents normal/normal-like fibroblasts, whereas the TGF- β -driven C-ECM-up signature identifies a poor prognosis CAF activation phenotype that is upregulated pan-cancer. We further show that tumours that are otherwise 'immunologically hot' tend to display higher levels of the C-ECM-up programme and possess genomic alterations that may minimise the fitness costs of the high levels of TGF- β signalling required to sustain the C-ECM-up programme. These statistical associations serve as suggestive early evidence supporting the continued experimental investigation of TGF- β in immune modulation across tissue types.

Given this, the depletion of CAFs may be a potential approach to enhance checkpoint blockade; however, in some cases, CAF depletion per se is paradoxically associated with worse outcomes⁴¹. This suggests approaches that seek to normalise the aberrant transcriptome in fibroblasts, possibly through TGF- β blockade, are likely to offer a more promising route to boost the

efficacy of checkpoint blockade. Consistent with this, recent preclinical studies have uncovered evidence that simultaneous targeting of both TGF- β and PD-L1 can result in markedly better tumour control in multiple mouse models⁴². Moreover, recent studies using genetically reconstituted murine colorectal cancers⁴³, and those based on transcriptional analyses of bladder cancers treated with PD-L1 blockade⁷, have arrived at similar conclusions and have demonstrated that TGF- β blockade can markedly potentiate antitumour immunity by modulating CAF-mediated T cell-exclusion phenotypes.

Our pan-cancer analyses establish the presence of a TGF- β -associated C-ECM programme that is correlated with immunotherapy failure and indicate the existence of subsets of C-ECM-high cancers independent of tissue type. Relevantly, prior work has suggested that negative selection in cancer is generally weak as a pan-cancer phenomenon, suggesting that the depletion of immunogenic mutations through immunoeediting is weak³³. The widespread occurrence of the C-ECM programme across cancer types may help explain, at least partially, why this is the case.

Our results suggest that tumours with elevated C-ECM programme across cancer types may likely benefit from combination immunotherapy with PD-1 blockade and TGF- β blockade to restore immune control of tumours. Future tissue-agnostic, C-ECM biomarker-guided basket trials are necessary to validate this hypothesis.

Finally, our findings linking TGF- β activity, CAFs, C-ECM signature and immunosuppression of otherwise immune 'hot' tumours are associative and further experiments would be required to prove a causal link.

Methods

General statistical procedures. All tests were two sided and were performed for unmatched samples measured once.

Analysis of ECM signatures in cancer. RNA-seq data were obtained in the form of RSEM estimates from SAGE Synapse. This was reduced to 15 cancer types that had >10 normal samples available and low expressed genes were filtered (average <1 CPM).

Limma-trend⁴⁴ was then used to compute DEGs between normal and cancer samples while controlling for cancer type at thresholds of FDR < 0.01, 2FC. Controlling for the cancer type was explicitly used in order to minimise the effects of class imbalance and also to derive a picture of ECM gene dysregulation independent of cancer type.

A reference gene set of extracellular matrix genes was derived using the gene ontology term 'extracellular matrix' (GO:0031032) biological process category (used by UniprotKB) and downloaded from geneontology.org. Gene ontology terms and enrichments in overlap were computed using Fisher's Exact Test.

Significant upregulated and downregulated ECM genes were used to define gene sets whose expression was summarised using ssGSEA¹⁵ for each sample. Clinical data were then derived from SAGE synapse for a large subset of these tumours (Accession: syn7343873) and ECM-up and -down scores were divided into quartiles for categorisation. Validation at the protein level was performed using Spearman's Rank Correlation between RNA-seq and mass spectrometry for matched BRCA and OVCA samples profiled by TCGA and CPTAC, with the mass spectrometric data derived from the corresponding publications.

A stratified Cox regression model was of the form Survival ~ ECM-up-quartile + ECM-down-quartile + strata(cancer type) + strata(stage) was fit to estimate impact on prognosis. Similarly, a combined analysis was performed by fusing ECM-up and -down quartile categories.

Microenvironmental analyses of ECM dysfunction. ABSOLUTE¹⁸ was used to derive purity and ploidy estimates for $n =$ cancers. ECM scores were tested for association with purity using Spearman's Rank Correlations. This analysis was then repeated using purity estimates from ASCAT¹⁹ derived from COSMIC and the median VAF of mutations was obtained from the MC3 MAF (Synapse: syn7214402)⁴⁵.

An ovarian cancer data set was then obtained from GEO (GSE40595) and normalised using fRMA⁴⁶. C-ECM genes were then projected onto the data set to visualise clustering by phenotype. Significant probes were then identified for cancer epithelium versus cancer stroma and cancer stroma versus cancer epithelium using limma. Fisher's Exact Tests were used to test for enrichment of ECM signatures among DEGs (defined using limma).

Cellular deconvolution analyses were carried out using estimated cellular fractions produced using MethylCIBERSORT using Spearman's Rank Correlations between the ECM-up and -down scores. These findings were validated using gene expression of associated markers (*ACTA2*, *CD14*, *CD8A* and *CYT*) using Spearman's Rank Correlation. Validation of CAF association was performed in single-cell RNA-seq data from HNSCC, selected owing to the availability of large numbers of CAFs ($n = 1440$) (GSE103322) by projecting ECM signatures onto $\log_2(\text{TPM} + 1)$ transformed matrices. Data were obtained in a TPM matrix from GEO and the cell-type allocations were taken from the corresponding publication, wherein known cellular markers were used to define populations. ssGSEA scores were computed and Mann–Whitney tests were used to test for differential expression.

Additional validation was performed using count data for colorectal cancer single cells from GSE81861. Flow-sorted samples and colorectal cancer xenograft data were obtained from GEO data sets (GSE39397 and GSE35144)^{6,7} and heatmapting was used to visualise associations between C-ECM gene expression and cell type.

Immunological correlates of ECM dysfunction. HLA type for MHC class I alleles was retrieved from The Cancer Immunome Atlas⁴⁷. Topiary (<https://github.com/openvax/topiary>) was then used in combination with the MC3 set of TCGA mutation calls (Synapse:syn7214402) to define variants and potential neoantigens, using nonamers and thresholds of binding affinity <500 nM. Mutation loads were also computed for the same set of samples.

Spearman's correlation and negative binomial models were used to test for associations between these genomic counts and ECM scores after adjusting for cancer type. M1/M2 macrophage fractions were computed using CIBERSORT²⁷.

The expression of immune checkpoints was computed as part of whole-transcriptome analyses from differential expression between the top and bottom quartiles of the ECM-up score, adjusting for cancer type. Mismatch repair analyses were carried out using TCGA-allocated status based on Sanger sequencing published previously⁴⁸.

Multiplatform correlates of ECM dysfunction. For RNA-seq and RPPA data, limma-trend was used to compute differentially regulated genes/proteins at thresholds of 2FC, FDR < 0.01 and 1.3FC, FDR < 0.01 by comparing samples in the top and bottom quartiles of ECM-up score, respectively.

IPA was used to perform a core pathway analysis using interactions experimentally verified in human tissues in general. Causal network analysis results were restricted to those themselves differentially expressed, while upstream regulatory analyses were not.

For mutation analyses, we retrieved a previously published catalogue of driver mutations³³ (positively selected) in cancer and carried out linear regression against ECM-up score with cancer type for a covariate. For copy-number data, we generated GISTIC³⁴ calls pan-cancer at intensity thresholds of 0.3 and FDR < 0.01. Events for each GISTIC peak were dichotomised into altered/wild type and were regressed against ECM-up scores using a linear model with cancer type as covariate.

Analysis of immunotherapy data sets. We assembled RNA-seq, mutation and neoantigen burden data from multiple immunotherapy data sets and standardised the expression data to consider transcripts quantified in all data sets. Then we computed ssGSEA scores for the C-ECM-up and -down scores for each data set and examined association with response (partial response/complete response/stable disease versus no response) by Response Evaluation Criteria in Solid Tumours criteria using Mann–Whitney tests. A pooled logistic regression was conducted using the C-ECM scores from the three data sets with cancer type and treatment as covariates.

Finally, 100 iterations of 0.632 bootstrapping⁴⁰ were carried out to evaluate the performance of logistic regression and Cohen's unweighted Kappa was used to evaluate model performance. For Random Forest-based analyses, we first subtracted study-specific effects using a linear model via the `removeBatchEffects` function in the `limma` R package. Mann–Whitney tests were used to compare model performance across resamples. The models fit were C-ECM + *CYT*, C-ECM alone, *TGFB1* alone, Mutational load + *CYT* and C-ECM, *CYT* and mutational load, a CAF signature derived from Tirosh et al.¹⁰ and an interferon-induced signature¹¹ (and in each case, the study/cohort was included as a covariate). Limma-trend was used to identify ECM genes differentially expressed between nonresponders and responders using restricted hypothesis testing at FDR < 0.1. Finally, for analysis of mechanistic hypotheses surrounding the relative contributions of CAFs, *TGF-β* and the C-ECM-up score to immunotherapy failure, we fit the following logistic regression models. (i) Response ~ *TGFB1* + *TGFB2* + $\log_2(\text{MutLoad})$ + T cell-inflamed GEP score + Cohort, (ii) Response ~ Tirosh CAF signature score + $\log_2(\text{MutLoad})$ + T cell-inflamed GEP score + Cohort, (iii) Response ~ C-ECM-up score + Tirosh CAF signature score + *TGFB1* + *TGFB2* + $\log_2(\text{MutLoad})$ + T cell-inflamed GEP score + Cohort. To test whether both *TGFB1* and CAFs were required for activation of the C-ECM-up signature, we defined the lowest quartile of the Tirosh CAF signature score, and the lowest quartile of *TGFB1* expression, as low and the rest as high. Pairwise Mann–Whitney

tests were computed between the four groups produced thus with Benjamini–Hochberg correction for FDR.

Code availability. Knit HTML R markdowns of the code used to generate the results are available on Zenodo at <https://doi.org/10.5281/zenodo.1410639>.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

TCGA data used in this paper are available from SAGE Synapse (<https://www.synapse.org>) at accessions syn7343873, syn4311114, syn4303551 and syn7214402. Single-cell RNA-seq and gene expression microarray data used in this study are available at GEO (<https://www.ncbi.nlm.nih.gov/geo/>) accessions GSE40595, GSE103322, GSE39397, GSE35144 and GSE81861. RData files of intermediate processed objects are available upon reasonable request from the authors.

Received: 1 June 2018 Accepted: 18 September 2018

Published online: 08 November 2018

References

- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Pickup, M. W., Mouw, J. K. & Weaver, V. M. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* **15**, 1243–1253 (2014).
- Shintani, Y. et al. IL-6 secreted from cancer-associated fibroblasts mediates chemoresistance in NSCLC by increasing epithelial-mesenchymal transition signaling. *J. Thorac. Oncol.* **11**, 1482–1492 (2016).
- Richards, K. E. et al. Cancer-associated fibroblast exosomes regulate survival and proliferation of pancreatic cancer cells. *Oncogene* **36**, 1770–1778 (2017).
- Moffitt, R. A. et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
- Isella, C. et al. Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47**, 312–319 (2015).
- Calon, A. et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47**, 320–329 (2015).
- Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
- Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
- Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
- Ayers, M. et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Investig.* **127**, 2930–2940 (2017).
- Naba, A. et al. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteomics* **11**, M111.014647 (2012).
- Mertins, P. et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
- Zhang, H. et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* **166**, 755–765 (2016).
- Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
- Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
- Palumbo, A. Jr. et al. Extracellular matrix secreted by reactive stroma is a main inducer of pro-tumorigenic features on LNCaP prostate cancer cells. *Cancer Lett.* **321**, 55–64 (2012).
- Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
- Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. USA* **107**, 16910–16915 (2010).
- Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
- Li, B. & Li, J. Z. A general framework for analyzing tumor subclonality using SNP array and DNA sequencing data. *Genome Biol.* **15**, 473 (2014).
- Yeung, T. L. et al. *TGF-β* modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res* **73**, 5016–5028 (2013).
- Chakravarty, A. et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat. Commun.* **9**, 3220 (2018).

24. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611.e24–1626.e24 (2017).
25. Li, H. & Courtois, E. T. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).
26. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
27. Newman, A. M., Liu, C. L. & Green, M. R. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
28. Verrecchia, F. et al. Smad3/AP-1 interactions control transcriptional responses to TGF-beta in a promoter-specific manner. *Oncogene* **20**, 3332–3340 (2001).
29. Xi, Q., He, W., Zhang, X. H., Le, H. V. & Massague, J. Genome-wide impact of the BRG1 SWI/SNF chromatin remodeler on the transforming growth factor beta transcriptional program. *J. Biol. Chem.* **283**, 1146–1155 (2008).
30. Caja, L., Dituri, F., Mancarella, S. & Caballero-Diaz, D. TGF-beta and the tissue microenvironment: relevance in fibrosis and cancer. *Int. J. Mol. Sci.* **19**, E1294 (2018).
31. Drabsch, Y. & ten Dijke, P. TGF-beta signalling and its role in cancer progression and metastasis. *Cancer Metastasis Rev.* **31**, 553–568 (2012).
32. Siegel, P. M. & Massague, J. Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer. *Nat. Rev. Cancer* **3**, 807–821 (2003).
33. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029.e21–1041.e21 (2017).
34. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
35. Tumeh, P. C. et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
36. Wei, S. C. et al. Distinct cellular mechanisms underlie anti-CTLA-4 and anti-PD-1 checkpoint blockade. *Cell* **170**, 1120.e17–1133.e17 (2017).
37. Snyder, A. & Nathanson, T. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: an exploratory multi-omic analysis. *PLoS Med.* **14**, e1002309 (2017).
38. Hugo, W. et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35–44 (2016).
39. Riaz, N. et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934.e15–949.e15 (2017).
40. Efron, B. & Tibshirani, R. Improvements on cross-validation: the .632+ bootstrap method. *J. Am. Stat. Assoc.* **92**, 548–560 (1997).
41. Özdemir, B. C. et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with diminished survival. *Cancer Cell* **25**, 719–734 (2014).
42. Lan, Y. & Zhang, D. Enhanced preclinical antitumor activity of M7824, a bifunctional fusion protein simultaneously targeting PD-L1 and TGF-beta. *Sci. Transl. Med.* **10**, eaan5488 (2018).
43. Tauriello, D. V. F. et al. TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* **554**, 538–543 (2018).
44. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
45. Ellrott, K. et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst.* **6**, 271.e7–281.e7 (2018).
46. McCall, M. N., Bolstad, B. M. & Irizarry, R. A. Frozen robust multiarray analysis (fRMA). *Biostatistics* **11**, 242–253 (2010).
47. Charoentong, P. et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **18**, 248–262 (2017).
48. Cortes-Ciriano, I., Lee, S., Park, W. Y., Kim, T. M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. *Nat. Commun.* **8**, 15180 (2017).

Acknowledgements

A.C. is supported by The Guglietti Fellowship in Tumor Immune Therapy from the Princess Margaret Cancer Foundation. Work in D.D.D.C.'s laboratory is supported by grants from the Cancer Research Society (CRS19092 and CRS19091), Canadian Cancer Society (CCSRI 703279 and CCSRI 703716), NSERC (489073), Ontario Institute for Cancer Research (OICR) with funds from the province of Ontario, J.P. Bickell Foundation, the Helen M Cooke professorship from Princess Margaret Cancer Foundation, Canada Research Chair (950-231346), the CIHR New Investigator Salary award (201512MSH-360794-228629) and CIHR Foundation Grant (FDN 148430). P.B. is funded by Ohlson Research Initiative, University of Calgary. N.P.B. and L.K. are funded by O'Brien Centre Summer Studentships at the University of Calgary.

Author contributions

A.C., L.K. and N.P.B. performed data analyses, A.C., P.B. and D.D.C. conceived the project, P.B. and D.D.D.C. supervised the project and A.C. and D.D.D.C. wrote the paper with input from P.B. All authors have read and approved the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-06654-8>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018