Pan-cancer analysis connects tumor matrisome to immune response

Su Bin Lim, Melvin Lee Kiang Chua, Joe Poh Sheng Yeong, Swee Jin Tan, Wan-Teck Lim and Chwee Teck Lim

Recent sequencing efforts unveil genomic landscapes of tumor microenvironment. A key compartment in this niche is the extracellular matrix (ECM) and its related components – matrisome. Yet, little is known about the extent to which matrisome pattern is conserved in progressive tumors across diverse cancer types. Using integrative genomic approaches, we conducted multi-platform assessment of a measure of deregulated matrisome associated with tumor progression, termed as tumor matrisome index (TMI), in over 30,000 patient-derived samples. Combined quantitative analyses of genomics and proteomics reveal that TMI is closely associated with mutational load, tumor pathology, and predicts survival across different malignancies. Interestingly, we observed an enrichment of specific tumor-infiltrating immune cell populations, along with signatures predictive of resistance to immune checkpoint blockade immunotherapy, and clinically targetable immune checkpoints in TMIhigh tumors. B7-H3 emerged as a particularly promising target for anti-tumor immunity in these tumors. Here, we show that matrisomal abnormalities could represent a potential clinically useful biomarker for prognostication and prediction of immunotherapy response.

npj Precision Oncology (2019) 3:15; https://doi.org/10.1038/s41698-019-0087-0

INTRODUCTION
The extracellular matrix (ECM) is a complex multi-spatial meshwork of macromolecules with structural, biochemical and biomechanical cues, influencing virtually all fundamental aspects of cell biology. Although tightly controlled during normal development, ECM is frequently altered in many diseases, including cancer. Despite clear evidence of abnormal tumor matrix in cancer, characterization and understanding their functional role in tumors have been challenging, possibly owing to complex nature of ECM proteins and their associated factors, or matrisome. Little is known about the extent to which matrisome pattern is shared across various carcinomas or unique in tumors of differing metastatic potential; it remains unknown as to whether there exist subclasses of tumor matrisome that modulates tumor initiation and response to therapy, particularly in the context of immune response.

Tumor matrisome index (TMI) is a 29-matrisome-gene expression based classifier that has been validated for its predictive value in prognosis and adjuvant therapy response in early-stage non-small cell lung cancer (NSCLC). TMI comprises exclusively of a small set of matrisome gene signatures, primarily encoding non-core matrisome proteins, which were found to be most differentially expressed in NSCLC relative to matched normal epithelium. In essence, TMI is calculated by the sum of the expression level that is multiplied by predefined Cox proportional hazards model coefficient of each TMI gene (see Methods).

Here we hypothesized that this specific pattern of deregulated matrisome could be a common determinant of tumor aggression irrespective of tumor origin. Given the prognostic significance of TMI in predicting the response to adjuvant chemotherapy (ACT) in NSCLC, we investigated if TMI would be associated with signatures predictive of clinical response to immune checkpoint inhibitor (ICI) treatments, including total mutational burden (TMB), PD-L1 expression, tumor-infiltrating lymphocytes (TILs), and immune gene signatures. Through parallel analyses of whole-transcriptomic and matched proteomic data, we report robust associations of TMI with TMB, histopathological and clinical features, and reveal the immune landscape of matrisome-deregulated tumors. Our resource of curated compendiums of 8,386 genome-wide profiles, and molecular and clinical associations of TMI may advance our understanding on the underlying role of biophysical matrix changes that is common across cancer types.

RESULTS
Matrisome is commonly deregulated in human cancers
The present study was conducted to extend the investigation of TMI performance into 11 major cancer types (Fig. 1a). To facilitate multi-platform parallel analyses with RNA-seq-acquired TCGA data, we first generated a unified, cancer type-specific, merged microarray-acquired dataset (MMD), comprising 8386 transcriptome profiles of tumor and tumor-free tissues, using our

1NUS Graduate School for Integrative Sciences & Engineering, National University of Singapore, Singapore, Singapore; 2Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore; 3Division of Radiation Oncology, National Cancer Centre Singapore, Singapore, Singapore; 4Division of Medical Sciences, National University of Singapore, Singapore, Singapore; 5Oncology Academic Clinical Programme, Duke-NUS Medical School, Singapore, Singapore; 6Department of Anatomical Pathology, Singapore General Hospital, Singapore, Singapore; 7Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, Singapore, Singapore; 8Sysmex Asia Pacific Pte Ltd, Singapore, Singapore; 9Division of Medical Oncology, National Cancer Centre Singapore, Singapore, Singapore; 10Office of Clinical Sciences, Duke-NUS Medical School, Singapore, Singapore; 11Mechanobiology Institute, National University of Singapore, Singapore, Singapore and 12Institute for Health Innovation and Technology, National University of Singapore, Singapore, Singapore

Correspondence: Wan-Teck Lim (darren.lim.w@singhealth.com.sg) or Chwee Teck Lim (ctlim@nus.edu.sg)

Received: 19 November 2018 Accepted: 10 April 2019
Published online: 22 May 2019

Published in partnership with The Hormel Institute, University of Minnesota
previously developed R pipeline\textsuperscript{7} (Supplementary Data files S1 and S2; see Methods). As TMI biomarkers were derived from lung cancer, we first queried the extent to which genome-wide expression patterns of lung cancer would be shared across various cancers.

Using MMD and TCGA, we found that the differential expression patterns of lung cancers, relative to non-tumor controls, were comparable to that of breast, ovarian, bladder, colorectal, and prostate cancers, regardless of assayed platforms (Tables S1 and S2), as recently observed in systematic pan-cancer analyses.\textsuperscript{10,11} Black and gray lines represent the ranked position of generic and lung-specific TMI signature, respectively. TMI in tumor vs. non-tumor tissues across 11 cancer types. The black horizontal line indicates the mean of the samples. ***Mann–Whitney U-test \( P < 0.001 \), **\( P < 0.01 \), *\( P < 0.05 \).

Using MMD and TCGA, we found that the differential expression patterns of lung cancers, relative to non-tumor controls, were comparable to that of breast, ovarian, bladder, colorectal, and prostate cancers, regardless of assayed platforms (Tables S1 and S2), as recently observed in systematic pan-cancer analyses.\textsuperscript{10,11} Several of the 29 TMI genes were consistently placed in the top ranked differentially expressed gene (DEG) list across multiple cancer types at the individual gene level (Fig. 1b and Table S3). We identified a subset of genes that was significantly enriched in lung adenocarcinoma (LUAD), prostate adenocarcinoma (PRAD), kidney renal papillary cell carcinoma (KIRP), stomach adenocarcinoma (STAD), colon adenocarcinoma (COAD), breast invasive carcinoma (BRCA), liver hepatocellular carcinoma (LIHC) and bladder urothelial carcinoma (BLCA), and we termed it as a generic TMI signature (Supplementary Data file S3; see Methods). This initial screen supported our hypothesis that altered matrisome dynamics might represent a common phenotype across different malignancies.

TMI distinguishes cancers from normal tissues

Except for TCGA PAAD (pancreatic adenocarcinoma), which had an insufficient number of normal samples (\( n = 4 \)), all cancer types demonstrated significant difference in the TMI between tumor and non-tumor tissues (Fig. 1c and Supplementary Data files S1, S4 and S5). Expectedly, NSCLC-derived TMI achieved near-perfect diagnostic accuracy in lung cancer datasets, where the area under the receiver operating characteristic (ROC) curve (AUC) at the optimal cut-offs was 0.946, 0.999, and 0.999 in lung MMD, TCGA LUAD, and LUSC (lung squamous cell carcinoma), respectively (Fig. 1d). Across all malignancies, we observed an overall classification accuracies of 82% and 92% using MMD and TCGA datasets, respectively (Table S4). Of note, the observed difference between the two assayed platforms (MMD vs. TCGA) is possibly due to different number of TMI genes filtered in TCGA for final index computation (Supplementary Data file S2; see Methods).

To perform a pan-cancer analysis with minimal technical variability across samples, we next analyzed an independent data source in which tumors were procured under standard conditions (Fig. S1 and Supplementary Data file S6; see Methods). Although over a narrower range than for lung cancer, 1492 carcinomas spanning 10 cancer types displayed a wide and diverse TMI distribution, suggesting a varying degree of matrisome abnormalities at different stages of cancer progression and a specific TMI spectrum for each cancer type.

TMI is predictive of patient survival

We obtained a total of 72 independent validation cohorts annotated with clinical information, including overall survival (OS), disease-specific survival (DSS), disease-free survival (DFS), relapse-free survival (RFS), metastasis-free survival (MFS), and/or progression-free survival (PFS) data. As previously described,\textsuperscript{8} a cut-off index was determined for each dataset to stratify patients into TMI\textsubscript{low} and TMI\textsubscript{high} groups. Univariable survival analyses revealed a cancer-specific association of TMI in predicting time to death, recurrence, and distant metastasis (Fig. 2a, b; Supplementary Data file S7; see our earlier work for lung cancer\textsuperscript{9}). TMI\textsubscript{high} was an unfavorable prognostic factor for OS in colon, liver, renal, and breast cancers, whereas it appeared to confer a favorable prognosis in ovarian and gastric cancers, even after adjustment for clinical parameters on multivariable analyses using TCGA...
datasets (Fig. 2c and Supplementary Data file S8). Comparing with traditional clinical features between TMI\textsubscript{low} and TMI\textsubscript{high} patients from TCGA datasets, we found that the TMI\textsubscript{high} group had a higher proportion of patients staged pathologically as T3 or T4, diagnosed as lymph node and distant metastasis positive, and classified as late stage (Fig. 2d).

TMI is associated with pathological and molecular features

We next asked if there would be a change in TMI during multistep carcinogenesis. In breast, colorectal, and pancreatic cancers, we observed increasing TMI values corresponding to the progressive steps of oncogenesis (normal to adenoma to carcinoma in situ to invasive carcinoma; Fig. 3a and Supplementary Data file S9). Next, we further investigated for an association between TMI and known prognostic molecular phenotypes in breast cancers (Luminal A/B, HER2\textsuperscript{+} and basal subtypes).\textsuperscript{12,13} Our prognostic TMI was highest among the most adverse molecular phenotypes (basal and HER2\textsuperscript{+} tumors), which were known to harbor the worst prognosis (Fig. 3b and Supplementary Data file S10).

We also explored if TMI expression would be correlated with genomic mutational burden, and analyzed matched whole-exome sequencing-derived data of nine TCGA cohorts (pancreas, stomach, prostate, colon, lung, breast, kidney, and bladder), for which TMI had previously been computed (Table S5). TMI was positively correlated with somatic TMB across cancer types (Spearman’s rho test $r_s = 0.23$ and $P = 1.77\times10^{-14}$ (Breast) to $r_s = 0.48$ and $P = 1.41\times10^{-11}$ (Pancreas); Fig. 3c, top; Supplementary Data file S11). Additionally, based on a previously defined cut-off for discretization into TMI\textsubscript{high} and TMI\textsubscript{low}, the former harbored significantly higher TMB in all cancers (Fig. 3c, bottom), except for renal (Mann–Whitney U-test $P = 0.104$) and liver cancers (Mann–Whitney U-test $P = 0.065$).

TMI in the context of immune response

Given the evidence in support of TMB as a robust determinant of tumor immunogenicity in many solid tumors,\textsuperscript{14,15} and that TMI is closely associated with TMB, we next asked if there would be any direct association of this specified matrisomal pattern with the cancer immune landscape. Applying machine learning-based CIBERSORT to MMD and TCGA cancer patient samples (see Methods), we observed a close correlation between TMI and the composition of specified TIL populations for several cancers (Fig. 4a). Enrichment of these TILs related to both innate and adaptive immunity was diverse and cancer type-specific. Relative abundance of M0 and M1 macrophages, neutrophils, activated mast cells, regulatory T cells (Treg), and T follicular helper (Tfh) cells, activated CD4\textsuperscript{+} memory T cells increased, while resting CD4\textsuperscript{+} memory T cells, mast cells, naive B cells, and resting dendritic cells decreased with tumor-promoting matrisomal change (Supplementary Data file S12). Impact on immune infiltrates was particularly pronounced in gastric, lung, colorectal and breast cancers, having significant positive and negative correlations with TMI, thus highlighting the influence on the immune contexture during matrisome remodeling in this subset of cancers.

Early works suggest that signatures of T cell states, particularly that of CD8\textsuperscript{+} T cells, may predict clinical response to ICI-based immunotherapy.\textsuperscript{16–19} Of all cancer types analyzed, breast (BRCA) and pancreatic (PAAD) cancers demonstrated pronounced negative correlation with the estimated abundance of CD8\textsuperscript{+} T cells, indicating that TMI\textsubscript{low} tumors harbored higher CD8\textsuperscript{+} T-cell
infiltration levels in these cancers. To validate their differential expression at the protein level, we assessed matched proteomes of TCGA samples provided by the NCI Clinical Proteomic Tumor Analysis Consortium for breast (BRCA) cancers, for which samples were previously classified as either TMI low or TMI high at the transcriptomic level (CPTAC; see Methods).

A matched-cohort assessment of 108 BRCA samples revealed higher protein levels of CIBERSORT-defined CD8+ T cell signatures, including CD8A, GZMB, LIME1, and RASA3, in TMI low tumors (Fig. 4b). Interestingly, differential expression analysis of whole proteomes comparing the two groups (TMI high vs. TMI low) identified MAGEA3, a frequently expressed tumor-specific antigen, as the second most highly expressed protein in TMI low tumors (Fig. 4c). The functional contribution of spontaneously occurring MAGEA3-reactive CD8+ T cells to favorable prognosis may explain the better patient outcomes consistently observed in TMI low breast tumors.

Abundance of CD8+ T cells has been associated with better response to immunotherapies. Given the enriched CD8+ T cell signatures at both transcript and protein levels in TMI low tumors, we next investigated the association between TMI and recent reported predictive signatures for immunotherapy response. The responder signature comprises 161 genes, which were highly expressed in anti-PD-1 responding melanoma patients compared to non-responding patients (Fig. 5a and Supplementary Data file S13). Unexpectedly, the data revealed strong correlations of TMI with B7-H3 expression in all (MMD) and nearly all (TCGA) tumor samples regardless of cancer types. Having observed mRNA-protein expression correlation using the processed proteomic data in TCGA BRCA cohort (Fig. 5b; Spearman’s rho test $r_s = 0.38$ and $P = 4.19E−05$), we found markedly higher levels of B7-H3 protein expression in TMI high breast tumors (Fig. 5c). We thus speculate that the present combined quantitative analyses may help to unveil specific immune targets that could be selectively efficacious in a subset of tumors across diverse carcinomas.

B7-H3 as a potential immune target for TMI high tumors

We next correlated the index with the expression of 20 potentially targetable immune checkpoints—that are currently in preclinical or clinical trial stages, and/or FDA-approved (Fig. 5a).

**Fig. 3** Tumor pathology and molecular features associated with TMI. **a** Tumor pathology associations with TMI in breast, colon and pancreas cancers. **b** TMI stratified by molecular subtypes in breast cancer. **c** Correlation of TMI with total mutational burden (TMB) in TCGA cohorts. Patient samples in each dataset are stratified into TMI low or TMI high group based on the optimal predefined TMI cut-offs. Linear regression lines are drawn (black line) with 95% CI (gray zone); $n =$ number of samples analyzed; $r_s =$ Spearman’s correlation coefficient; Mann–Whitney U-test $P$-values are stated. For **a** and **b**, Kruskal–Wallis $P$-values are stated. For **a**–**c**, box hinges represent 1st and 3rd quartiles, and middle represents the median. The upper and lower whiskers extend from hinges up and down indicate the most extreme values that are within 1.5*IQR* (interquartile range) of the respective hinge. The short horizontal lines represent the standard deviations.
Promising predictive value for immunotherapy and other potential targets

We questioned if TMI could further be associated with innate PD-1 resistance signature (IPRES), including genes involved in immunosuppression, angiogenesis, monocyte and macrophage chemotaxis, and EMT. To comparatively with other commercially available multi-gene tests (MGTs). Given that TMI signatures were derived from lung cancer, we chose lung cancer-derived prognostic MGTs and their predictive values for immunotherapy were assessed using lung cancer datasets (lung MMD and TCGA LUAD), with MGT-specific predefined cut-off thresholds for risk stratification (see Methods). The Myriad myplan Lung Cancer and Pervenio Lung RS tests are the only currently available prognostic MGTs for lung cancers, which comprise 31 cell cycle progression (CCP) genes and 11 cancer-related pathway genes, respectively. For each validation dataset, IPRES-enriched patients were identified as previously described (see Methods), and demonstrated significant differences in TMI, but not in other prognostic indices, with the exception of Pervenio in the MMD cohort (Fig. 5d).

To identify potential therapeutic targets for both TMIlow and TMIhigh groups, we performed differential proteomic analyses comparing the two groups and investigated if there would be any overlapping targets identified by other commercially available MGTs. Of note, as matched proteomic data were available only for breast cancer (TCGA BRCA), we analyzed two breast cancer MGTs (Oncotype DX and MammaPrint) in addition to the two lung cancer MGTs (Fig. 5e; Supplementary Data file S14; see Methods).

Interestingly, while both breast and lung cancer MGTs demonstrated several differentially expressed proteins between the two groups, we observed new potential targets that were exclusive to either lung or breast cancers ( Fig. 5f; Supplementary Data file S15).

Further, a total of 4 (CASP14, CENPI, PAX9, and TDRD12) and 12 (FAM174A, UBXN10, BBS5, C9orf40, CACNA2D2, TFF1, UBXN10, RIPK4, CALHM6, WDFC2, NAGPA) proteins were consistently enriched in MGThigh and MGTlow tumors in TCGA BRCA cohort, respectively (Fig. 5g).

**DISCUSSION**

This work is based on multi-platform evaluation of TMI at multiple molecular levels across 11 major cancer types. By curating a total of 8836 patient-derived tumor and tumor-free samples, we
generated 11 cancer type-specific MMDs annotated with clinical features and deposited the data at ArrayExpress (see Data availability). This approach minimizes biased selection of validation cohorts and allows parallel analyses with other high-throughput data sources such as TCGA.

Considering that the index was computed under the assumptions of each gene having concordant regulation as observed in lung cancer, it is noteworthy that tumors with highly dynamic and constantly remodeling ECM could have ubiquitously altered expression of matrisome genes across genetically and phenotypically diverse epithelial tumors. The data may imply the presence of shared signal transduction pathways activated across these selected tumors, such as recently reported TGFβ or HIF1α/VEGF pathways, in controlling ECM balance or cell-intrinsic mechanism regulating the expression of set of ECM genes contributing to tumor development.23

The functions of TMI matrisome genes, including secreted factors and metalloproteinases, in forming the pre-metastatic niche at specific target site remain poorly defined despite their well-established role in local matrix degradation and regulation.24,25 Exceptions are MMP1, which induces vascular permeability and mediates breast cancer metastasis to the lung,24,26 and S100 proteins, which generate pro-inflammatory phenotypes and recruit non-resident, bone marrow-derived cells.24,27,28 Intriguingly, our survival analyses revealed that TMI was correlated with a tropism for metastasis development to the lungs, but not to the bones in breast cancer patients, albeit in a subset of studies (Fig. 2b). This may suggest that the TMI signatures could inform on the tumor promoting microenvironment that facilitates metastasis to respective tissue organs. Through a matched proteomic data analysis, we further identified a small number of specific genes consistently high expressed at protein level in both risk groups classified by TMI and other commercially available MGTs. Although their expression levels and prognostic values in malignant tumors remain inconclusive, high expression of CASP14 and CENPI have been associated with poor prognosis in breast cancer, consistent with our observations in TMIhigh patient group.

In addition to close associations with patient survival, we provide a comprehensive overview of commonly deregulated matrisome pattern in the context of tumor genotypes, molecular phenotypes, and immune response. Particularly, our combined quantitative analyses of matched proteomes revealed significant
enrichment of MAGEA3 and CD8+ T cell signatures in TMI$_{low}$ tumors. MAGEA3 is a cancer-germline gene highly expressed in various carcinomas, but the gene is silent in normal adult tissues, except for the testes and placenta.\textsuperscript{31} Independent of tumor burden, the prognostic value of MAGEA3-reactive CD8+ T cells for overall survival was reported in esophageal squamous cell carcinoma.\textsuperscript{20} Further, the study observed elevated expansion of functional MAGEA3-specific CD8+ T cells both in vitro and in vivo upon anti-PD-1 treatment.\textsuperscript{32} It is thus tempting to speculate that the genes included in the TMI signature might have functional roles for immune surveillance in the tumor microenvironment, given that MAGEA3 was the second most differentially expressed gene and anti-PD-1 therapy responder signature was highly enriched in TMI$_{low}$ tumors.

To assess if the present approach could provide new indications for immunotherapy, we next associated TMI directly with the previously reported predictive biomarkers for ICI-based treatments, such as TMB, PD-L1 expression, TIL density, peripheral blood markers and other immune gene signatures.\textsuperscript{9} Of 20 targetable immune checkpoints, B7-H3 was the only gene with higher protein expression in TMI$_{high}$ tumors in TCGA BRCA cohort. Consistent with our findings, increasing evidence suggests B7-H3 as a negative predictor of patient outcomes in solid tumors, including breast cancer.\textsuperscript{32–35} Further, its functional contribution to T cell inhibition and immune evasion is increasingly being recognized, making the molecule an appealing target as a novel potential drug target.\textsuperscript{36}

METHODS

Data preprocessing and MMD generation

Published in partnership with The Hormel Institute, University of Minnesota npj Precision Oncology (2019) 15

TIL abundance alone might not serve as an ideal predictive biomarker.\textsuperscript{14,15} Taken together, we propose the design of combinatorial approaches whereby tumor agnostic biomarkers are integrated with tumor-specific molecular indices, so as to optimize the accuracy of predicting immunotherapy response in specific cancer types.
are recorded in Supplementary Data files S4 and S5, respectively. TMI of tumor and normal groups were compared using the Mann–Whitney–Wilcoxon test. To further statistically evaluate the diagnostic accuracy of the TMI in deciding the presence of the disease, we computed the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, and specificity with the best threshold (Table S4) using the pROC package.58 All ROC curves generated were subjected to binormal smoothing for illustration (Fig. 1c).

**TMI spectrum quantification**

We analyzed the Expression Project for Oncology (expO) data provided by the International Genomics Consortium (IGC, USA, www.intgen.com) for TMI spectrum quantification, for tumor samples procured and processed under standard conditions, resulting in minimal non-biological variations across multiple cancer types. All tumor specimens annotated with prostate, lung, renal, liver, gastric, breast, ovarian, pancreatic, colorectal, and bladder cancer from the expO dataset were included in the spectrum quantification (Fig. S1 and Supplementary Data file S6).

**Survival analyses**

A statistical summary of datasets used in survival analyses is recorded in Supplementary Data file S7. Kaplan–Meier (KM) survival curves were derived for OS and DSS and other multiple endpoints using the survival package in R (http://CRAN.R-project.org/package=survival). Clinical data for TCGA datasets were collected via the embedded DownloadBinspecimenClinicalData function in the TCGA-Assembler package.47 Multivariate Cox regression analyses were performed to adjust confounding factors including age, race, gender, pt, pn, and pm status (Supplementary Data file S8). For all cancer types, patients with available survival data and TMI were all included in the KM analyses.

**Molecular subtyping in breast cancer**

Of analyzed datasets, five datasets (BRCA, GSE20711, GSE21653, GSE19615, and GSE50567) were annotated with either ER, PR, and HER2 status or predefined molecular subtypes of breast cancers (normal breast-like, luminal A, luminal B, HER2 positive, and basal-like). Tumors harboring positive status of ER and/or negative status of HER2 were classified as luminal cancers while the basal-like tumors were defined as ER-, PR-, and HER2-negative cancers (Supplementary Data file S10).

**Total mutational burden**

The mutational load of TCGA tumor samples across nine epithelial cancer types, for which TMI was previously computed, was obtained from the National Cancer Institute GDC Data Portal (http://portal.gdc.cancer.gov/projects/). The portal defines the mutational load as the total number of simple somatic mutations. Only tumors harboring at least one mutation were included and log 10 transformed to compute Spearman correlations between TMI and mutational load. A descriptive statistics including the number of patient samples available for both TMI gene expression and mutational load data, cut-off points used for patient stratification, and key outputs of Spearman correlation tests in each dataset is shown in Supplementary Data file S11.

**IPRES and GSVA**

Gene signatures previously associated with resistance to PD-1 immunotherapy, termed Innate PD-1 RESistance (IPRES), were obtained from Broad MSigDB (http://software.broadinstitute.org/gsea/msigdb) and Supplementary Data provided by the original publication.21 The gene set variation analysis (GSVA) scores of the signatures were computed for each MMD and TCGA patient sample. GSVA scores were transformed to z-scores, and were correlated with TMI. As previously described,21 a cut-off of > 0.35 was applied to the mean z-score for a patient to be determined as IPRES-enriched. As these signatures were derived from non-responders (resistant) tumors, we further defined responder signatures as 161 highly expressed genes (log FC > 2 and Mann–Whitney P < 0.1) in responders compared to non-responders using the list of 693 DE genes provided by the original work.21

**CIBERSORT**

The estimated fraction of individual immune cell types was computed using the beta version of CIBERSORT (http://cibersort.stanford.edu/). For any given sample, we calculated Spearman correlation between TMI and relative abundance of each immune cell type using our 11 generated MMDs and 11 TCGA cohorts (LUAD, OV, SKCM, BLCA, LIHC, BRCA, COAD, STAD, KIRC, PRAD, and PAAD). As our MMDs of breast, colorectal, and lung cancer exceeded maximum load capacity (500 MB), 1000 tumors were randomly selected for input data files. We selected LM22 (22 immune cell types) for signature gene file, 100 for permutations, and disabled quantile normalization for all runs.

**Quantitative proteomic analysis**

Relative protein abundance data of 10,625 protein-coding genes were generated by the National Cancer Institute Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC, https://cptc.xferui.sgeorgetown.edu/publicData/Phase_II_Data/TCGA_Breast_Cancer/). Log ratios (base 2), representing relative abundance of each sample compared to the pooled reference sample, were obtained from TCGA_Breast_BI_Proteome.itraq.tsv for a total of 108 TCGA BRCA breast cancer samples, for which TMI was previously computed.

**Differential expression analysis and heatmaps**

R/Bioconductor limma package49 was used to assess differential protein expression between two patient groups classified by five MGTs. FC > 1.5 or FC < −1.5 and limma P < 0.05 were applied to determine DE proteins. A descriptive list of all DEGs at the protein level found in both groups is provided in Supplementary Data file S15. Heatmaps used in this study were generated using Morpheus (http://software.broadinstitute.org/morpheus/).

**DATA AVAILABILITY**

Our generated cancer type-specific MMDs are available at ArrayExpress under accession codes E-MTAB-6690 (pancreatic cancer), E-MTAB-6691 (ovarian cancer), E-MTAB-6692 (renal cancer), E-MTAB-6693 (gastric cancer), E-MTAB-6694 (prostate cancer), E-MTAB-6695 (liver cancer), E-MTAB-6696 (bladder cancer), E-MTAB-6697 (melanoma cancer), E-MTAB-6698 (colorectal cancer), E-MTAB-6699 (lung cancer), and E-MTAB-6703 (breast cancer). The accession codes of all public datasets analyzed in this study are listed in Supplementary Data file S1.

**CODE AVAILABILITY**

The R codes used to preprocess, merge and batch-effect correct datasets for generation of all 11 MMDs can be found in Figshare (https://doi.org/10.6084/m9.figshare.7870896).

**ACKNOWLEDGEMENTS**

This work was conceived and carried out at the Mechanobiology Engineering laboratory at the Department of Biomedical Engineering, National University of Singapore. We acknowledge support provided by the National Research Foundation, Prime Minister’s Office, Singapore under its Research Centre for Excellence, and Mechanobiology Institute at NUS. W.-T.L. is supported by the National Medical Research Council (NMRC/CSA/040/2012 and NMRC/CSA-INV/0025/2017). S.B.L. is supported by NUS Graduate School for Integrative Sciences and Engineering (NG5), Daewoong Foundation, and MOGAM Science Scholarship Foundation. We extend our appreciation to P.Z. and K.I. for proofreading the article.

**AUTHOR CONTRIBUTIONS**

S.B.L., S.J.T., W.-T.L. and C.T.L. conceived and designed the study. S.B.L. performed bioinformatics analyses. S.B.L., M.C., J.Y., W.-T.L. and C.T.L. analyzed and interpreted the data. J.Y. contributed from immunopathology perspective. S.B.L., M.C., J.Y., S.J.T., W.-T.L. and C.T.L. reviewed and contributed to the manuscript.

**ADDITIONAL INFORMATION**

Supplementary Information accompanies the paper on the npj Precision Oncology website (https://doi.org/10.1038/s41698-019-0087-0).

**Competing interests:** S.B.L. and C.T.L. have filed a patent for the 29-gene assay. The remaining authors declare no competing interests.
REFERENCES

1. Hynes, R. O. The extracellular matrix: not just pretty fibrils. Science 326, 1216–1219 (2009).
2. Bateman, J. F., Boot-Handford, R. P. & Lamande, S. R. Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. Nat. Rev. Genet. 10, 173–183 (2009).
3. Lu, P., Weaver, V. M. & Werb, Z. The extracellular matrix: a dynamic niche in cancer progression. J. Cell Biol. 196, 395–406 (2012).
4. Naba, A. et al. The extracellular matrix: tools and insights for the “omics” era. Matrix Biol. 49, 10–24 (2016).
5. Hynes, R. O. & Naba, A. Overview of the matrixome—an inventory of extracellular matrix constituents and functions. Cold Spring Harb. Perspect. Biol. 4, a004903 (2012).
6. Naba, A. et al. The matrixome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol. Cell Proteom. 11 (M111), 014647 (2012).
7. Lim, S. B., Tan, S. J., Lim, W.-T. & Lim, C. T. A merged lung cancer transcriptome dataset for clinical predictive modeling. Sci. Data 5, 180136 (2018).
8. Lim, S. B., Tan, S. J., Lim, W.-T. & Lim, C. T. An extracellular matrix-related prognostic and predictive indicator for early-stage non-small cell lung cancer. Nat. Commun. 8, 1734 (2017).
9. Gibney, G. T., Weiner, L. M. & Atkins, M. B. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. Lancet Oncol. 17, e542–e551 (2016).
10. Uhlen, M. et al. A pathology atlas of the human cancer transcriptome. Science 357, eaan2507, https://doi.org/10.1126/science.aan2507 (2017).
11. Hoadley, K. A. et al. Multiplatform analysis of 12 cancer types reveals molecular alterations and immune target enrichment following epithelial-to-mesenchymal transition enabling rapid biomarker cutoff optimization. PLoS ONE 7, e51038 (2014).
12. Mak, M. P. et al. Pre-metastatic niches: organ-specific homes for metastases. Nat. Rev. Cancer 17, 302–317 (2017).
13. Lu, X. et al. ADAMTS1 and MMP1 proteolytically engage EGF-like ligands in an osteolytic signaling cascade for bone metastasis. Genes Dev. 23, 1882–1894 (2009).
14. Mine, A. J. et al. Genes that mediate breast cancer metastasis to lung. Nature 436, 518–524 (2005).
15. Gupta, G. P. et al. Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. Nature 446, 765–770 (2007).
16. Peinado, H. et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. Nat. Med. 18, 883–891 (2012).
17. Hanae, T. et al. Caspase14 expression is associated with triple negative phenotypes and cancer stem cell marker expression in breast cancer patients. J. Surg. Oncol. 116, 706–715 (2017).
18. Chen, X. et al. Dual TGF-beta and PD-1 blockade synergistically enhances MAGE-A3-specific immune responses in advanced melanoma. Cancer Immunol. Immunother. 65 (2016).
19. Huang, A. C. et al. T-cell invigoration to tumour burden ratio associated with anti-CTLA-4 and anti-PD-1 checkpoint blockade. J. Immunother. 41, 357–365 (2018).
20. Arigami, T. et al. B7-h3 ligand expression by primary breast cancer and associated with regional nodal metastasis. Ann. Surg. 252, 1044–1051 (2010).
21. Castellanos, J. R. et al. B7-H3 role in the immune landscape of cancer. Am. J. Clin. Exp. Immunol. 6, 66–75 (2017).
22. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. Nat. Rev. Cancer 12, 252–264 (2012).
23. Chen, Y. W., Teke, C. & Foldstad, O. The immunoregulatory protein human B7H3 is a tumor-associated antigen that regulates tumor cell migration and invasion. Curr. Cancer Drug Targets 8, 404–413 (2008).
24. Teke, C. et al. B7-H3 contributes to the metastatic capacity of melanoma cells by modulation of known metastasis-associated genes. Int. J. Cancer 130, 2282–2290 (2012).
25. Ricci-Vitiani, L. et al. Tumor vascularization via endothelial differentiation of glioblastoma stem-like cells. Nature 468, 824–828 (2010).
26. Taube, J. M. et al. Association of PD-1, PD-1 ligands, and other features of the tumor immune microenvironment with response to anti-PD-1 therapy. Clin. Cancer Res. 20, 5064–5074 (2014).
27. Clugger, H. M. et al. Characterization of PD-L1 expression and associated T-cells infiltrates in metastatic melanoma samples from variable anatomic sites. Clin. Cancer Res. 21, 3052–3060 (2015).
28. Lim, S. B., Tan, S. J., Lim, W.-T. & Lim, C. T. Cross-platform meta-analysis reveals common matrisome variation associated with tumor genotypes and immunophenotypes in human cancers. Preprint at https://doi.org/10.1101/353698 (2018).
29. Lim, S. B., Tan, S. J., Lim, W.-T. & Lim, C. T. Compendium of cancer transcription for machine learning applications. Preprint at https://doi.org/10.1101/353706 (2018).
30. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy–affymetrix analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20, 307–315 (2004).
31. Taminau, J. et al. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. BMC Bioinf. 13, 35 (2012).
32. Zhu, Y., Qiu, P. & Ji, Y. TCGA-assembler: open-source software for retrieving and processing TCGA data. Nat. Methods 11, 599–600 (2014).
33. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
34. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015).
35. Gu, Z., Gu, L., Ellis, R., Schlesner, M. & Brors, B. Cilsrate implements and enhances circular visualization in R. Bioinformatics 20, 2811–2812 (2014).
36. Tan, T. Z. et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. EMBO Mol. Med. 6, 1279–1293 (2014).
37. Budzies, J. et al. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. PLoS ONE 7, e51862 (2012).
38. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. 12, 77 (2011).