Deconstruction of a metastatic tumor microenvironment reveals a common matrix response in human cancers

Oliver M T Pearce[*1], Robin Delaine-Smith[*1,2], Eleni Maniati[*1], Sam Nichols[**1], Jun Wang[**1], Steffen Böhm[1], Vinothini Rajeeve[1], Dayem Ullah[1], Probir Chakravarty[4], Roanne R. Jones[1], Anne Montfort[1], Tom Dowe[1], John Gribben[1], J. Louise Jones[1], Hemant M Kocher[1], Jonathan S Serody[5], Benjamin G Vincent[5], John Connelly[6,3], James D. Brenton[7], Claude Chelala[1], Pedro R. Cutillas[1], Michelle Lockley[1], Conrad Bessant[8], Martin Knight[2,3], Frances R Balkwill[1,3 ***]

*These authors contributed equally

** These authors contributed equally

*** Corresponding author and lead contact f.balkwill@qmul.ac.uk

1. Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, London, ECIM 6BQ, UK
2. School of Engineering and Materials Science, Queen Mary University of London, Mile End, London E1 4NS, UK
3. Institute of Bioengineering, Queen Mary University of London, Mile End, London E1 4NS, UK
4. Bioinformatics Core, The Francis Crick Institute, 44 Lincoln's Inn Field, London WC2A 3LY, UK
5. UNC Lineberger Comprehensive Cancer Centre, Chapel Hill, North Carolina, USA
6. Blizard Institute, Queen Mary University of London, Newark St, London E1 2AT, UK
7. Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK
8. School of Biological and Chemical Sciences, Queen Mary University of London, Mile End, London E1 4NS, UK

Abstract

We have profiled, for the first time, an evolving human metastatic microenvironment, measuring gene expression, matrisome proteomics, cytokine and chemokine levels, cellularity, ECM organization and biomechanical properties, all on the same sample. Using biopsies of high-grade serous ovarian cancer (HGSOC) metastases that ranged from minimal to extensive disease, we show how non-malignant cell densities and cytokine networks evolve with disease progression. Multivariate integration of the different components allowed us to define for the first time, gene and protein profiles that predict extent of disease and tissue stiffness, whilst also revealing the complexity and dynamic nature of matrisome remodeling during development of metastases. Although we studied a single metastatic site from one human malignancy, a pattern of expression of 22 matrisome genes distinguished patients with a shorter overall survival in ovarian and twelve other primary solid cancers, suggesting that there may be a common matrix response to human cancer.

Keywords

Tumor microenvironment, extracellular matrix, biomechanics, cytokines, tumor-infiltrating leukocytes, ovarian cancer, cancer-associated fibroblasts, data integration

Significance

Conducting multi-level analysis with data integration on biopsies with a range of disease involvement identifies important features of the evolving TME. The data suggest that despite the large spectrum of genomic alterations, some human malignancies may have a common and potentially targetable matrix response which influences the course of disease. (50 words)

Introduction

Solid tumors consist of malignant cells surrounded and infiltrated by a variety of host cells that are recruited and 'corrupted' by the cancer, aiding its growth and spread (1,2). A dynamic network of soluble factors, cytokines, chemokines, growth factors and adhesion molecules drive the interactions between malignant and non-malignant cells to create this tumor microenvironment (TME) (3,4). The TME network stimulates extracellular matrix (ECM) remodeling, expansion of abnormal vascular and lymphatic networks and migration of cells into and out of the tumor mass (5,6). Solid tumors are also typically stiffer than the surrounding tissue due to aberrant ECM deposition and organization that has a major influence on cell and tissue mechanics (7,8).

While the TME is of critical importance during initiation and spread of cancer, relatively little is known about its evolution or the relationship between the molecular mechanisms of disease progression and higher-order features such as the extent of disease, non-malignant cell density and tissue stiffness. Studies on molecular mechanisms of human cancer have mainly focused on large-scale genomic and transcriptomic analysis of primary tumors (9) and the immune cell landscape (1). Human cancer evolution is also now being studied in multiple metastatic sites e.g.(10,11) but mainly in terms of the genomics of the malignant cells. Also, most of these analyses focus on one stage of a cancer.

Here, for the first time, we have used multi-layered TME profiling of a metastatic site, omental metastases of high-grade serous ovarian cancer (HGSOC), to identify molecular changes that predict the higher-order TME features. Our study differs from other genomic and transcriptomic studies in two important ways: first, we have integrated data from six different TME parameters from each metastatic sample studied and second, we have studied the evolution of metastases by including samples that vary in the extent of disease.

HGSOC is one of the most lethal of the peritoneal cancers: less than 30% of patients currently survive more than five years after diagnosis with little improvement in overall survival in the past 40 years (12). As poor prognosis is mainly due to early dissemination into the peritoneal cavity (12,13) and HGSOC metastases have a complex TME (14), there is a need for an integrated understanding of its different components (12). We chose to study the omental TME because it is the most frequent site for HGSOC metastases and is routinely resected during debulking surgery.

Using samples ranging from minimal to extensive disease we conducted cellular, biomechanical and molecular analyses on each biopsy. Integration of the different components using multivariate analyses allowed us to define for the first time, gene and protein profiles that predicted extent of disease and tissue stiffness whilst also revealing how the ECM is remodeled during metastases development. Of particular interest was an ECM-associated molecular signature, that we termed the matrix index, that predicted both extent of disease and tissue stiffness in our sample set. This novel signature distinguished patients with shorter overall survival not only in ovarian cancer, but also in twelve other cancer types irrespective of patient age, stage or response to primary treatment, suggesting a common matrix response to human primary and metastatic cancers.

Results

Study Design

We analysed six different parameters of omental biopsies from 36 HGSOC patients: the extent of disease, densities of non-malignant cells, tissue mechanics, cytokines, matrisome protein and RNA profiles (Figure 1A). The samples ranged from uninvolved or minimally-diseased omentum to biopsies with extensive disease (Table S1, Figure 1B and Figure S1A). The extent of disease in each biopsy was measured by digital histopathology on haematoxylin and eosin (H&E) stained sections and was calculated as the percentage of tissue area occupied by malignant cells and stroma. We termed this the 'disease score'. Remodeling of the omentum was extensive when malignant cells were present and the malignant cells comprised a minor proportion of the tissue (Figure 1B). In order to monitor for any significant changes in sample architecture during tissue processing for the different analyses we took serial sections for H&E staining. We did not observe any major changes in disease score between the different areas analyzed.

The density of the major non-malignant cell populations was measured by immunohistochemistry (IHC) and digital histopathology in the same specimens. The biomechanical properties of the tissues were measured using a mechanical indentation methodology (15) that gave us the tissue modulus of each sample.

Twenty-nine cytokine and chemokines were measured in protein lysates using an electro-chemiluminescence assay (Table S2). For proteomic analysis of the same biopsies we focused on the ECM and associated molecules using a method that enriches whole tissue lysates for the matrisome protein compartment (16). Using this technique we detected 145 proteins (Table S3). The term 'matrisome' is defined as the ensemble of all core ECM proteins (collagens, proteoglycans, glycoproteins) and associated molecules (the secretome, ECM-regulators and ECM-affiliated molecules) of tissue extracellular matrices (17). After

alignment and filtering, RNA sequencing identified 15,441 protein-coding genes (Table S4). We then used univariate analyses and a multivariate regression method – partial least squares (PLS) (18) – to model the relationships between these different components of the metastases (Figure 1C).

The relationship between cell density and disease score

Using a tissue microarray constructed from the biopsies we quantified the adipocytes, fibroblasts and leukocytes which were the major non-malignant components in the specimens and related this to disease score. The area occupied by adipocytes, the major cell type of the normal omentum, decreased with disease score (Figure 1B) and this was in part due to a reduction in the diameter of the adipocytes (Figure 2A) which may reflect research showing that adipocytes can provide energy for ovarian cancer cell growth (13). Using α-smooth muscle actin (α-SMA) and α-fibroblast activation protein (α-FAP) as markers of cancer-associated fibroblasts (CAFs) (19), we assessed the area of the tissue occupied by α-SMA+ and α-FAP+ cells and found a strong positive correlation with disease score for both markers (Figure 2B, C).

We then stained and counted six major leukocyte subtypes and plotted cell density against disease score. In all cases there was a significant positive correlation between leukocyte density and disease score ($p<0.001$) (Figure 2D, Figure S1B). Densities of T cells with the surface markers CD3, CD4, CD8 and CD45RO strongly correlated with each other ($p<0.001$, $r>0.6$) but CD68+ macrophage density only weakly correlated with the other leukocytes ($p<0.05$, $r<0.5$) (Figure 2E, Table S5).

Therefore, as metastases developed in the omentum, the fatty tissue was replaced by a combination of fibroblasts, lymphocytes and macrophages. The cellular composition that we observed illustrates the changes from a normal omental tissue, primarily composed of

adipocytes with minimal immune cell infiltrate and little fibroblastic reaction, to heavily diseased tissues with profound tumor-associated inflammation and a large increase in all types of leukocytes and fibroblasts. However, whilst there was a general increase in stromal cell density across all of the markers we studied, we observed larger variance with high disease score samples for all immune cell counts. This is to be expected as it is already well-documented that the TME of advanced HGSOC biopsies ranges from sparse to dense leukocyte infiltration.

The immune cell densities significantly correlated with their corresponding immune gene expression signatures extracted from the associated RNAseq data for each sample (Table S6) and levels of the adipogenic transcription factor PPARγ mRNA declined with disease score (Figure S1C). Thus, the cell density scores were validated by the gene transcription data.

Leukocyte density and cytokine networks in the TME

Next, we correlated leukocyte density against levels of 29 different cytokines and chemokines in protein lysates of the metastases. There were eight significant correlations (Figure 2F, Table S7), the strongest of which was, unexpectedly, an association between IL16, a chemoattractant and modulator of T cell function, and the density of CD3, CD45RO and CD8 cells. These correlations became stronger with the 10 samples with the highest disease score (Figure S1D). IHC revealed IL16 protein in both malignant and stromal areas, with a higher density in the former (Figure 2G). There was also a high positive correlation between global cell proliferation assessed by Ki67 and LTA, IL17A, IL15, CXCL10 (Figure 2F).

As cytokine networks are major determinants of leukocyte density and phenotype in the TME (3,20), we asked if the cytokine proteins and genes we detected in the tissue lysates could inform us about the networks that regulate omental metastases. We constructed

heatmaps showing pairwise comparisons of cytokine protein and gene transcription levels (Figure 2H, Figure S1E, Tables S8 and S9). Overall the protein gene correlation was 30%, in line with other studies e.g. (21). The heatmaps show five significant co-expressions at both gene and protein level: IL6 with IL1A, IL1B, and IL8, CSF2 with IL8, and CCL4 with CCL3. IL6 was of particular interest as we previously identified this as a major mediator of cytokine networks in ovarian cancer (20,22). Finally, we asked if levels of any of the cytokines and chemokines associated with disease score. There were weak but significant associations with disease score with IL12B, IL13, IL16, VEGF, CCLs 11, 26, and CXCL10.

These results suggest that malignant cell-derived cytokine and chemokine networks in the omental metastases regulate leukocyte density and overall proliferative index. Unexpectedly, we identified the CD4 ligand IL16 as a potential major mediator of the leukocyte infiltrate. It is interesting that increased tissue and serum levels of IL16 have been reported during tumor development in laying hen models of ovarian cancer and in ovarian cancer patients (23).

These cytokine and cellularity data confirm and extend previous research on the ovarian cancer and other tumor microenvironments and we believe validate our approach. Extensive study of another TME component, the matrisome, has recently become possible through proteomics methods that focus on these proteins (17). There is currently little information on how the matrisome evolves with disease progression. Therefore, our next aim was to study the ECM-associated proteins and genes, collectively termed the matrisome, in the same biopsies.

How the matrisome changes with disease progression

Using our matrisome-focused proteomic technique (16) and the RNAseq data we quantitatively assessed matrisome proteins and genes. In terms of relative mass ratios, the major matrix proteins in samples with the lowest disease score were collagen 1, 6 and 3, the glycoprotein fibrillin, the ECM regulator alpha-2-macroprotein, and the basement membrane proteoglycans lumican and heparin sulphate proteoglycan-2. In contrast, biopsies with the highest disease score had an expansion of ECM-glycoproteins fibrinogen and fibronectin, as well as increases in proteoglycans, secreted factors, and affiliated proteins, (FDR <0.1) (Figure 3A).

Extending the analysis to the entire sample set we found that as disease score increased levels of some matrisome proteins decreased and others increased. Comparing the relative mass ratio of all matrisome proteins with disease score, we found that 18 proteins decreased and 49 proteins increased with disease progression (Figure 3B, Table S10). After these univariate analyses, we used the multivariate regression PLS method to rank genes and proteins according to their influence on disease score, and a permutation-derived threshold was applied to determine those that were most strongly associated with disease score (24,25). Of these, 58 proteins ranked top in PLS modeling of disease score ($r^2 = 0.70$), defining a matrisome protein signature of disease score (Figure 3C, Table S11).

412 of the 764 matrisome genes detected in our transcriptomics dataset also predicted disease score (Table S12). The top 60 genes are shown in Figure 3D with 27 ECM-associated molecules predicting disease score at both the gene and protein level (Figure 3E, Figure S2A). We used IHC to detect four of these proteins, FN1, COMP, CTSB and COL11A1, in HGSOC omentum detecting all four within stromal regions (Figure 3F). As collagen organisation strongly influences cell behavior and tissue mechanics (26,27), we utilised two-photon microscopy to visualise collagen fibres using second harmonic generation (SHG) label-free illumination (Figure 3G). In low disease score tissues collagen fibres were thin and

arranged mostly around the adipocytes. In high disease score tissues, there were denser arrays of long collagen bundles with an apparent micro-scale orientation preference. Collagen orientation correlated strongly with disease score (Figure 3G). It should be noted that Figure 3A shows relative matrisome protein abundance as mass ratios whereas Figure 3G depicts the alignment of collagen in the tissues. While the amount of collagen does increase in diseased tissue, its relative abundance goes down as other matrisome proteins are induced as the tissue becomes more diseased. Representative images for COL1A1 IHC staining are shown in Figure S2B.

Other biological processes associated with disease score

We then analysed the RNAseq data to find other biological processes associated with disease score (Table S13). Significantly associated pathways included cell metabolism, adhesion, communication as well as ECM organization and immune response pathways (Figure S2C, Table S14).

We have described here, for the first time, how the matrisome changes with disease progression. As some of the strongest correlations with disease score were found with these ECM-associated proteins and genes, and increased tissue stiffness has been linked with tumor progression (28,29), we next investigated how the changes in matrisome genes and proteins related to the biomechanical properties of the biopsies.

Relationships between tissue modulus (stiffness) and disease score

We used mechanical indentation (15) to determine tissue modulus (which describes material stiffness independent of sample size) and the stress-relaxation behavior of the samples. We measured disease score from histological sections of the area of the specimen that was

indented (Figure 4A, Figure S3). Biopsies with a high disease score displayed a non-linear loading response and greater stress relaxation while there was a relatively linear loading response in low disease score tissue (Figure 4B, Figure S3C, Table S15). Tissue modulus values in high disease score biopsies were one-two orders of magnitude higher than in low disease biopsies. There were significant positive correlations between tissue modulus and malignant cell area, the stromal area and disease score of each biopsy (Figure 4C, Figure S3D). Thus, there was a significant log relationship between tissue modulus and disease score in the evolving TME suggesting a close association of tissue stiffness with disease progression.

The matrisome, tissue stiffness and disease score

Using the PLS method, we identified 64 matrisome proteins, mainly glycoproteins, that accurately predicted tissue modulus ($r^2 = 0.69$) (Figure 4D, Figure S4A, Table S16). We then used 764 matrisome genes detected by RNAseq and identified 405 that predicted tissue modulus (Figure 4E, Figure S4B, Table S17) of which 38 also featured as proteins in Figure 4D. Thus, as with disease score, the tissue modulus could be predicted by a subset of ECM-associated genes and proteins of the matrisome.

We also modeled tissue modulus against the entire transcriptome of the metastases (Figure S4C, Table S18). Genes associated with cell metabolism, cell communication, wound healing, ECM organization, as well as development, correlated with tissue modulus (Figure S4D, Table S19). Figure 4F shows the PLS prediction plot and the top 50 genes from this signature. As expected there was a strong overlap with disease score-associated genes and proteins (74% and 75% respectively) and these were significantly associated with tissue modulus (Table S20).

Collectively, the experiments described above demonstrated the complexity and dynamic nature of matrisome evolution during development of HGSOC metastases and the close relationship between tissue stiffness and extent of disease.

A subset of matrisome molecules models both disease score and tissue modulus and has prognostic significance in HGSOC

We next asked how many matrisome genes and proteins significantly defined both disease score and tissue modulus in our sample set (Figure 5A, Table S21). Twenty-two molecules were highly significant across all of our analyses with a gene:protein concordance of 68% (Figure 5A, Figure S5A). Thirteen of the 22 proteins had documented protein:protein interactions (Figure 5B). Using the ChEA database (30) we found that the 22 genes shared a range of common transcription factors including RUNX2, STAT3, SMAD4, WT1, JUN and TP53. These reflect pathways associated with Wnt signalling pathway, inflammation and osteogenesis, whilst TP53 is of course the most frequently mutated gene in HGSOC, a major genetic driver of the disease (Figure S5B and Table S22).

Using these 22 most significant molecules, we measured the ratio between the mean expression levels of the positively regulated genes and the mean expression levels of the negatively regulated genes. We termed this the matrix index because these molecules are all components of the matrisome (17). As would be expected, the matrix index of each sample significantly correlated with disease score and tissue modulus in our set of samples ($p<0.0001$) as would be expected (Figure 5C). There were also significant positive and negative correlations between matrix index and immune cell signatures in the corresponding RNAseq data (Figure 5D, Table S23), notably Treg and Th2 cell signatures, cell subtypes associated with tumor promotion and immune suppression (31) and a modest statistically significant relationship between disease score and entropy as a measure of clonal abundance

for T and B cells (Table S23). These data suggest that matrix index molecules may influence expansion of specific infiltrating cell populations. In support of these findings, there were significant linear correlations between the matrix index and the data in Figure 2 in terms of CD4+ (pearson r=0.523, p=0.001) and FOXP3+(pearson r=0.52, p=0.001) cells but there was no correlation between matrix index and CD8+ cells (pearson r=0.29, p=0.094).

Relevance of matrix index to other stages of HGSOC and prognosis

As the matrix index positively correlated with disease score, tissue modulus and immune-suppressive signatures in our sample set, we wondered if it would distinguish ovarian cancer patients with a poorer prognosis in transcriptomic data from untreated primary tumors. We extracted expression values from two publicly available HGSOC gene expression datasets and calculated the matrix index for each sample. The high and low matrix index groups were determined using a method described previously (32). High matrix index significantly correlated with shorter overall HGSOC patient survival in both the ICGC and TCGA gene expression datasets, as well as in our original sample set (Figure 5E, Figure S5B-E). To test that the clinical outcome association of the matrix index was not a random finding we conducted 200,000 simulations and found that the association was significantly above that expected from random signatures.

In order to account for the higher relative abundance of tumor cells compared with stroma present in the TCGA ovarian cancer samples compared to our samples, we plotted the correlation between matrix index and % tumor cell or % stroma in each TCGA sample. In both cases we observed no association with matrix index (Figure S5F). Taken together with the immune cell correlations, this further suggests the matrix index is not only a measure of the tissue remodeling that accompanies HGSOC, but is also a measure of a matrisome composition that better supports tumor progression.

Interrogating the TCGA ovarian cancer dataset, we next evaluated the power of the matrix index against nine other well-known prognostic gene expression signatures in ovarian and other cancers, including signatures for stromal and immune responses (33-41). In terms of hazard-ratio scores, matrix index was in the top three after the 26-gene breast cancer stromal signature reported by Finak *et al* (41) and the 193-transcriptional signature from TCGA (9) (Figure 5F, left panel). However, using multivariate analysis, matrix index was the single significant predictor of ovarian cancer survival independently of age, stage, grade and treatment outcome (Figure 5F, right panel and Table S24).

At the protein level, we used matrix index to examine the recently released TCGA/CPTAC ovarian cancer proteomics dataset. Whilst the study was not focused on detecting ECM proteins, which requires matrisome protein enrichment prior to analysis, as described above, we found that there were 12 proteins from the matrix index with a significant association with survival (10 with $p < 0.05$ and a further 2 with $p < 0.1$).

Matrix Index in other human cancers

ECM remodeling is a common feature of many human cancers and significant desmoplasia and ECM deposition is found in other solid tumors. Since we hypothesize that the matrix index is a measure of a tumor-promoting matrisome in HGSOC, we wondered if it may also be a feature associated with poor outcome in other cancer types. We calculated matrix index values in 30 other publicly available gene expression datasets from epithelial, mesenchymal and haematologic malignancies analysing data from 9215 human cancer biopsies. High matrix index was an indicator of poor prognosis in epithelial and mesenchymal cancers but not in haematological cancers, melanoma and glioblastoma (Figure 6A and Figure S6A). Using univariate analysis, high matrix index predicted shorter overall patient survival in 15 datasets representing 13 major cancer types ($p < 0.05$) (Figure S6B, Table S25). The range of

matrix index values across all these cancer databases had a median value close to 1.0 (Figure S6C). We believe this provides further evidence that the pattern of ECM-associated gene expression determined by the matrix index may be a common feature of human cancers. Remarkably, multivariate analysis showed that the prognostic value of the matrix index was independent of age, stage, grade and response to primary treatment in 15 of the datasets representing 13 major cancer types ($p < 0.05$) (Figure 6B).

Using IHC, we confirmed the presence of four of the upregulated matrix index proteins FN1, COL11A1, CTSB, and COMP, in three tissue microarrays from triple negative breast cancer (TNBC), pancreatic ductal adenocarcinoma (PDAC), and diffuse large B-cell lymphoma (DLBCL) (Figure 6C). These cancers reflected the range of hazard ratios for high matrix index in Figure 6B. Digital microscopy analysis showed the highest staining level in TNBC (Figure 6D), in keeping with the matrix index score for this cancer (Figure S6C). FN1, COMP, and CTSB were present in stroma and fibroblastic cells of all tumors. COL11A1 was located within the malignant cells in all biopsies. FN1 was also found in malignant PDAC cells and in immune cells in DLBCL. CTSB was located in macrophages in TNBC and PDAC, and tumor cells in DLBCL.

Data resource

All data in this paper are provided in a mine-able web-based resource http://www.canbuild.org.uk currently under construction. Users are able to download, visualize, analyse and integrate across datasets.

Discussion

In this paper we have profiled, for the first time, an evolving human metastatic microenvironment, using analysis that includes gene expression, matrix proteomics,

cytokine/chemokine expression, ECM organization and biomechanical properties, all performed on the same sample. This gives a unique and informative snapshot of the evolving metastatic state of one type of ovarian cancer. Integration of the most significant features of this microenvironment may have identified a matrix response that is conserved in other cancers of epithelial or mesenchymal origins.

Our study has also shown that conducting multi-level analysis with data integration on well-characterized cancer biopsies with a range of disease involvement, and multiple analyses per sample, can identify important features representative of the evolving TME. This approach is complementary to 'omic' molecular cancer datasets that have larger numbers of samples. The data presented here provide a unique resource regarding molecular, cellular, and mechanical regulation in the tumor microenvironment and a template for bioengineers who are building complex tumor microenvironment models.

Molecular genetics has revealed great intra- and inter-tumor heterogeneity. It is now accepted that malignant cell clones undergo Darwinian evolution, resulting in a high level of molecular heterogeneity. In contrast, this study shows that the interactions between malignant cells and the host to remodel the tissue matrix may be more consistent. It is already known that high lymphocyte density is a common indicator of good prognosis at different stages of disease in many malignancies including HGSOC (14). We suggest that another common feature of TMEs may be patterns of matrisome genes and proteins and that these also have prognostic significance.

The up-regulated genes that were most significantly related to disease score in our analysis, *COL11A1, COMP, VCAN, FN1, COL1A1* and *CTSB* have all been associated with cancer progression, poor prognosis and malignant cell invasion in ovarian and/or other cancers (42-49). For example, fibronectin promotes ovarian cancer invasion and metastasis through an α5β1-integrin/c-Met/FAK/Src-dependent signaling pathway (44) and *COL11A1* and *VCAN*

feature in a 10-gene poor prognostic signature of collagen-remodeling genes regulated by TGF-β signaling in ovarian cancer (45). More recently gene expression of COL11A1 showed a positive association with poor prognosis in several epithelial cancers (49). Importantly, the matrix index appears to correlate with certain immune cell signatures that are also known to influence prognosis. For example, the matrix index does not positively correlate with CD8 molecular signatures, which are associated with good prognosis in HGSOC, but significantly correlates with Treg and Th2 signatures.

We have not focused on proteases in this study. However, in our samples, normal tissue ECM stained highly positive and evenly for COL1A1, but COL1A1 staining in diseased sections, whilst still strong, appeared more uneven. In particular there was reduced staining around malignant cell areas. This may be due to expression of proteases such as MMP13, which degrade collagen structures (50), that we identified in our PLS analysis. Other matrix remodeling proteases that we identified such as MMP7, MMP11, CTSB, and ST14 are also able to degrade collagens, albeit to a lesser extent, but may also be capable of degrading matrix proteoglycans and glycoproteins (51,52). In addition to protease activity, the apparent reduction in COL1A1 by mass ratio (Figure 3A) was in part due to the relative increase in matrisome complexity in diseased tissues. Also, we used a modified matrisome analysis method which increased our protein detection coverage, at the cost of underestimating the absolute amount of large fibrous core matrisome proteins such as COL1A1.

As we have shown, there was a core group of matrisome molecules that best predicted tissue modulus. This appears to be through an expansion of matrisome glycoproteins and proteoglycans (Figure 3A) and a reorganization of fibrillar collagens (Figure 3G). The expansion of the glycoprotein and proteoglycan compartment increases the potential for post-translational modification within the extracellular space which could significantly alter the mechanical properties of the tissue, particularly through glycosylation and cross-linking (53).

For example, glycosaminoglycans on proteoglycans contribute to hydration, which contributes to tissue stiffness (54). Additionally, specific ECM cross-linking molecules, including the pro-lysyl hydroxylases which cross-link matrix proteins through collagen-like peptides, were associated with increasing disease in our samples (55). Molecules like COL11A1 and COMP are normally only present in stiffer tissues such as cartilage and bone, while VCAN plays a role in the morphogenesis of these stiffer tissues. In addition, collagens 11 and 6 play roles in collagen fibril organization and matrix integrity.

Tumor mechanics have a profound effect on fibroblasts and cancer cells and can promote tumor progression and metastasis (56). Stiffening of ECM creates a feed-forward self-reinforcing loop that contributes to the activation state of the fibroblast (57). Elevated mechano-signalling in PDAC cells as a result of elevated tissue stiffening, promotes tumor progression and aggression (8). In breast tumors, the stiffest regions are located at the invasive margins and tumors harboring the stiffest regions are the most aggressive (58). There is a significant desmoplastic response in most solid tumors, but given that there is large intra- and inter-tumor heterogeneity it is important to understand why our index of matrisome gene expression defines patients with poor prognosis in multiple human cancers. The matrix index does not seem to be simply a measure of the amount of desmoplasia or stromal component that accompanies cancer growth. We believe that it is a measure of a type matrisome composition that is more able to promote tumor growth. We found a strong association between the density of α-SMA and α-FAP positive cells, two markers commonly associated with activation of cancer associated fibroblasts, and disease score and there are several examples in the literature of poor prognostic fibroblast, desmoplastic, wound healing and stromal signatures in individual cancer types e.g. (39,59). As fibroblasts are the predominant matrix producing cells in many tissues this may, at least partially, explain the commonality of the matrix index in different cancers. It is also interesting that some of the

matrix index molecules we found to be down-regulated as disease increases (LAMB1, LAMC1, LAMA4, COL15A and HSPG2) are associated with the basement membrane which is vital for maintaining tissue homeostasis.

Malignant cell response to tumor-associated fibrosis, and the stromal cell phenotypes that contribute to ECM deposition, can vary within and between major cancer types. This was shown in great detail recently in a study of experimental and human pancreatic cancers where a distinct malignant cell genotype modulated the fibrotic phenotype of the tissue and pathology (8). This does not argue against our finding because we have found the matrix index is variable between different cases of each cancer.

As the predictive power of the matrix index was independent of age, stage and response to primary treatment, we suggest that the pattern of change in the matrisome may reflect increased propensity of the malignant cells to establish metastases. It is intriguing that five of the six up-regulated ECM genes in our matrix index (*COMP, VCAN, FN1, COL1A1* and *CTSB*) are typical of pre-metastatic niches (60). Another explanation for the association with poor prognosis could be that this configuration of ECM molecules prevents infiltration or effector function of host anti-tumor immune cells. A stiffened matrix can compromise T-cell antigen presentation and proliferation as well as Th1-cell differentiation (61). In addition, the ECM acts as a reservoir for angiogenic factors and is important for migration of endothelial cells during neo-angiogenesis and vascular remodeling seen in cancer (62).

Many of the matrix index molecules described above circulate systemically as fragments from protease remodeling, sometimes as neo-epitopes (63). Therefore, further investigation of the matrix index may have potential as a cancer diagnostic/prognostic blood test.

If we have identified a common and especially detrimental signature of the tumor-associated matrisome, then agents that could target or reconfigure the cancer matrisome could have wide

applicability in solid cancers and may enhance the action of immunotherapies, especially given the association of high matrix index with immunosuppressive T cell signatures. Some molecules of the matrix index may also prove good targets for drug delivery to the tumor site. A recent study demonstrated collagen 1 targeting of an anti-EGFR mAb showed increase therapeutic efficacy (64). Targeting matrix index molecules which are not as ubiquitous as collagen 1 may provide a significant advancement to such strategies.

Methods

Ovarian cancer patient samples

Patient samples were kindly donated by women with high-grade serous ovarian cancer (HGSOC) undergoing surgery at Barts Health NHS Trust between 2010 and 2014. Blood and tissue that was deemed by a pathologist to be surplus to diagnostic and therapeutic requirement were collected together with associated clinical data under the terms of the Barts Gynae Tissue Bank (HTA license number 12199. REC no: 10/H0304/14). Each patient gave written informed consent and the study was approved by a UK national review board. The studies were conducted in accordance with the Declaration of Helsinki and International Ethical Guidelines for Biomedical Research Involving Human Subjects (CIOMS).

RNA isolation and sequencing

Total RNA was extracted from 10 x 50 μm cryosections from frozen tissue sections and placed directly into the RLT Plus buffer (Qiagen) and rigorously vortexed. Samples were then processed using Qiagen RNeasy Plus Micro kit according to manufacturer's instructions. RNA quality was analyzed on agilent bioanalyzer 2100 using RNA PicoChips according to manufacturer's instructions. RNA integrity numbers (RIN) were between 8.1 and 9.9. RNA-Seq was performed by Oxford Gene Technology (Benbroke, UK) to ~42x mean depth on the

Illumina HiSeq2500 platform, strand-specific, generating 101bp paired-end reads, as previously described (65). The detailed methods regarding RNA sequencing and bioinformatic analysis are provided in the Supplementary Methods.

Quantitative Proteomics

The ECM component was enriched from frozen whole tissue sections (20 x 30 μm sections, approximately 40-50 mg of tissue) as previously described (17). The extracted proteins were reduced, alkylated and digested with trypsin. Peptides were separated by nanoflow ultra-high pressure liquid chromatography (UPLC, NanoAcquity, Waters) and analyzed by mass spectrometry using a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific). The detailed methods of ECM component enrichment, peptide preparation, mass spectrometry and bioinformatics analysis are provided in the Supplementary Methods.

Cytokine and chemokine analysis

Cytokine and chemokines were assayed using Mesoscale Discovery Platform (MSD SI2400) according to manufacturer's instructions. Cytokine panel 1(Human) K15050D, Proinflammatory panel 1(human) K0080087, and Chemokine panel 1(Human) K0080125 were used. Samples used were lysates from the ECM-enrichment protocol (described above). The amount of total protein used from each sample was between 1 and 3 μg.

Mechanical characterization

Mechanical characterisation was performed on whole tissues using a flat-punch indentation methodology on an Instron ElectroPuls E1000 (Instron, UK) equipped with a 10 N load cell (resolution = 0.1 mN) in order to measure the modulus of the tissue samples(15)(Delaine-

Smith et al., 2016). The detailed method of quantification is provided in the Supplementary Methods.

Histochemical analysis

Frozen tissues that were subsequently used for RNA, proteomics and cytokine analysis were cryosectioned to 8-10 µm slices. Sections were fixed in in 4 % paraformaldehyde (PFA) and stained with haematoxylin and eosin using standard methods. Tissues used in mechanical characterization were cut in half at the center of the tissue dye marked area and perpendicular to the direction of indentation while still frozen. Tissue was then fixed in 4 % PFA for 24 h and paraffin embedded and sectioned (8 µm) using standard procedures followed by H&E staining. All tissue sections were scanned using a 3DHISTECH Panoramic 250 digital slide scanner (3DHISTECH, Hungary) and the resulting scans were analysed using Definiens software (Definiens AG, Germany). Disease scores were determined firstly by manually defining regions of interest in the tissue that represented tumor, stroma, fat (adipocytes) or other (lymphatic structure) and then training the software to recognize these regions of interest. Disease score was expressed as a percentage of the whole tissue area that contained tumor and/or stroma (Figure 1B). Detailed methods of immunohistochemical analysis for quantification of immune cells, α-SMA and α-FAP positive cells, adipocyte diameters, ECM proteins and second-harmonic generation microscopy are provided in the Supplementary Methods.

RNA in situ hybridization

Chromogenic *in situ* hybridization for VCAN (Probe-Hs-VCAN, Cat No. 430071, Advanced Cell Diagnostics Inc. USA) was performed using the RNAscope 2.5 HD Detection Reagent

kit (Advanced Cell Diagnostics Inc.) according to the manufacturer's instructions (see Supplementary Methods).

Statistical and bioinformatics analysis

All statistical analyses and graphics were performed in the statistical programming language R (version 3.1.3). Detailed methodology for PLS regression models, Matrix index and its clinical association across cancer types is provided in the Supplementary Methods.

Accession Numbers

RNA-Seq data have been deposited in Gene Expression Omnibus (GEO) under the accession number GSE71340. Proteomic data are available via the PRIDE database accession number PXD004060.

Data availability

All of the primary data are deposited at http://www.canbuild.org.uk.

Disclosure of Potential Conflicts of Interest

The authors have no conflicts of interest to declare.

Grant Support

This project was funded by the European Research Council (ERC322566) and Cancer Research UK (A16354, A13034, A19694).

Acknowledgements

We thank Barts Trust Oncology Surgeons for sample provision and Prof. Kairbaan Hodivala-Dilke for useful discussion. We also thank Andrew Clear, Dr Joanne ChinAleong, Dr Prabhu Arumugam and Dr Sally Dreger for technical help with the tissue microarrays, George Elia and the BCI Pathology Core, Christof Smith and Dr Dante Bortone for help with bioinformatics analysis of the immune cell signatures and Dr Jackie McDermott for histopathological analysis of the TMA samples. Finally, we express our gratitude to the patients for donating the samples without which this work would not have been possible.

References

1. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* **2012**;12:298-306
2. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **2012**;21:309-22
3. West NR, McCuaig S, Franchini F, Powrie F. Emerging cytokine networks in colorectal cancer. *Nat Rev Immunol* **2015**;15:615-29
4. Crusz SM, Balkwill FR. Inflammation and cancer: advances and new agents. *Nat Rev Clin Oncol* **2015**
5. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* **2013**;19:1423-37
6. Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **2010**;141:52-67
7. Ingber DE. Mechanobiology and diseases of mechanotransduction. *Ann Med* **2003**;35:564-77
8. Laklai H, Miroshnikova YA, Pickup MW, Collisson EA, Kim GE, Barrett AS, *et al.* Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression. *Nat Med* **2016**
9. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. *Nature* **2011**;474:609-15
10. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **2012**;366:883-92
11. Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **2015**;521:489-94
12. Bowtell DD, Bohm S, Ahmed AA, Aspuria PJ, Bast RC, Jr., Beral V, *et al.* Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer* **2015**;15:668-79
13. Nieman KM, Kenny HA, Penicka CV, Ladanyi A, Buell-Gutbrod R, Zillhardt MR, *et al.* Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nature medicine* **2011**;17:1498-503
14. Nelson BH. New insights into tumor immunity revealed by the unique genetic and genomic aspects of ovarian cancer. *Curr Opin Immunol* **2015**;33:93-100
15. Delaine-Smith RM, Burney S, Balkwill FR, Knight MM. Experimental validation of a flat punch indentation methodology calibrated against unconfined compression tests for determination of soft tissue biomechanics. *J Mech Behav Biomed Mater* **2016**;60:401-15
16. Naba A, Pearce OMT, Del Rosario A, Ma D, Ding H, Rajeeve V, *et al.* Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics. *J Proteome Res* **2017**;16:3083-91

17. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular & cellular proteomics : MCP* **2012**;11:M111 014647

18. Wold S, Ruhe,A., Wold,H. and Dunn,III,W.J. The collinearity problem in linear regression. the partial least squares approach to generalized inverses. *SIAM J Sci Stat Comput* **1984**;5:735–43

19. Kalluri R, Zeisberg M. Fibroblasts in cancer. *Nat Rev Cancer* **2006**;6:392-401

20. Kulbe H, Chakravarty P, Leinster DA, Charles KA, Kwong J, Thompson RG, *et al.* A Dynamic Inflammatory Cytokine Network in the Human Ovarian Cancer Microenvironment. *Cancer research* **2012**;72:66-75

21. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **2012**;13:227-32

22. Coward J, Kulbe H, Chakravarty P, Leader D, Vassileva V, Leinster DA, *et al.* Interleukin-6 as a Therapeutic Target in Human Ovarian Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2011**;17:6083-96

23. Yellapa A, Bitterman P, Sharma S, Guirguis AS, Bahr JM, Basu S, *et al.* Interleukin 16 expression changes in association with ovarian malignant transformation. *Am J Obstet Gynecol* **2014**;210:272 e1-10

24. Johansson D, Lindgren P, Berglund A. A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription. *Bioinformatics* **2003**;19:467-73

25. Mehmood T, Liland KH, Snipen L, Saebo S. A review of variable selection methods in Partial Least Squares Regression. *Chemometr Intell Lab* **2012**;118:62-9

26. Trappmann B, Gautrot JE, Connelly JT, Strange DG, Li Y, Oyen ML, *et al.* Extracellular-matrix tethering regulates stem-cell fate. *Nat Mater* **2012**;11:642-9

27. Delaine-Smith RM, Green NH, Matcher SJ, MacNeil S, Reilly GC. Monitoring fibrous scaffold guidance of three-dimensional collagen organisation using minimally-invasive second harmonic generation. *PLoS One* **2014**;9:e89761

28. Krouskop TA, Wheeler TM, Kallel F, Garra BS, Hall T. Elastic moduli of breast and prostate tissues under compression. *Ultrason Imaging* **1998**;20:260-74

29. Levental KR, Yu H, Kass L, Lakins JN, Egeblad M, Erler JT, *et al.* Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **2009**;139:891-906

30. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **2010**;26:2438-44

31. Singh M, Loftus T, Webb E, Benencia F. Minireview: Regulatory T Cells and Ovarian Cancer. *Immunol Invest* **2016**:1-9

32. Mihaly Z, Kormos M, Lanczky A, Dank M, Budczies J, Szasz MA, *et al.* A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. *Breast cancer research and treatment* **2013**;140:219-32

33. Bonome T, Levine DA, Shih J, Randonovich M, Pise-Masison CA, Bogomolniy F, *et al.* A gene signature predicting for survival in suboptimally debulked patients with overian cancer. *Cancer Res* **2008**;68:5478-86

34. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **2012**;489:519-25

35. Palmer C, Diehn M, Alizadeh AA, Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* **2006**;7:115

36. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **2013**;39:782-95

37. Yoshihara K, Tajima A, Yahata T, Kodama S, Fujiwara H, Suzuki M, *et al.* Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. *PLoS One* **2010**;5:e9615

38. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* **2015**;47:1168-78

39. Iglesia MD, Vincent BG, Parker JS, Hoadley KA, Carey LA, Perou CM, *et al.* Prognostic B-cell signatures using mRNA-seq in patients with subtype-specific breast and ovarian cancer. *Clin Cancer Res* **2014**;20:3818-29

40. Yoshihara K, Tsunoda T, Shigemizu D, Fujiwara H, Hatae M, Fujiwara H, *et al.* High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway. *Clin Cancer Res* **2012**;18:1374-85

41. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, *et al.* Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med* **2008**;14:518-27

42. Yeung TL, Leung CS, Wong KK, Samimi G, Thompson MS, Liu J, *et al.* TGF-beta modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res* **2013**;73:5016-28

43. Ruan H, Hao S, Young P, Zhang H. Targeting Cathepsin B for Cancer Therapies. *Horiz Cancer Res* **2015**;56:23-40

44. Kenny HA, Chiang CY, White EA, Schryver EM, Habis M, Romero IL, *et al.* Mesothelial cells promote early ovarian cancer metastasis through fibronectin secretion. *J Clin Invest* **2014**;124:4614-28

45. Cheon DJ, Tong Y, Sim MS, Dering J, Berel D, Cui X, *et al.* A collagen-remodeling gene signature regulated by TGF-beta signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clin Cancer Res* **2014**;20:711-23

46. Englund E, Bartoschek M, Reitsma B, Jacobsson L, Escudero-Esparza A, Orimo A, *et al.* Cartilage oligomeric matrix protein contributes to the development and metastasis of breast cancer. *Oncogene* **2016**

47. Suwiwat S, Ricciardelli C, Tammi R, Tammi M, Auvinen P, Kosma VM, *et al.* Expression of extracellular matrix components versican, chondroitin sulfate, tenascin, and hyaluronan, and their association with disease outcome in node-negative breast cancer. *Clin Cancer Res* **2004**;10:2491-8

48. Waalkes S, Atschekzei F, Kramer MW, Hennenlotter J, Vetter G, Becker JU, *et al.* Fibronectin 1 mRNA expression correlates with advanced disease in renal cancer. *BMC Cancer* **2010**;10:503

49. Jia D, Liu Z, Deng N, Tan TZ, Huang RY, Taylor-Harding B, *et al.* A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets. *Cancer Lett* **2016**;382:203-14

50. Leeman MF, Curran S, Murray GI. The structure, regulation, and function of human matrix metalloproteinase-13. *Crit Rev Biochem Mol Biol* **2002**;37:149-66

51. Eckhard U, Huesgen PF, Schilling O, Bellac CL, Butler GS, Cox JH, *et al.* Active site specificity profiling of the matrix metalloproteinase family: Proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses. *Matrix Biol* **2016**;49:37-60

52. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* **2014**;15:786-801

53. Barry-Hamilton V, Spangler R, Marshall D, McCauley S, Rodriguez HM, Oyasu M, *et al.* Allosteric inhibition of lysyl oxidase-like-2 impedes the development of a pathologic microenvironment. *Nat Med* **2010**;16:1009-17

54. Schaefer L, Schaefer RM. Proteoglycans: from structural compounds to signaling molecules. *Cell Tissue Res* **2010**;339:237-46

55. Gilkes DM, Bajpai S, Wong CC, Chaturvedi P, Hubbi ME, Wirtz D, *et al.* Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. *Mol Cancer Res* **2013**;11:456-66

56. Northey JJ, Przybyla L, Weaver VM. Tissue Force Programs Cell Fate and Tumor Aggression. *Cancer Discov* **2017**;7:1224-37

57. Calvo F, Ege N, Grande-Garcia A, Hooper S, Jenkins RP, Chaudhry SI, *et al.* Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat Cell Biol* **2013**;15:637-46

58. Acerbi I, Cassereau L, Dean I, Shi Q, Au A, Park C, *et al.* Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration. *Integr Biol (Camb)* **2015**;7:1120-34

59. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, *et al.* Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **2016**;166:755-65

60. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, *et al.* Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* **2015**;17:816-26

61. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep* **2014**;15:1243-53

62. Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. *J Cell Biol* **2012**;196:395-406

63. Leeming DJ, Karsdal MA, Byrjalsen I, Bendtsen F, Trebicka J, Nielsen MJ, *et al.* Novel serological neo-epitope markers of extracellular matrix proteins for the detection of portal hypertension. *Aliment Pharmacol Ther* **2013**;38:1086-96

64. Liang H, Li X, Wang B, Chen B, Zhao Y, Sun J, *et al.* A collagen-binding EGFR antibody fragment targeting tumors with a collagen-rich extracellular matrix. *Sci Rep* **2016**;6:18205

65. Bohm S, Montfort A, Pearce OM, Topping J, Chakravarty P, Everitt GL, *et al.* Neoadjuvant Chemotherapy Modulates the Immune Microenvironment in Metastases of Tubo-Ovarian High-Grade Serous Carcinoma. *Clin Cancer Res* **2016**;22:3025-36

Figure 1. Study design and sample description

a) Overview of the samples and the analyses conducted on the same tissue specimen. b) Bar plot shows results from digital analysis of architecture of haematoxylin and eosin (H&E)-stained samples based on percentage of malignant cell area (tumor), stroma, and adipocyte area, coloured blue, green and red respectively. The combined percentage area occupied by tumor and stroma was used to determine the 'disease score' of each sample. Each G number represents one sample. Upper microscope images show H&E staining of a biopsy and the same biopsy pseudo-coloured as malignant cell area (tumor) blue, stroma green and adipocyte area, red. Bottom images show four different H&E-stained samples representative of sample range with increasing disease score. c) Schematic of the PLS regression method used to define higher-order features of the tumor microenvironment from molecular components.

Figure 2. The cells of the TME change with disease score

a) Adipocyte diameter negatively correlated with increasing disease score. Top panel, microscope images representative of low (left) and high (right) disease score tissue sections (stained for α-SMA by IHC) showing adipocytes. Scale-bar corresponds to 100µm. Bottom left panel, scatter plot illustrating mean ± sd of digitally quantified adipocyte diameter (linear regression, N = 16, R^2 = 0.66, p = 0.0001). b) Correlation of α-SMA positive cells against disease score. Top panel, representative low (left) and high (right) disease score tissue sections stained for α-SMA by IHC. Scale-bar corresponds to 100µm. Bottom panel, quantification of α-SMA+ area % against disease score (linear regression, N = 30, R^2 = 0.83, p < 0.0001). c) Correlation of α-FAP positive cells against disease score. Top panel, representative low (left) and high (right) disease score tissue sections stained for α-FAP by IHC. Scale-bar corresponds to 100µm. Bottom panel, quantification of α-FAP+ area % against disease score (power regression, N = 32, R^2 = 0.77, p < 0.0001). d) Cleveland plots of immune cell counts against disease score (Spearman's correlation, N = 34). e-f) Heatmap of pairwise Pearson's correlation coefficients of e) immune cell counts (N = 34), f) MSD-quantified cytokine/chemokine correlations against immune cell counts (N = 32). h) IHC of IL16 in HGSOC omental biopsies. Scale-bars correspond to 100µm. g) Heatmap of pairwise Pearson's correlation coefficients of MSD-quantified cytokine/chemokine (N = 32).

Figure 3. Identification of matrisome proteins and genes that define tissue architecture
a) Matrisome data displayed as relative mass ratios. Top panels show individual matrisome proteins identified in low and high disease score tissue; bottom panels show the relative proportions of each of the major classes of matrisome proteins in lowest (N = 6) versus highest disease score (N = 10). b) Line graphs illustrating normalized protein abundance and local polynomial regression fitted trend lines of proteins that either decrease (top panel), or increase (bottom panel) with disease score. c) PLS-identified matrisome proteins and d) matrisome genes that define disease score. e) Scatter plot of gene and protein correlation with disease score, highlighted molecules denote significant correlations (Pearson's correlation, N = 33, $p < 0.05$). f) IHC staining for four matrisome proteins, FN1, COMP, CTSB, COL11A1 identified from PLS analysis as highly significantly related to disease score. Scale-bars correspond to 200µm. g) Collagen fiber alignment; top panel shows representative images of high and low disease score tissue sections visualised using second harmonic generation, and bottom panel, semi-quantification of fiber alignment from images plotted as number of fiber occurrences per angle bin (predominant fibre direction normalized to 0 degrees) with local polynomial regression fitted lines and disease color-coding.

Figure 4. Identification of molecular components that define tissue modulus

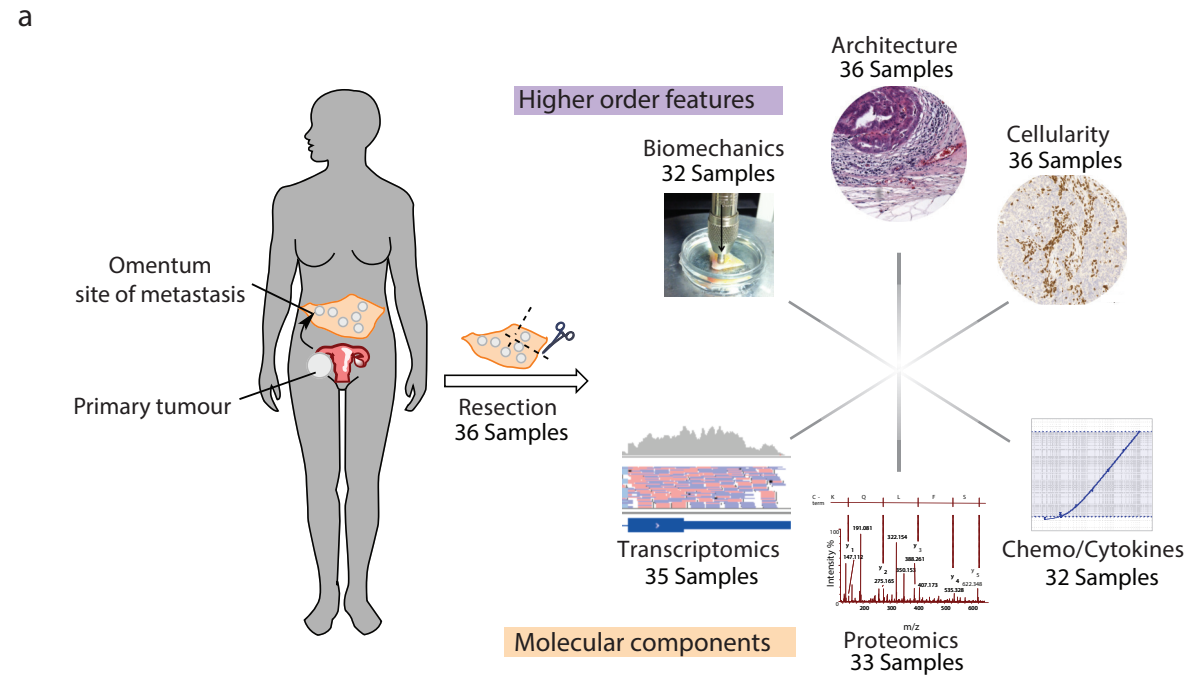
a) Orientation of flat-punch indentation showing representative low and high disease score samples stained with H&E, dashed line indicates tissue area analysed for determining disease score. b) Representative load-displacement curve from loading phase obtained from high and low disease score samples. c) Optimal tissue modulus correlated against combined % tumor plus stroma (disease score) ($N = 32$, $p < 0.05$). d-f) Cross-validation plot of measured versus predicted tissue modulus values (diagonal line represents measured = predicted) and heatmap of PLS-identified d) matrisome proteins, e) matrisome genes, and f) all coding gene components that describe tissue modulus. Heatmap columns correspond to individual samples ordered by increasing tissue modulus. ($N = 29$, 30 and 30, respectively). Rows ordered by decreasing model weight values.

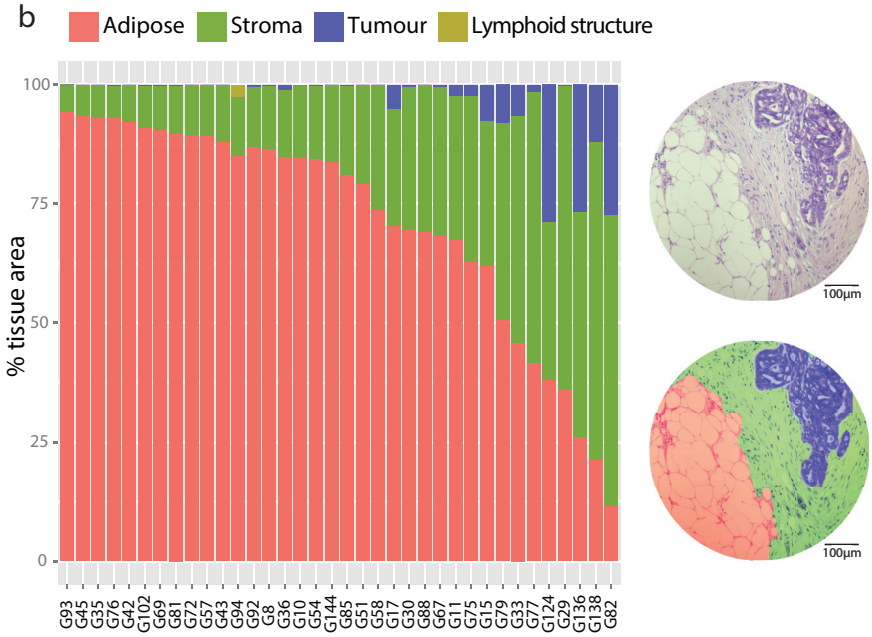
Figure 5. A matrix signature that predicts survival in ovarian cancer

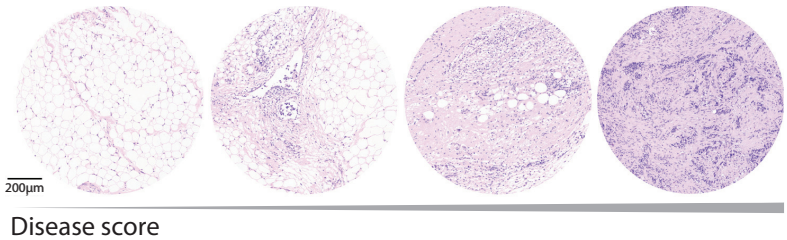
a) Venn diagram showing the overlap of PLS-identified molecules associated to tissue modulus and disease score (DS) at both gene and protein level. A total of 22 ECM-associated molecules overlapped across all analyses, red colour denotes positive association and blue colour negative association of each molecule at gene (G) and protein (P) level with disease score and tissue modulus. b) Network of known protein:protein interactions from IntAct and BioGRID within the 22 ECM-associated. Visualisation was carried out using Cytoscape v.3.3.0. c) Based on gene expression levels of these molecules we calculated a matrix index as the ratio of average level of expression of genes positively associated to those negatively associated with disease score and tissue modulus. Scatter plots show the correlation of matrix index with tissue modulus (linear regression, $N = 30$, $R^2 = 0.74$, $p < 0.0001$) and disease score (linear regression, $N = 35$, $R^2 = 0.76$, $p < 0.0001$). d) Association of matrix index with immune gene signature expression. Barplot illustrates Spearman p-values, FDR corrected using the Benjamini & Hochberg method. Red denotes positive correlations, blue denotes negative and gray denotes insignificant associations. The dotted line specifies the significance cutoff $p = 0.05$. e) Kaplan-Meier survival curves with overall survival of TCGA and ICGC dataset for HGSOC divided by high or low matrix index. The x-axis is in the unit of years. f) Comparison of hazard ratio scores (HR, with 95% CI) derived from Cox proportional hazards model for matrix index and the indicated gene expression signatures extracted from literature on the ovarian TCGA dataset. Left panel corresponds to univariate analysis, right panel corresponds to multivariate analysis taking into account age, tumor stage, grade and treatment (i.e., primary therapy outcome success). The asterisks represent the significance in the KM analysis between the high- and low-index groups (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$ and ▪$0.05 < p < 0.1$).

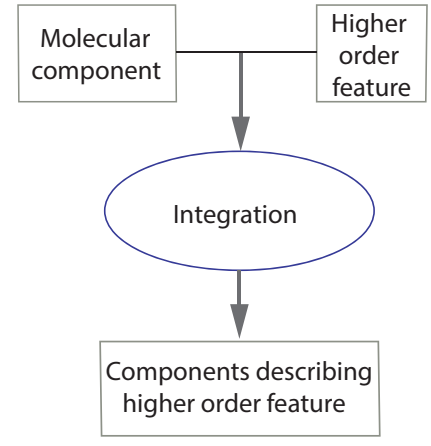
Figure 6. Matrix index reveals a common stromal reaction across cancers

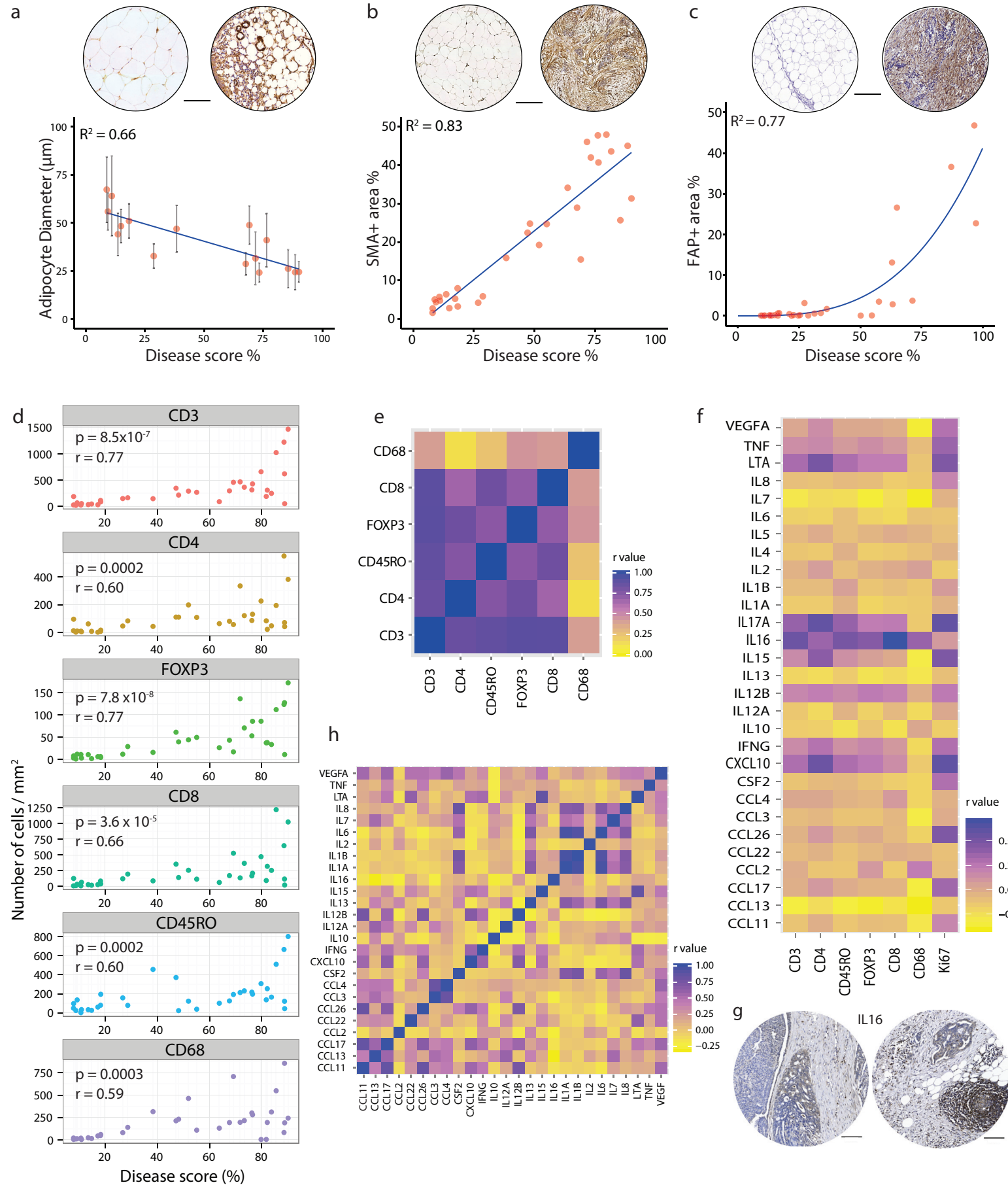
a) Kaplan-Meier survival curves with overall survival from the indicated datasets divided by high or low matrix index. The x-axis is in the unit of years. b) Multivariate hazard ratio (HR, with 95% CI) derived from a Cox proportional hazards regression model across cancer types / datasets using the matrix index. In each cancer, patients were split into high and low index groups, and their association with the overall survival (OS) was tested taking into account age, stage, grade (T-factor), and treatment factors. Asterisks represent the significance in the KM analysis between the high- and low-index groups (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$ and ▪$0.05 < p < 0.1$). HR > 1 means that high index is inversely correlated with OS, while HR < 1 means high index positively correlated OS. c) Example IHC images from TNBC, PDAC and DLBCL biopsies digitally quantified using Definiens™ software on cancer tissue array cores for matrix index proteins FN1, COL11A1, CTSB, and COMP. High intensity staining = red, medium = orange, low = yellow. d) Quantification of IHC staining on tissue arrays from TNBC, PDAC and DLBCL biopsies using Definiens™ software. Box plots illustrate the percentage area of high intensity staining for each marker. Scale bar = 500µm. COL11A1 and FN1, N = 30, 36, 54; CTSB, N = 28, 35, 52; COMP, N = 29, 35, 54; for TNBC, PDAC and DLBCL respectively.


a
Higher order features
Architecture
36 Samples
Biomechanics
32 Samples
Cellularity
36 Samples
Omentum
site of metastasis
Primary tumour
Resection
36 Samples
Transcriptomics
35 Samples
Chemo/Cytokines
32 Samples
Molecular components
Proteomics
33 Samples


b
Adipose
Stroma
Tumour
Lymphoid structure
% tissue area
100
75
50
25
0
G93 G45 G35 G76 G42 G102 G69 G81 G72 G57 G43 G94 G92 G8 G36 G10 G54 G144 G85 G51 G58 G17 G30 G88 G67 G11 G75 G15 G79 G33 G77 G124 G29 G136 G138 G82
100µm
100µm


200µm
Disease score


c
Molecular component
Higher order feature
Integration
Components describing higher order feature


a
b
c
Adipocyte Diameter (µm)
R² = 0.66
Disease score %
SMA+ area %
R² = 0.83
Disease score %
FAP+ area %
R² = 0.77
Disease score %
d
CD3
p = 8.5x10⁻⁷
r = 0.77
CD4
p = 0.0002
r = 0.60
FOXP3
p = 7.8 x10⁻⁸
r = 0.77
CD8
p = 3.6 x 10⁻⁵
r = 0.66
CD45RO
p = 0.0002
r = 0.60
CD68
p = 0.0003
r = 0.59
Number of cells / mm²
Disease score (%)
e
CD68
CD8
FOXP3
CD45RO
CD4
CD3
r value
f
VEGFA
TNF
LTA
IL8
IL7
IL6
IL5
IL4
IL2
IL1B
IL1A
IL17A
IL16
IL15
IL13
IL12B
IL12A
IL10
IFNG
CXCL10
CSF2
CCL4
CCL3
CCL26
CCL22
CCL2
CCL17
CCL13
CCL11
CD3
CD4
CD45RO
FOXP3
CD8
CD68
Ki67
r value
g
IL16
h
r value

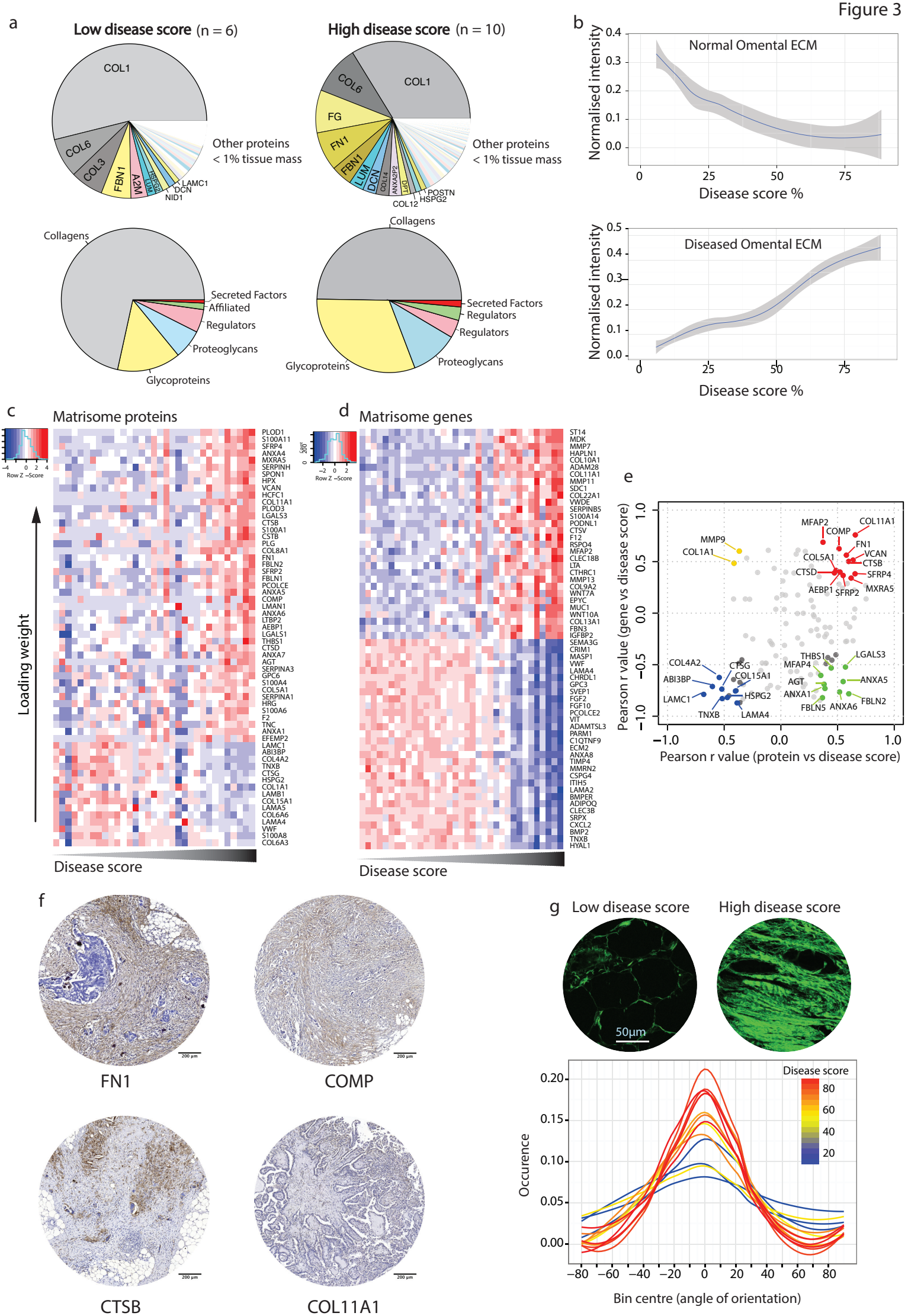
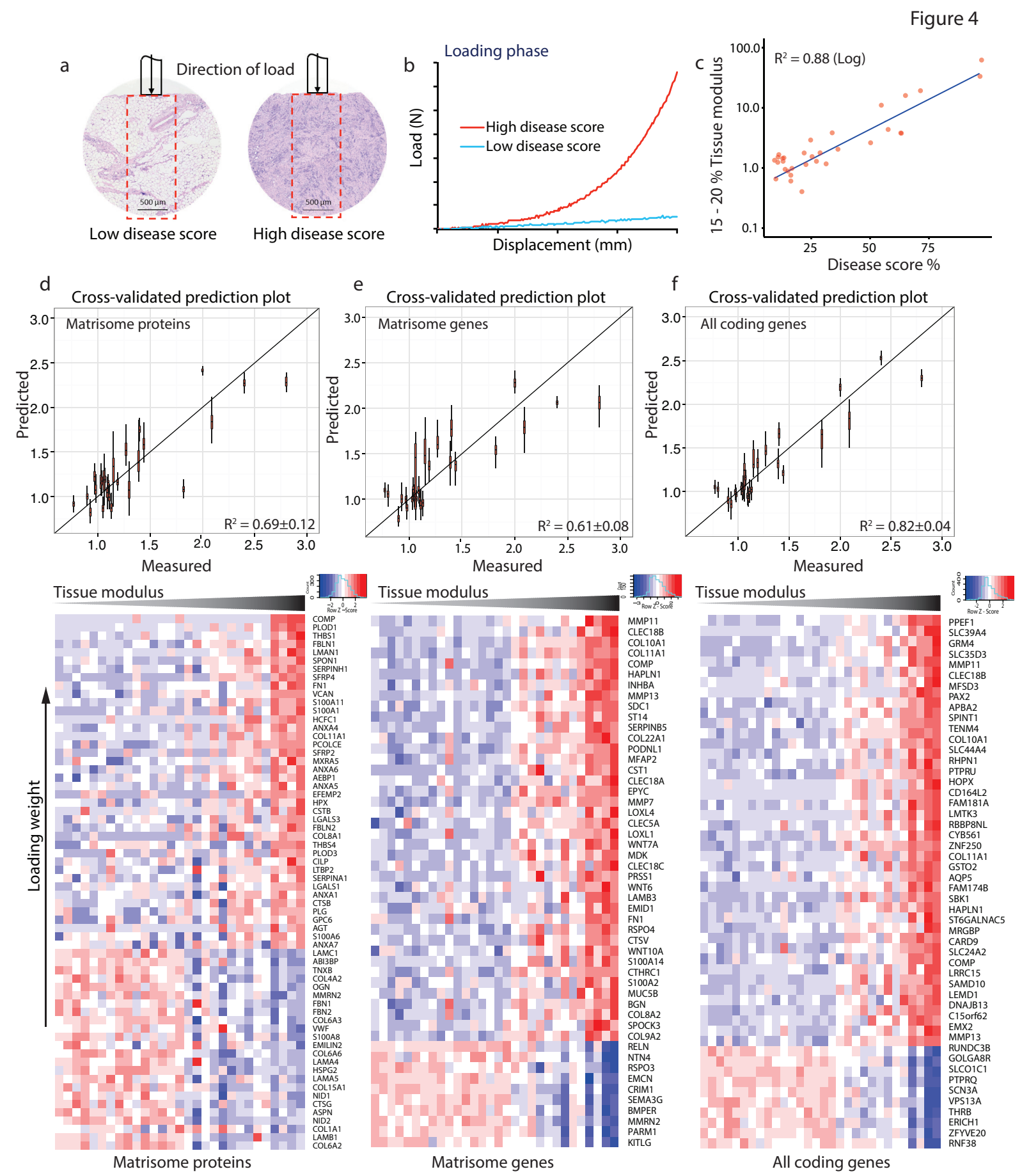
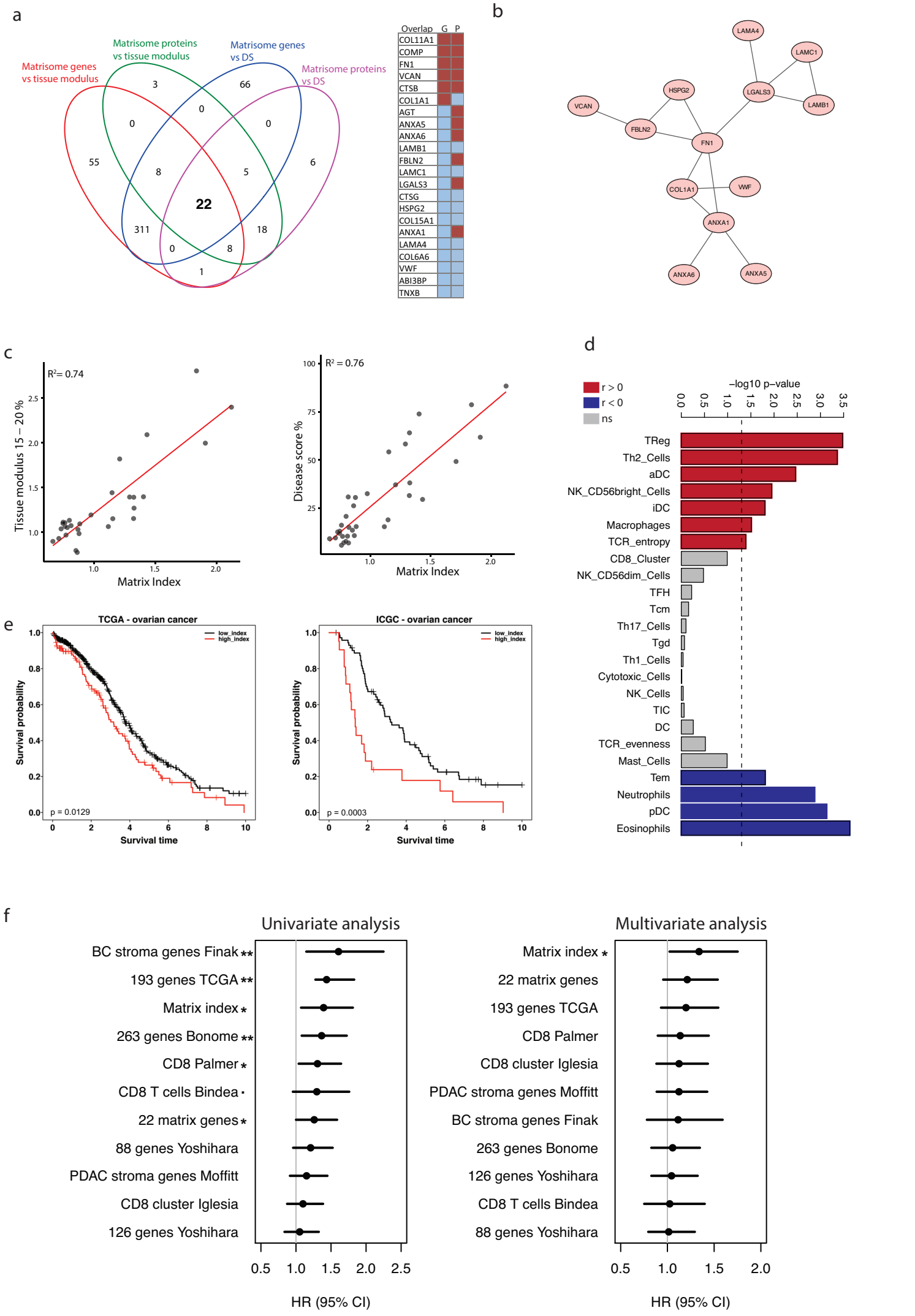
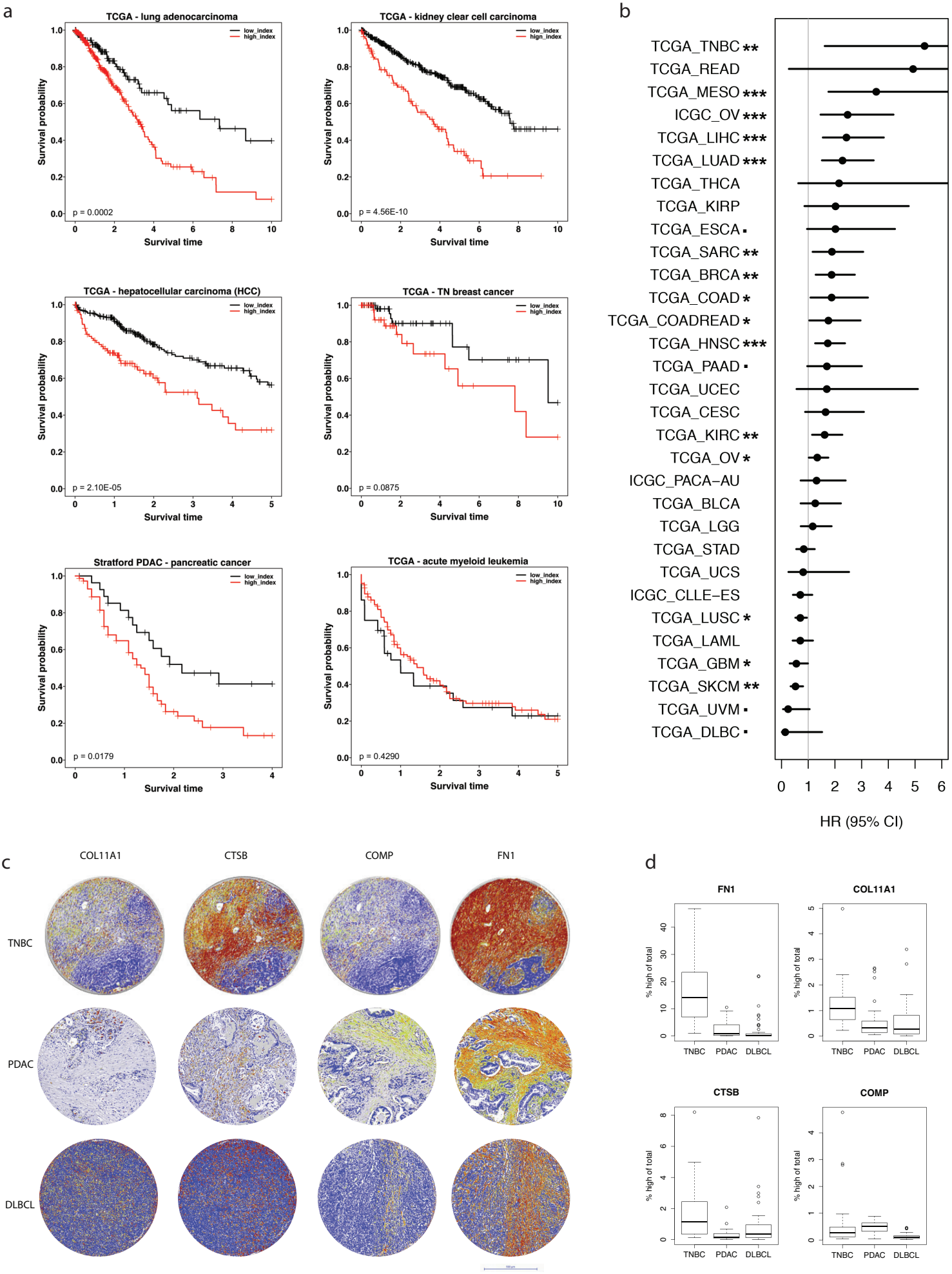

a
Low disease score (n = 6)
High disease score (n = 10)
COL1
COL6
COL3
FBN1
A2M
HSPG2
LUM
LAMC1
DCN
NID1
Other proteins
< 1% tissue mass
COL1
COL6
FG
FN1
FBN1
LUM
DCN
COL14
ANXA2P2
DPT
COL12
POSTN
HSPG2
Other proteins
< 1% tissue mass
Collagens
Secreted Factors
Affiliated
Regulators
Proteoglycans
Glycoproteins
Collagens
Secreted Factors
Regulators
Regulators
Proteoglycans
Glycoproteins
b
Normal Omental ECM
Normalised intensity
Disease score %
Diseased Omental ECM
Normalised intensity
Disease score %
c
Matrisome proteins
Row Z –Score
Loading weight
Disease score
PLOD1
S100A11
SFRP4
ANXA4
MXRA5
SERPINH
SPON1
HPX
VCAN
HCFC1
COL11A1
PLOD3
LGALS3
CTSB
S100A1
CSTB
PLG
COL8A1
FN1
FBLN2
SFRP2
FBLN1
PCOLCE
ANXA5
COMP
LMAN1
ANXA6
LTBP2
AEBP1
LGALS1
THBS1
CTSD
ANXA7
AGT
SERPINA3
GPC6
S100A4
COL5A1
SERPINA1
HRG
S100A6
F2
TNC
ANXA1
EFEMP2
LAMC1
ABI3BP
COL4A2
TNXB
CTSG
HSPG2
COL1A1
LAMB1
COL15A1
LAMA5
COL6A6
LAMA4
VWF
S100A8
COL6A3
d
Matrisome genes
Row Z –Score
Disease score
ST14
MDK
MMP7
HAPLN1
COL10A1
ADAM28
COL11A1
MMP11
SDC1
COL22A1
VWDE
SERPINB5
S100A14
PODNL1
CTSV
F12
RSPO4
MFAP2
CLEC18B
LTA
CTHRC1
MMP13
COL9A2
WNT7A
EPYC
MUC1
WNT10A
COL13A1
FBN3
IGFBP2
SEMA3G
CRIM1
MASP1
VWF
LAMA4
CHRDL1
GPC3
SVEP1
FGF2
FGF10
PCOLCE2
VIT
ADAMTSL3
PARM1
C1QTNF9
ECM2
ANXA8
TIMP4
MMRN2
CSPG4
ITIH5
LAMA2
BMPER
ADIPOQ
CLEC3B
SRPX
CXCL2
BMP2
TNXB
HYAL1
e
Pearson r value (gene vs disease score)
Pearson r value (protein vs disease score)
MFAP2
COL11A1
COMP
MMP9
COL1A1
FN1
VCAN
COL5A1
CTSB
CTSD
SFRP4
AEBP1
SFRP2
MXRA5
THBS1
LGALS3
COL4A2
CTSG
MFAP4
COL15A1
ABI3BP
ANXA5
AGT
LAMC1
HSPG2
ANXA1
FBLN5
ANXA6
FBLN2
TNXB
LAMA4
f
FN1
COMP
CTSB
COL11A1
g
Low disease score
High disease score
50µm
Disease score
80
60
40
20
Occurence
Bin centre (angle of orientation)

Figure 4


a
Direction of load
500 µm
500 µm
Low disease score
High disease score
b
Loading phase
Load (N)
High disease score
Low disease score
Displacement (mm)
c
R² = 0.88 (Log)
15 - 20 % Tissue modulus
100.0
10.0
1.0
0.1
25
50
75
Disease score %
d
Cross-validated prediction plot
Matrisome proteins
Predicted
Measured
R² = 0.69±0.12
e
Cross-validated prediction plot
Matrisome genes
Predicted
Measured
R² = 0.61±0.08
f
Cross-validated prediction plot
All coding genes
Predicted
Measured
R² = 0.82±0.04
Tissue modulus
Loading weight
COMP
PLOD1
THBS1
FBLN1
LMAN1
SPON1
SERPINH1
SFRP4
FN1
VCAN
S100A11
S100A1
HCFC1
ANXA4
COL11A1
PCOLCE
SFRP2
MXRA5
ANXA6
AEBP1
ANXA5
EFEMP2
HPX
CSTB
LGALS3
FBLN2
COL8A1
THBS4
PLOD3
CILP
LTBP2
SERPINA1
LGALS1
ANXA1
CTSB
PLG
GPC6
AGT
S100A6
ANXA7
LAMC1
ABI3BP
TNXB
COL4A2
OGN
MMRN2
FBN1
FBN2
COL6A3
VWF
S100A8
EMILIN2
COL6A6
LAMA4
HSPG2
LAMA5
COL15A1
NID1
CTSG
ASPN
NID2
COL1A1
LAMB1
COL6A2
Matrisome proteins
MMP11
CLEC18B
COL10A1
COL11A1
COMP
HAPLN1
INHBA
MMP13
SDC1
ST14
SERPINB5
COL22A1
PODNL1
MFAP2
CST1
CLEC18A
EPYC
MMP7
LOXL4
CLEC5A
LOXL1
WNT7A
MDK
CLEC18C
PRSS1
WNT6
LAMB3
EMID1
FN1
RSPO4
CTSV
WNT10A
S100A14
CTHRC1
S100A2
MUC5B
BGN
COL8A2
SPOCK3
COL9A2
RELN
NTN4
RSPO3
EMCN
CRIM1
SEMA3G
BMPER
MMRN2
PARM1
KITLG
Matrisome genes
PPEF1
SLC39A4
GRM4
SLC35D3
MMP11
CLEC18B
MFSD3
PAX2
APBA2
SPINT1
TENM4
COL10A1
SLC44A4
RHPN1
PTPRU
HOPX
CD164L2
FAM181A
LMTK3
RBBP8NL
CYB561
ZNF250
COL11A1
GSTO2
AQP5
FAM174B
SBK1
HAPLN1
ST6GALNAC5
MRGBP
CARD9
SLC24A2
COMP
LRRC15
SAMD10
LEMD1
DNAJB13
C15orf62
EMX2
MMP13
RUNDC3B
GOLGA8R
SLCO1C1
PTPRQ
SCN3A
VPS13A
THRB
ERICH1
ZFYVE20
RNF38
All coding genes


a
Matrisome proteins vs tissue modulus
Matrisome genes vs DS
Matrisome genes vs tissue modulus
Matrisome proteins vs DS
3
66
0
0
0
55
8
5
6
22
311
18
0
8
1
Overlap G P
COL11A1
COMP
FN1
VCAN
CTSB
COL1A1
AGT
ANXA5
ANXA6
LAMB1
FBLN2
LAMC1
LGALS3
CTSG
HSPG2
COL15A1
ANXA1
LAMA4
COL6A6
VWF
ABI3BP
TNXB
b
LAMA4
LAMC1
HSPG2
LGALS3
LAMB1
VCAN
FBLN2
FN1
COL1A1
VWF
ANXA1
ANXA6
ANXA5
c
R²= 0.74
Tissue modulus 15 – 20 %
Matrix Index
R² = 0.76
Disease score %
Matrix Index
d
r > 0
r < 0
ns
–log10 p–value
TReg
Th2_Cells
aDC
NK_CD56bright_Cells
iDC
Macrophages
TCR_entropy
CD8_Cluster
NK_CD56dim_Cells
TFH
Tcm
Th17_Cells
Tgd
Th1_Cells
Cytotoxic_Cells
NK_Cells
TIC
DC
TCR_evenness
Mast_Cells
Tem
Neutrophils
pDC
Eosinophils
e
TCGA - ovarian cancer
low_index
high_index
Survival probability
p = 0.0129
Survival time
ICGC - ovarian cancer
low_index
high_index
Survival probability
p = 0.0003
Survival time
f
Univariate analysis
BC stroma genes Finak **
193 genes TCGA **
Matrix index *
263 genes Bonome **
CD8 Palmer *
CD8 T cells Bindea ·
22 matrix genes *
88 genes Yoshihara
PDAC stroma genes Moffitt
CD8 cluster Iglesia
126 genes Yoshihara
HR (95% CI)
Multivariate analysis
Matrix index *
22 matrix genes
193 genes TCGA
CD8 Palmer
CD8 cluster Iglesia
PDAC stroma genes Moffitt
BC stroma genes Finak
263 genes Bonome
126 genes Yoshihara
CD8 T cells Bindea
88 genes Yoshihara
HR (95% CI)


a
TCGA - lung adenocarcinoma
p = 0.0002
TCGA - kidney clear cell carcinoma
p = 4.56E-10
TCGA - hepatocellular carcinoma (HCC)
p = 2.10E-05
TCGA - TN breast cancer
p = 0.0875
Stratford PDAC - pancreatic cancer
p = 0.0179
TCGA - acute myeloid leukemia
p = 0.4290
low_index
high_index
Survival probability
Survival time
b
TCGA_TNBC **
TCGA_READ
TCGA_MESO ***
ICGC_OV ***
TCGA_LIHC ***
TCGA_LUAD ***
TCGA_THCA
TCGA_KIRP
TCGA_ESCA .
TCGA_SARC **
TCGA_BRCA **
TCGA_COAD *
TCGA_COADREAD *
TCGA_HNSC ***
TCGA_PAAD .
TCGA_UCEC
TCGA_CESC
TCGA_KIRC **
TCGA_OV *
ICGC_PACA–AU
TCGA_BLCA
TCGA_LGG
TCGA_STAD
TCGA_UCS
ICGC_CLLE–ES
TCGA_LUSC *
TCGA_LAML
TCGA_GBM *
TCGA_SKCM **
TCGA_UVM .
TCGA_DLBC .
HR (95% CI)
c
COL11A1
CTSB
COMP
FN1
TNBC
PDAC
DLBCL
d
FN1
COL11A1
CTSB
COMP
% high of total
TNBC
PDAC
DLBCL