



Pan-Cancer analysis of the expression and regulation of matrisome genes across 32 tumor types



Valerio Izzi^a, Juho Lakkala^a, Raman Devarajan^a, Anni Kääriäinen^a, Jarkko Koivunen^a, Ritva Heljasvaara^{a,b} and Taina Pihlajaniemi^a

a - Oulu Center for Cell-Matrix Research and Biocenter Oulu, Faculty of Biochemistry and Molecular Medicine, University of Oulu, P. O. Box 5000, FIN-90014 Oulu, Finland

b - Centre for Cancer Biomarkers (CCBIO), Department of Biomedicine, University of Bergen, N-5009 Bergen, Norway

Correspondence to Valerio Izzi: valerio.izzi@oulu.fi
<https://doi.org/10.1016/j.mbplus.2019.04.001>

Abstract

The microenvironment plays a central role in cancer, and neoplastic cells actively shape it to their needs by complex arrays of extracellular matrix (ECM) proteins, enzymes, cytokines and growth factors collectively referred to as the matrisome. Studies on the cancer matrisome have been performed for single or few neoplasms, but a more systematic analysis is still missing. Here we present a Pan-Cancer study of matrisome gene expression in 10,487 patients across 32 tumor types, supplemented with transcription factors (TFs) and driver genes/pathways regulating each tumor's matrisome. We report on 919 TF-target pairs, either used specifically or shared across tumor types, and their prognostic significance, 40 master regulators, 31 overarching regulatory pathways and the potential for druggability with FDA-approved cancer drugs. These results provide a comprehensive transcriptional architecture of the cancer matrisome and suggest the need for development of specific matrisome-targeting approaches for future therapies.

© 2019 University Of Oulu. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

The crucial role played by the microenvironment in cancer development, progression and dissemination has been extensively investigated [1–5] and evidence begin to accumulate for Pan-Cancer “micro-environmental signatures” impacting patients' prognosis and survival [6–8]. Cancer microenvironment is a complex entity though, populated by different cell types embedded in a mesh of structural and functional extracellular matrix (ECM) and ECM-associated molecules, ECM-remodeling enzymes, cytokines and growth factors collectively referred to as the matrisome [9–11]. Altered expression of matrisome proteins is intimately linked to cancer progression, as *e.g.* in the case of collagen I, II, III, V and IX deposition in a vast array of solid tumors [12,13] and impaired laminin adhesion in breast and lung cancer [14,15]. Likewise, ECM remodeling *via* metalloproteinases and impaired or ectopic interac-

tion with ECM receptors such as CD44 are reportedly major Pan-Cancer drivers of dissemination and resistance to therapy [13,16,17].

On the other hand, understanding and manipulating the peritumoral matrisome composition shows promising therapeutic potential. In example, over-expressed matrisome proteins can be used as “beacons” for antibody-drug conjugate therapy to direct the chemotherapeutic cargo just next to the cancer cells, thus maximizing therapeutic effects while minimizing toxicity [18], and ECM-cancer interaction mechanisms can be manipulated to increase the efficacy of antimetastatic drugs [19,20]. The burgeoning of cancer “omics” data during the last years, led by The Cancer Genome Atlas (TCGA) Research Network and culminated in the Pan-Cancer Atlas (the most comprehensive collection of integrated data on gene expression, methylation, mutation and copy-number variation in human cancers available to date), has provided an

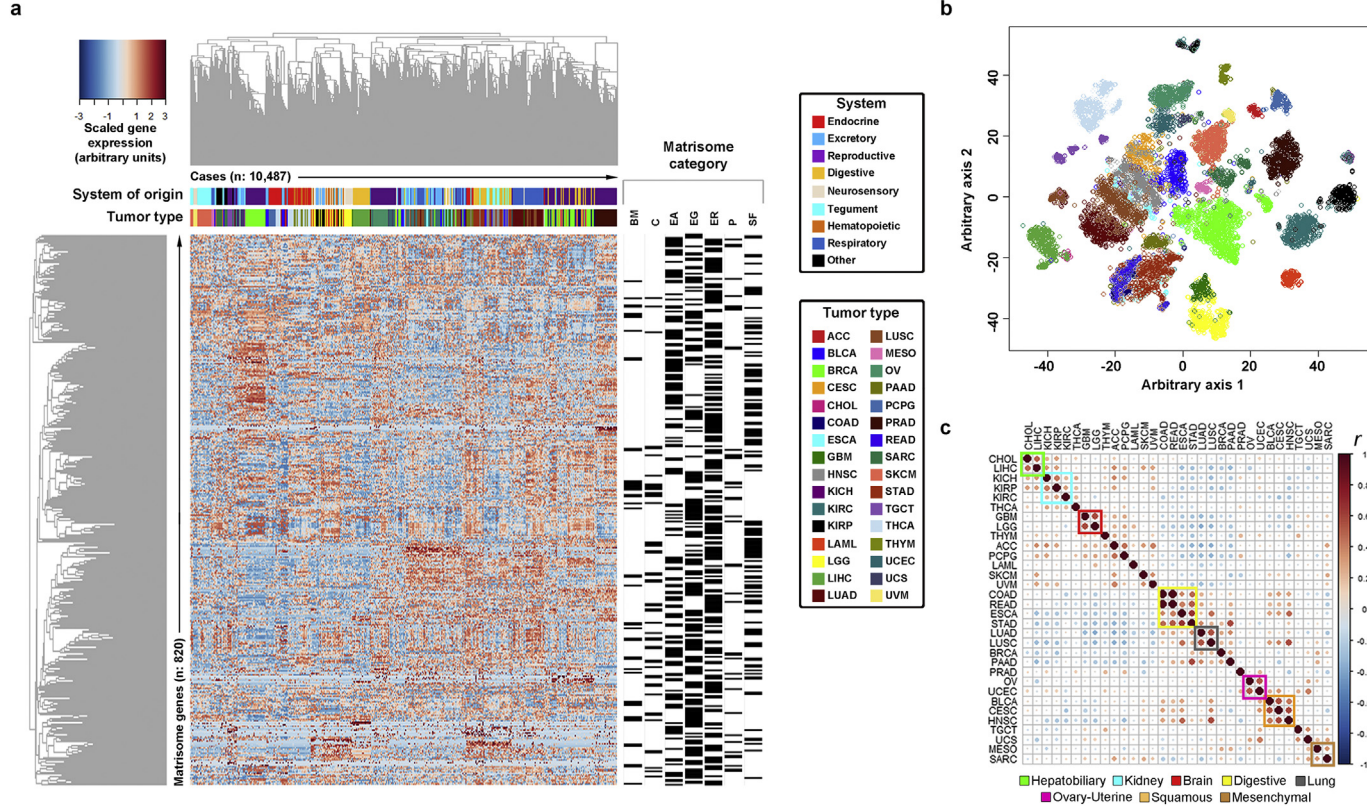


Fig. 1. The cancer matrisome. (a) Heatmap of the expression of 820 matrisome genes in 10,487 cases from 32 tumor types. The matrisome category of each gene is given on the right as black bars, while tumor types and system of the body from which each tumor originates are color-coded and placed on top of the heatmap. The distance metrics for hierarchical clustering is 1-Pearson correlation (r). *Abbreviations:* ACC, adrenocortical cancer; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical and endocervical cancer; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head & neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney clear cell carcinoma; KIRP, kidney papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, brain lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma & paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumor; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrioid carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma. BM, basement membranes; C, collagens; EA, ECM-affiliated; EG, ECM glycoproteins; ER, ECM regulators; P, proteoglycans; SF, secreted factors. (b) Multidimensional scaling of the data in (a) by T-distributed Stochastic Neighbor Embedding (t-SNE). (c) Inter-tumor correlation (Pearson correlation r) based on the 100 most-expressed matrisome genes in each tumor type. Clusters of highly correlated tumor types are reported as colored squares.

incredible momentum to any field of cancer research [21–23]. Several “layers” of information from TCGA have obviously already entered the matrisome field, contributing to the identification of subtle patient phenotypes and molecular mechanisms [24–27], but a comprehensive view of the whole “cancer matrisome”, systematically defining gene expression and its transcriptional regulatory circuitry, is still missing.

Considering the centrality of matrisome in cancer, we have attempted to fill this gap by compiling a systematic atlas of matrisome data from the TCGA Pan-Cancer cohort to discover similarities and differences between different tumors and to infer regulatory relationships driving matrisome organization in different tumors. We show here that the composition of each “cancer matrisome” is a combination of tumor-specific molecular phenotypes, tissue- and system-of-origin features and physiopathological mechanisms. We further show that different tumors share similar matrisome organization under similar master regulatory mechanisms, suggestive of common microenvironmental needs, and that matrisome expression has significant impact on Pan-Cancer and tumor-specific patient survival and offers possibilities for therapeutic interventions.

Results and discussion

We analyzed the expression of 820 matrisome genes [10] across 32 cohorts of The Cancer Genome Atlas (TCGA) Pan-Cancer initiative [28]. After filtering out samples with missing clinical or genetic information, 10,487 patients from different cohorts were included (Suppl Table 1).

At global level, matrisome gene expression separated neatly each tumor (Fig. 1a, b) with an average recall accuracy per tumor of 95% by 3 independent automatic classification algorithms (χ -square test <0.0001). Tissue-of-origin effects were typical confounding factors as already reported [29], determining misclassification of e.g. rectal (READ) as colon (COAD) cancer in 25% of the cases, COAD as READ in 8% of the cases, lung squamous (LUSC) as adenomatous (LUAD) cancer in 7% of the cases, and uterine carcinosarcoma (UCS) as uterine endometrial carcinoma (UCEC) in 8% of the cases and as sarcoma (SARC) in 9% of the cases (Suppl Table 2, 3 and Suppl Fig. 1). Restricting the analysis to the most-expressed genes in each tumor type exacerbated tissue-of-origin effects and highlighted clusters of highly correlated tumors, in which tissue- and organ-of-origin effects as well as histopathological features such as squamosity converged (Fig. 1c and Suppl Figs. 2, 3). Notably, the overall similarity between tumors was much lower than between healthy samples from matching tissues (Suppl Fig. 4, 5), suggesting a strong matrisome specialization

during cancer development [30]. We identified 29 tumor-specific matrisome signatures (see Suppl Table 4), which varied significantly in size and composition (χ -square test <0.001). With respect to global matrisome expression, the signatures had the benefit of being smaller and easier to further investigate and to concentrate on those matrisome genes which scored among the highest-ranking genes in each tumor type. Also, multi-tumor clustering could be performed with signatures, producing coherent high-ranking results in multiple cohorts and tests used for cross-validation (see Suppl Table 5 and Suppl Figs. 6–8).

Using a multi-source consensus approach, we then inferred a complete chart of TFs driving gene expression (each TF-target interaction is defined here as a “module”) for 28 tumor types. Samples in this and further analyses included almost only primary tumors (99.27% of total) while metastatic tumors accounted for 0.3% of total and recurrent tumors for 0.43% of total (see Suppl Fig. 9a). Before further modelling, we also evaluated global and tumor-specific patterns of matrisome expression in primary and metastatic or recurrent cancers, in light of suggested matrisome differences in the primary and metastatic niches [31,32]. Data in Suppl Fig. 9b, c show that, for what concerns matrisome expression, metastatic and recurrent tumors overlap with primary tumors, both at the global and tumor-specific level (BRCA, THCA, GBM and LGG are considered as examples), in line with similar recent findings by Naba et al. at proteomics level [33]. Next, we moved on to identify transcriptional modules differentially governing the matrisome components (Fig. 2a, b and Suppl Table 6); of these (919 unique modules in total), the majority (783 modules, 85.2%) was highly specific and only active in one tumor type, with the remainder (136 modules, 14.8%) active in two or more tumor types (Fig. 2c) independently of eventual clusters (Fig. 2d). In line with the expected sparsity of gene regulatory networks [34], approx. 89% of all identified TFs had only 1 or 2 matrisome targets and “hub” TFs with 3 or more connections were the minority (Fig. 2e and Suppl Fig. 10). The use of different TF families characterizes the different tumors (χ -square test <0.001) and aligns with the expected abundance of TFs from whole-genome binding studies and functional connections with the ECM [35–41]. *E.g.*, leaving aside the TFs which were not grouped into a specific subfamily (the “other” group), the C2H2 family of zinc-finger factors was the most represented through all tumors (on average 26% of all TFs), followed by leucine zipper/helix-loop-helix factors (bHLH+bZIP) and nuclear receptor TFs (14% and 13% of the total, respectively). Homeodomain-containing TFs were also abundant (10% of all TFs) while HMG/Sox factors and AP-2 were less represented (5% and 3% of the total, respectively), though their expression varied

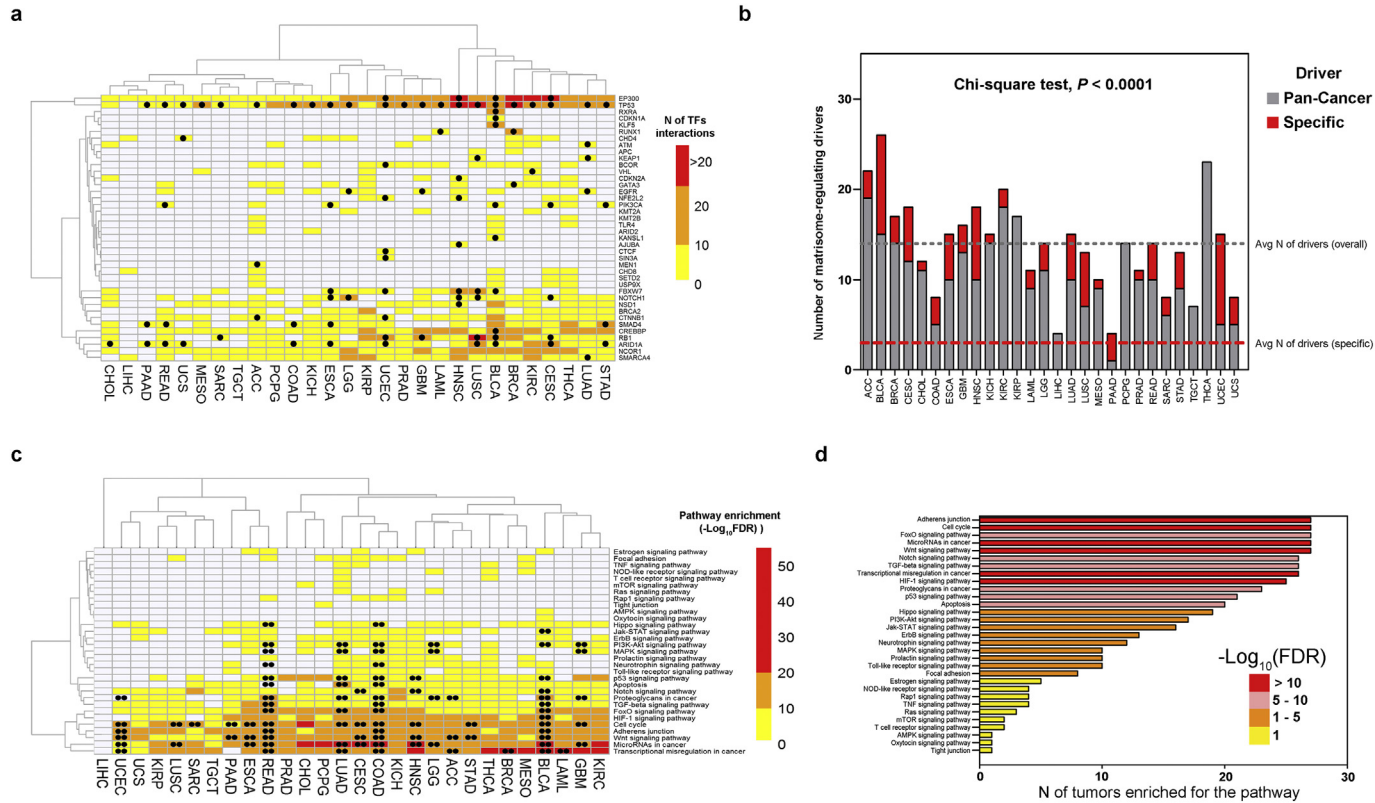


Fig. 3. Master regulators of the cancer matrisome. Identification of driver genes and pathways regulating the matrisome signatures of any given cancer. (a) Heatmap of all driver genes identified as regulators in this study and the number of transcription factors (TF) they regulate. Dots indicate that the regulator is a cancer-specific driver for the given neoplasm. (b) Number of drivers found per cancer type, including general (Pan-Cancer) and tumor-specific (specific) cancer drivers. Average number of total and specific drivers across the study is reported. (c) KEGG enrichment for pathways regulating the matrisome (master regulators + TFs) in any given cancer. Double dots indicate pathways with oncogenic mutations in at least two cancer-specific drivers for the given neoplasm. Values are shown as $-\text{Log}_{10}(\text{FDR})$. (d) Number of tumor types in which each matrisome-regulating pathway is enriched.

significantly between tumors. In fact, AP-2 reached approx. 10% in breast (BRCA), cervical/endocervical (CESC), lung squamous (LUSC) and prostate cancer (PRAD), and nearly 20% of all TFs in pheochromocytoma/paraganglioma (PCPG). HMG/Sox TF were more typically used by clusters of similar tumors, covering approx. 10% of total TFs in brain (GBM and LGG), colorectal (COAD and READ), uterine (UCEC and UCS) and head-and-neck neoplasms (HNSC). The abundant presence of the “other” group depends on the inclusion of several TFs which are crucial for multiple cancers, such as *PAX* and *NFKB* factors in the Rel subfamily, *RUNX1*, 2 and 3 in the Runt subfamily and *TEAD1*, 3 and 4 in the Tea subfamily [35,36,42] (Fig. 2f, g). In line with previous studies [43] and the proposed specialization of matrisome transcriptional programs during cancer development, we also notice that the TF-target correlation within a module for a given cancer is significantly higher (average approx. 8-fold) than the correlation between the same TF-target pair in the rest of the Pan-Cancer cohort (Suppl Fig. 11a). Furthermore, a sizeable proportion of the TFs and targets identified by our analysis stain at “high” or “medium” level (average 61 and 41%, respectively) in tumor samples from The Human Protein Atlas (Suppl Fig. 11b), facilitating future investigations.

We found 233 modules linked to differential patient survival (Suppl Table 7), with examples such as adrenocortical (ACC), bladder (BLCA), breast (BRCA), lung adenomatous cancer (LUAD), sarcomas (SARC) and kidney clear cell carcinomas (KIRC) showing approx. 30% or more modules with prognostic potential and extremes such as low-grade glioma (LGG) and pancreatic cancer (PAAD) showing >50% matrisome modules with significant prognostic value (Suppl Fig. 11c). Notably, in any case in which a TF controlled multiple targets through multiple prognostic modules, the “direction” of the prognostic association (towards either better or worse prognosis) was the same for all the modules, clearly defining noxious or favorable transcriptional circuits for each tumor (Suppl Table 7). Likewise, when the same matrisome target was under different TFs in multiple prognostic modules within the same tumor, the “direction” of the prognostic association was equal, marking noxious and favorable matrisome regulatory mechanisms as above (Suppl Table 7). In line with previous findings, we could not identify global differences in the module profiles of metastatic or recurrent tumors in respect to primary tumors, while rather only a few modules varied their expression to the plus or the minus side, again suggesting that matrisome differences in secondary tumors are limited and very specific (see Suppl Fig. 9d).

Analysis of master regulators identified 40 cancer-specific or Pan-Cancer driver genes [44] which physically interacted with the TFs presiding the

various modules and explained the majority of each cancer matrisome (Fig. 3a, Suppl Fig. 12 and Suppl Table 8). On average, tumors expressed about 14 matrisome-controlling master regulators, 3 of which cancer-specific (Fig. 3b), though in some cases such as bladder and uterine corpus neoplasms (BLCA and UCEC) the number of cancer-specific regulators increased notably, in line with recent reports [44]. The master regulators identified were typically hub proteins (such as TP53 and P300), governing large networks of TFs as expected [45], and the number of drivers in a cancer type correlated linearly with the number of TFs they controlled (Suppl Fig. 13). Cancer drivers occurring in at least two cancer types (“common” regulators) governed, on average, a higher number of TFs than those occurring in a single tumor type (“unique” regulators, see Suppl Fig. 14a–c), possibly providing a further explanation to the observed overlap of matrisome profiles between different tumor types within the same cluster. Also, cancer-specific drivers governed generally more targets than drivers with no specific association (Pan-Cancer - only, Suppl Fig. 14d), suggesting that the cancer matrisome is globally mastered by a relatively small set of regulators which acquire oncogenic mutations in multiple tumors [7,46]. Combinations of regulators are not otherwise interchangeable and, e.g., those characterizing the different clusters such as the digestive system (ESCA, STAD, COAD and READ), the brain (GBM and LGG) or the lung (LUAD and LUSC) show a clear ability to discriminate their cluster of origin (see Suppl Fig. 15). Mapping master regulators and TFs to biological pathways, finally, provided further insights into cancer matrisome regulation. In general, pathway activation by mutations (at least two cancer-specific master regulators carrying oncogenic mutations in the same pathway) was infrequent, though in some tumors (BLCA, COAD, LUAD, READ and UCEC) it accounted for the majority of active pathways. Unsurprisingly [47–55], cell-to-cell adhesion and transcriptional avenues (FoxO, Wnt and MicroRNAs) were the most common matrisome-controlling pathways (27/27, 100% tumor types), followed by other expected pathways such as TP53 and apoptosis, Notch and TGF- β signaling, and hypoxia [56–65] (Fig. 3c, d and Suppl Table 9).

Finally, we used the Drug Gene Interaction Database (DGIdb) [66] to assess the therapeutic implications of our findings. A wide array of drugs, small molecules, and more broadly-defined therapeutic approaches were found to potentially interact with the different levels of the cancer matrisome architecture (Suppl Fig. 16a). Among these, FDA-approved cancer drugs currently labelled for use in cancer show no inferable direct interaction with the matrisome and its circuitry, though many FDA-approved cancer drugs would seem to have such potential if used off-label in different cancers (Suppl

Fig. 16b and Suppl Table 10). We will here provide an example on the use of these data for the prioritization of future drug analyses on the cancer matrisome. As these data are inferred from the intersection of multiple databases and high-throughput screenings, however, we stress that robust, low-throughput validation of the results at the experimental level will be needed before their further use. According to our findings, Axitinib (tyrosine kinase inhibitor) and Dactinomycin (Actinomycin-D) could be re-purposed towards brain and kidney tumors (GBM, LGG and KIRC) to suppress *CSF1* activity. This would be particularly beneficial since *CSF1* expression, due to 3 modules in KIRC and to 1 in GBM and LGG, associates with poorer prognosis in these cancers (Suppl Table 5 and Suppl Fig. 17) in concert with the upregulation of its receptor (*CSF1R*), as already shown [67,68].

The findings we report represent the first attempt to systematize and understand matrisome regulation in cancer, providing an initial platform for further integration with the many aspects of cancer inherently linked to the microenvironment, such as genomic instability [29] and the immune system [69]. Furthermore, our report unveils prognostic matrisome circuitries in almost every tumor type assessed and informs on therapeutic approaches and the possible re-purposing of FDA-approved drugs for targeting the cancer matrisome. With the increasing understanding of the crucial role played by the tumor microenvironment in prognosis and response to therapy [69,70], our analysis might help in the identification of novel markers providing several valid and testable hypotheses for next exploration in cancer biology.

Experimental procedures

Data sources

The Cancer Genome Atlas (TCGA) Pan-Cancer RNA-seq V2 gene expression data and clinical data, data from matching healthy donors (TCGA) or unrelated healthy tissues (Genotype-Tissue Expression (GTEx) database), all from Toil [71] re-computing (RSEM, batch-normalized, $\log_2(x + 1)$ -transformed and upper quantile normalized), were downloaded from the UCSC Xena Browser (<https://xenabrowser.net/>). Textual transcription factor (TF)-target interactions were downloaded from TRRUST [72], ENCODE [73], Marbach [74] and MSigDB [75]. Cap analysis of gene expression (CAGE) data from FANTOM5 [76] were accessed and downloaded via the Semantic catalogue of Samples, Transcription initiation, And Regulations (SSTAR) [77]. FANTOM5 data were restricted to cancer cell samples matching, as much as possible, the tumor types from

TCGA as reported in Suppl Table 11 and only the reads from the most-supported promoter (“p1”) were considered. Tumor-specific and Pan-Cancer drivers were from Bailey et al. [44]. Human protein-protein interaction list was downloaded from BioGRID [78]. Staining profiles for proteins in human tumor tissue based on immunohistochemistry were downloaded from The Human Protein Atlas (THPA) [79]. Druggable genes and drug-genes interactions were from the Drug Gene Interaction Database (DGIdb) [66]. FDA-approved cancer drugs were from Sun et al. [80]. The list of matrisome genes was from the Matrisome Project [10].

Software

Majority of the analyses were performed in R x64 v3.3.2, with specific packages introduced next to relevant methods. Linear support vector machine for feature selection (LSVM) of master regulators at the cluster level was performed in IBM SPSS Modeler 18. Network analyses and gene ontology (GO) enrichment were performed in Cytoscape v3.3.0 with manual overlays. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment was performed in StringDB [81]. Majority of the graphics were produced with GraphPad Prism V5. Cross-validation of cluster signatures was performed in OncoPrint [82]. R codes for inferring TF, cross-validating results in THPA, evaluating module association with survival, identifying master regulators and evaluating druggability (as well as the data associated with the codes) are available upon request to the Authors.

Gene expression and Pan-Cancer classification

The initial data matrix from TCGA Pan-Cancer was intersected with the clinical data and with the list of matrisome genes to include only the genes present in all samples with complete clinical information. This filtering process resulted in a final database of 820 matrisome genes out of 1028 total (820/1028, 79.76%) and 10,487 samples out of 11,257 total, excluding healthy donors and samples with no annotation (10,487/11,257, 93.16%), spanning 32 different tumor types. The types of tumor and the number of samples per each type, as well as the system of origin of each neoplasm, are reported in Suppl Table 1. Heatmap was generated with the gplots package v3.0.1, using 1-Pearson correlation (r) as the distance metrics for hierarchical clustering and information about system of origin, tumor type and matrisome category of each gene as overlays. T-distributed Stochastic Neighbor Embedding (tSNE) using a Barnes-Hut implementation was calculated using the Rtsne package v0.13 and the same distance metrics as for the heatmap. Automated sample classification and cross-validation

was performed using 3 different algorithms: Support Vector Machine (SVM), perceptron-based Neural Network (NN), and the C5.0 classification and regression algorithm (C5.0). SVM analysis was performed using the *e1071* package v1.7-0, setting the analysis to C-classification, data scaling (standardization), radial basis function (RBF) kernel, RBF Gamma 0.1, Gamma 1/dimensions, and 100× cross-validation (80% training, 20% test) *via* random sample permutation. NN analysis was performed using the *neuralnet* package v1.33, data scaling (standardization), 250 as the maximum steps for the training and 10× cross-validation (80% training, 20% test) *via* random sample permutation. C5.0 analysis was performed using the C50 package v0.1.2, with default Quinlan's C5.0 algorithm and 10× cross-validation (80% training, 20% test) *via* random sample permutation. Results from each run of each algorithm were averaged and reported in Suppl Table 2. Total recall accuracy from the combination of the 3 models is reported in Suppl Table 3 and used to build cross-tumor similarities and misclassification tables. Contingency of cross-classified cases was analyzed using the *Stats* package v3.6.0, Chi-square test with Yates continuity correction.

For clustering based on correlation analysis, average expression was calculated per each gene in each tumor type and genes were ranked accordingly. Next, the 20, 50 and 100 most- and least-expressed genes were extracted from each tumor and two new Pan-Cancer lists were created. These lists were subjected to correlation (Pearson correlation (r)) analysis using the *corrplot* package v0.84. The same approaches to t-SNE and correlation-based clustering were followed for the healthy donor samples included in the original TCGA entry (727, after removing samples with missing information) and for GTEx (6456 samples). Matching GTEx-TCGA tissues are reported in Suppl Table 12. For GTEx only, the correlation analysis was repeated twice using both the whole “TCGA TARGET GTEx” cohort from the UCSC TOIL RNA-seq recompute pipeline and isolating GTEx samples, or the individual GTEx cohort from Xena. The two analyses were run in order to verify whether horizontal (cross-cohort) normalization (in the “TCGA TARGET GTEx” cohort) affects the results and the differences between TCGA and GTEx. We observe no differences between the comparisons of cross-normalized cohorts and individually-normalized TCGA and GTEx (not shown), and data are available upon request to the Authors.

To infer gene signatures, the expression of each gene in each tumor type was confronted against the same gene in healthy matching TCGA samples (if present) or in GTEx using the “TCGA TARGET GTEx” cohort. In parallel, each gene in each tumor type was also compared *vs.* the same gene in all other tumor types, excluding tumors belonging to the

same cluster as identified by correlation analysis, if present. Note that uveal melanoma (UVM) was dropped from the cases further analyzed at this stage because of no corresponding healthy tissue identifiable. All analyses were performed using the *Stats* package v3.6.0, two-sided multiple pairwise t -test with Benjamini-Hochberg adjustment for P values. Only the genes up-regulated in tumors *vs.* both healthy and rest of Pan-Cancer at $P < 0.05$ were kept as signature members. Contingency of signature composition in terms of matrisome categories was analyzed using the *Stats* package v3.6.0, Chi-square test with Yates continuity correction, and binary heatmap was generated using the *gplots* package v3.0.1, using 1-Pearson correlation (r) as the distance metrics for hierarchical clustering. Signatures for digestive (ESCA, STAD, COAD, READ), lung (LUAD, LUSC), brain (GBM, LGG), ovary-uterine (OV, UCEC), hepatobiliary (LIHC, CHOL) and kidney (KICH, KIRC, KIRP) clusters were inferred by comparing the single tumors and retaining the genes common to all members. Quasi-proportional Venn diagrams were calculated using *nVennR* package v0.2.1. The hepatobiliary and kidney cluster signatures were not further investigated because of their small size (8 and 2 genes, respectively). The remaining signatures were cross-validated by evaluating the ranking of their components in several tests (tumor *vs.* matching healthy samples) in multiple cohorts from Oncomine (Suppl Table 13). A test was considered positive if the gene in tumor had higher ranking than in healthy (reported as red in the heatmap from Oncomine).

Transcriptional regulation inference

Our approach leverages on a multistep consensus scheme to infer and cross-validate transcription factor (TF) - matrisome binary interactions (modules), upon which a “master regulator” layer is built based on known protein-protein interactions and oncogenicity features of master regulators. To define modules, in the first step a model was built by mining all TF interactions with the matrisome genes in each cancer signature from four databases (TRRUST, ENCODE, Marbach and MSigDB). In parallel, the resulting TFs were searched for occurrence in matching FANTOM5 samples and signature genes were investigated for over-representation of TF binding sites. Note that ovarian cancer (OV) and thymoma (THYM) were dropped from the cases further analyzed at this stage because of no corresponding FANTOM5 samples identifiable. Results were matched and only the modules present in two inference schemes (one of which is obligatory FANTOM5) were retained. TF-target Pearson correlation (r) was then calculated for each retained module in the cancer of interest and in the rest of the Pan-Cancer data, modules with negative correlation

values were removed and correlation coefficients were z-transformed (zCor) using Fisher transformation to ensure their normal distribution for subsequent analyses. A zCor threshold to separate the cancer of interest from the rest of Pan-Cancer was then computed as the Youden Index [83] of the whole zCor values (confidence interval = 0.95), all modules below the threshold were removed and only the remaining modules for which zCor in the cancer of interest is $> zCor$ in Pan-Cancer were retained. For each matrisome gene with >1 TF in the resulting model, then, gene expression values of any TF-target pair were regressed using adaptive lasso regression [84] and only the modules whose coefficients were not shrunk were kept (adaptive lasso pruning). The results of the above procedure (“model 1”) were then passed in parallel to two further inference schemes, namely sparse Bayesian network [85] (“model 2”) and mixed graphical model [86] (“model 3”), each with $10\times$ cross-validation. Finally, results from the models were compared and only modules present in at least two models were retained and reported in the study (Suppl Table 6). The final consensus pipeline yielded networks that are, on average, 27% the size of the original models based on data mining only, thus eliminating $>70\%$ of all mined interactions (data not shown, available on request to the Authors). Cross-validation of results into THPA was possible for 22 types of tumors (Suppl Table 14). For these, TFs and targets from each module were checked against THPA staining profiles and each single protein was considered as cross-validated if Percentile Positive Staining (PPS, % of assayed samples that show “High” or “Medium” staining for a given TF or target) > 0 . All heatmaps were generated using the pheatmap package v1.0.1, using Pearson correlation (r) as the distance metrics for hierarchical clustering.

For each tumor, a “master regulator” layer (cancer driver genes interacting with the TFs identified) was built upon the module structure as follows. Pan-Cancer drivers [44] were filtered so that, for each cancer, their frequency of mutation in the tumor of interest was $>$ average frequency of mutation of the same driver in the whole Pan-Cancer cohort. Cancer-specific drivers [44] were filtered so that their frequency of mutation in the tumor of interest was $>5\%$, well above the lower limit of intermediate-frequency drivers [44] and the Pan-Cancer frequency of mutation in drivers (4.16%). Next, Pan-Cancer and cancer-specific drivers passing the filtering stage were searched for physical interaction with the TFs identified in the given cancer against the BioGrid database, and only drivers interacting with $>10\%$ of all TFs (corresponding to the top 50% of the mined drivers) were considered true “master regulators” and retained (Suppl Table 8). This procedure yielded 380 unique protein-protein physical interactions covering all the 347 master regulator-TF

interactions identified across the different tumors. Of these, 41.05% were Affinity Capture Western, 20.79% Reconstituted Complex, 11.84% Two-hybrid, 6.32% Biochemical Activity, 5.26% Affinity Capture MS, 3.95% Co-localization, 2.11% Protein-peptide, 1.84% Co-crystal Structure, 1.58% Affinity Capture Luminescence, 1.32% Co-fractionation, 0.79% Co-purification, 0.79% Far Western, 0.79% FRET, 0.79% Synthetic Lethality, 0.53% PCA and 0.26% Phenotypic Enhancement. Furthermore, at the end of the pipeline, the weighted average of TFs interacting with each master was 7.96 (ranging from 5.7 in the lower quartile to 36.7 in the upper), in line with the values reported by [45]. Contingency of Pan-Cancer and cancer-specific regulators per each tumor was analyzed using the Stats package v3.6.0, Chi-square test with Yates continuity correction. The overall regulatory infrastructure above matrisome genes (TFs and master regulators) was then enriched for KEGG pathways using StringDB. Total results are restricted to the pathways shown in Fig. 3 for clarity. To infer the most important regulators at the cluster level, the lists of master regulators from each member of the cluster were first intersected, entries not common to all cluster members discarded and linear support vector machine (LSVM) analysis was performed setting epsilon: 0.1, lambda: 0.1, L2 penalty function, predictor importance calculation (for feature extraction) and $10\times$ cross-validation (80% training, 20% test) *via* random sample permutation. Gene expression of extracted predictors was used to classify digestive, lung and brain neoplasms into the appropriate cluster *vs.* rest of the 22 tumors and Area under the curve (AUC) of the Receiver Operating Characteristic (ROC) graph was performed in the ROCR package v1.0-7. All heatmaps were generated using the pheatmap package v1.0.1, using Pearson correlation (r) as the distance metrics for hierarchical clustering.

Prognostic value and druggability of matrisome modules

For each cancer type for which module analysis was possible, we then evaluated the association of module activity with overall patient survival and the possibilities for therapeutic interventions. For survival, univariate analysis (Kaplan-Meier method, Log-Rank test) was performed using the survminer package v0.4.3 and the survival package v2.42-6. To this aim, gene expression of TF and matrisome gene of each module were averaged together and all individuals of any cohort were dichotomized into “high” (above the average) or “low” (below the average) expression of any module identified. For each module in each cohort, finally, survival analysis was performed as above. To evaluate druggability, the whole matrisome architecture (proper matrisome

genes + TFs + master regulators) of each cancer type was searched for drugs and therapeutic approaches against DGldb and all entries without an associated mechanism were removed. From these lists, FDA-approved drugs [80] were deduced and divided into drugs targeting the master regulators, TFs or the matrix proper, and whether their suggested use was on-(FDA-approved for the given cancer) or off-(not FDA-approved for the given cancer) label.

Competing interests

The authors declare no conflict of interests.

Acknowledgements

This work was supported by funds from The Academy of Finland (grants 284605, 294617, and 308867), The Jane and Aatos Erkko Foundation, The Sigrid Jusélius Foundation, The Cancer Society of Finland, The Emil Aaltonen Foundation, and The Finnish Cultural Foundation. Authors would like to thank Prof. M. Sillanpää, University of Oulu, for helping with the transcriptional inference scheme.

Contributions

V.I., R.H. and T.P. designed the research and wrote the manuscript. V.I. performed and supervised computational analyses and developed integrated tools for analysis. J.L., R.D. and A.K. performed computational analyses. J.K. wrote the manuscript.

Appendix A. Supplemental data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.mbplus.2019.04.001>.

Received 26 November 2018;

Received in revised form 2 April 2019;

Accepted 2 April 2019

Available online 06 April 2019

Keywords:

Pan-Cancer;
Tumor;
Matrix proper;
Gene regulation;
Prognostic;
Therapy

References

- [1] V. Gkretsi, T. Stylianopoulos, Cell Adhesion and Matrix Stiffness: Coordinating Cancer Cell Invasion and Metastasis, *Front. Oncol.* 8 (2018) 145.
- [2] C. Rianna, P. Kumar, M. Radmacher, The role of the microenvironment in the biophysics of cancer, *Semin. Cell Dev. Biol.* 73 (2018) 107–114.
- [3] M. Wang, B. Yin, H.Y. Wang, R.F. Wang, Current advances in T-cell-based cancer immunotherapy, *Immunotherapy* 6 (2014) 1265–1278.
- [4] R. Iozzo, R.A. Gubbiotti, Extracellular matrix: The driving force of mammalian diseases, *Matrix Biol.* 71–72 (2018) 1–9.
- [5] T.L. Whiteside, The tumor microenvironment and its role in promoting tumor growth, *Oncogene* 27 (2008) 5904–5912.
- [6] O.M.T. Pearce, R.M. Delaine-Smith, E. Maniati, S. Nichols, J. Wang, S. Bohm, V. Rajeeve, D. Ullah, P. Chakravarty, R.R. Jones, A. Montfort, T. Dowe, J. Gribben, J.L. Jones, H.M. Kocher, J.S. Serody, B.G. Vincent, J. Connelly, J.D. Brenton, C. Chelala, P.R. Cutillas, M. Lockley, C. Bessant, M.M. Knight, F.R. Balkwill, Deconstruction of a Metastatic Tumor Microenvironment Reveals a Common Matrix Response in Human Cancers, *Cancer. Discov.* 8 (2018) 304–319.
- [7] Z. Cao, S. Zhang, An integrative and comparative study of pan-cancer transcriptomes reveals distinct cancer common and specific signatures, *Sci. Rep.* 6 (2016) 33398.
- [8] D. Jia, Z. Liu, N. Deng, T.Z. Tan, R.Y. Huang, B. Taylor-Harding, D.J. Cheon, K. Lawrenson, W.R. Wiedemeyer, A.E. Walts, B.Y. Karlan, S. Orsulic, A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets, *Cancer Lett.* 382 (2016) 203–214.
- [9] A. Naba, O.M.T. Pearce, A. Del Rosario, D. Ma, H. Ding, V. Rajeeve, P.R. Cutillas, F.R. Balkwill, R.O. Hynes, Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics, *J. Proteome Res.* 16 (2017) 3083–3091.
- [10] A. Naba, K.R. Clauser, H. Ding, C.A. Whittaker, S.A. Carr, R. O. Hynes, The extracellular matrix: Tools and insights for the "omics" era, *Matrix Biol.* 49 (2016) 10–24.
- [11] F.R. Balkwill, M. Capasso, T. Hagemann, The tumor microenvironment at a glance, *J. Cell Sci.* 125 (2012) 5591–5596.
- [12] D.M. Gilkes, G.L. Semenza, D. Wirtz, Hypoxia and the extracellular matrix: drivers of tumour metastasis, *Nat. Rev. Cancer* 14 (2014) 430–439.
- [13] P. Lu, V.M. Weaver, Z. Werb, The extracellular matrix: a dynamic niche in cancer progression, *J. Cell Biol.* 196 (2012) 395–406.
- [14] J.R. Todd, K.A. Ryall, S. Vyse, J.P. Wong, R.C. Natrajan, Y. Yuan, A.C. Tan, P.H. Huang, Systematic analysis of tumour cell-extracellular matrix adhesion identifies independent prognostic factors in breast cancer, *Oncotarget* 7 (2016) 62939–62953.
- [15] N.E. Reticker-Flynn, D.F. Malta, M.M. Winslow, J.M. Lamar, M.J. Xu, G.H. Underhill, R.O. Hynes, T.E. Jacks, S.N. Bhatia, A combinatorial extracellular matrix platform identifies cell-extracellular matrix interactions that correlate with metastasis, *Nat. Commun.* 3 (2012) 1122.
- [16] J. Cathcart, A. Pulkoski-Gross, J. Cao, Targeting Matrix Metalloproteinases in Cancer: Bringing New Life to Old Ideas, *Genes Dis.* 2 (2015) 26–34.
- [17] A.K. Ravindranath, S. Kaur, R.P. Wernyj, M.N. Kumaran, K. E. Miletti-Gonzalez, R. Chan, E. Lim, K. Madura, L.

- Rodriguez-Rodriguez, CD44 promotes multi-drug resistance by protecting P-glycoprotein from FBXO21-mediated ubiquitination, *Oncotarget* 6 (2015) 26308–26321.
- [18] L. Angenendt, S. Reuter, D. Kentrup, A.S. Benk, F. Neumann, J. Huve, A.C. Martens, C. Schwoppe, T. Kessler, L.H. Schmidt, T. Sauer, C. Brand, J.H. Mikesch, G. Lenz, R.M. Mesters, C. Muller-Tidow, W. Hartmann, E. Wardelmann, D. Neri, W.E. Berdel, C. Roesli, C. Schliemann, An atlas of bloodstream-accessible bone marrow proteins for site-directed therapy of acute myeloid leukemia, *Leukemia* 32 (2018) 510–519.
- [19] F.A. Venning, L. Wullkopf, J.T. Erler, Targeting ECM Disrupts Cancer Progression, *Front. Oncol.* 5 (2015) 224.
- [20] A.W. Holle, J.L. Young, J.P. Spatz, In vitro cancer cell-ECM interactions inform in vivo cancer treatment, *Adv. Drug Deliv. Rev.* 97 (2016) 270–279.
- [21] S.V. Vasaiakar, P. Straub, J. Wang, B. Zhang, LinkedOmics: analyzing multi-omics data within and across 32 cancer types, *Nucleic Acids Res.* 46 (2018) D956–D963.
- [22] K. Tomczak, P. Czerwinska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, *Contemp. Oncol. (Pozn)* 19 (2015) A68–A77.
- [23] The Cancer Genome Atlas Research Network, Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia, *N. Engl. J. Med.* 368 (2013) 2059–2074.
- [24] J.J. Hao, X. Zhi, Y. Wang, Z. Zhang, Z. Hao, R. Ye, Z. Tang, F. Qian, Q. Wang, J. Zhu, Comprehensive Proteomic Characterization of the Human Colorectal Carcinoma Reveals Signature Proteins and Perturbed Pathways, *Sci. Rep.* 7 (2017) 42436.
- [25] V. Izzi, J. Lakkala, R. Devarajan, H. Ruotsalainen, E.R. Savolainen, P. Koistinen, R. Heljasvaara, T. Pihlajaniemi, An extracellular matrix signature in leukemia precursor cells and acute myeloid leukemia, *Haematologica* 102 (2017) e245–e248.
- [26] L.E. Stevens, W.K.C. Cheung, S.J. Adua, A. Arnal-Estape, M. Zhao, Z. Liu, K. Brewer, R.S. Herbst, D.X. Nguyen, Extracellular Matrix Receptor Expression in Subtypes of Lung Adenocarcinoma Potentiates Outgrowth of Micrometastases, *Cancer Res.* 77 (2017) 1905–1917.
- [27] J. Hutchenreuther, K. Vincent, C. Norley, M. Racanelli, S.B. Gruber, T.M. Johnson, D.R. Fullen, L. Raskin, B. Perbal, D. W. Holdsworth, L.M. Postovit, A. Leask, Activation of cancer-associated fibroblasts is required for tumor neovascularization in a murine model of melanoma, *Matrix Biol.* 74 (2018) 52–61.
- [28] Cancer Genome Atlas Research Network, J.N. Weinstein, E. A. Collisson, G.B. Mills, K.R. Shaw, B.A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J.M. Stuart, The Cancer Genome Atlas Pan-Cancer analysis project, *Nat. Genet.* 45 (2013) 1113–1120.
- [29] K.A. Hoadley, C. Yau, D.M. Wolf, A.D. Cherniack, D. Tamborero, S. Ng, M.D.M. Leiserson, B. Niu, M.D. McLellan, V. Uzunangelov, J. Zhang, C. Kandoth, R. Akbani, H. Shen, L. Omberg, A. Chu, A.A. Margolin, L.J. Van't Veer, N. Lopez-Bigas, P.W. Laird, B.J. Raphael, L. Ding, A.G. Robertson, L.A. Byers, G.B. Mills, J.N. Weinstein, C. Van Waes, Z. Chen, E.A. Collisson, Cancer Genome Atlas Research Network, C.C. Benz, C.M. Perou, J.M. Stuart, Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin, *Cell* 158 (2014) 929–944.
- [30] D. Aran, M. Sirota, A.J. Butte, Systematic pan-cancer analysis of tumour purity, *Nat. Commun.* 6 (2015) 8971.
- [31] T.R. Cox, J.T. Erler, Molecular pathways: connecting fibrosis and solid tumor metastasis, *Clin. Cancer Res.* 20 (2014) 3637–3643.
- [32] T.R. Cox, J.T. Erler, Fibrosis and Cancer: Partners in Crime or Opposing Forces? *Trends Cancer.* 2 (2016) 279–282.
- [33] A. Naba, K.R. Clauser, C.A. Whittaker, S.A. Carr, K.K. Tanabe, R.O. Hynes, Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver, *BMC Cancer* 14 518-2407-14-518, 2014.
- [34] R.D. Leclerc, Survival of the sparsest: robust gene networks are parsimonious, *Mol. Syst. Biol.* 4 (2008) 213.
- [35] S. Manandhar, Y.M. Lee, Emerging role of RUNX3 in the regulation of tumor microenvironment, *BMB Rep.* 51 (2018) 174–181.
- [36] J.E. Hunter, J. Leslie, N.D. Perkins, c-Rel and its many roles in cancer: an old story with new twists, *Br. J. Cancer* 114 (2016) 1–6.
- [37] M. Cassandri, A. Smirnov, F. Novelli, C. Pitolli, M. Agostini, M. Malewicz, G. Melino, G. Raschella, Zinc-finger proteins in health and disease, *Cell Death Dis.* 3 (2017), 17071.
- [38] E. Ayroldi, L. Cannarile, D.V. Delfino, C. Riccardi, A dual role for glucocorticoid-induced leucine zipper in glucocorticoid function: tumor growth promotion or suppression, *Cell. Death Dis.* 9 (2018), 463-018-0558-1.
- [39] B.P. Purkayastha, J.K. Roy, Cancer cell metabolism and developmental homeodomain/POU domain transcription factors: a connecting link, *Cancer Lett.* 356 (2015) 315–319.
- [40] J. Youssef, M. Badr, Peroxisome proliferator-activated receptors and cancer: challenges and opportunities, *Br. J. Pharmacol.* 164 (2011) 68–82.
- [41] D. Radisky, J. Muschler, M.J. Bissell, Order and disorder: the role of extracellular matrix in epithelial cancer, *Cancer Investig.* 20 (2002) 139–153.
- [42] S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T.R. Hughes, M.T. Weirauch, The Human Transcription Factors, *Cell* 172 (2018) 650–665.
- [43] T. Schacht, M. Oswald, R. Eils, S.B. Eichmuller, R. Konig, Estimating the activity of transcription factors by the effect on their target genes, *Bioinformatics* 30 (2014) i401–i407.
- [44] M.H. Bailey, C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, A. Weerasinghe, A. Colaprico, M.C. Wendl, J. Kim, B. Reardon, P.K. Ng, K.J. Jeong, S. Cao, Z. Wang, J. Gao, Q. Gao, F. Wang, E.M. Liu, L. Mularoni, C. Rubio-Perez, N. Nagarajan, I. Cortes-Ciriano, D.C. Zhou, W.W. Liang, J.M. Hess, V.D. Yellapantula, D. Tamborero, A. Gonzalez-Perez, C. Suphavitai, J.Y. Ko, E. Khurana, P.J. Park, E.M. Van Allen, H. Liang, MC3 Working Group, Cancer Genome Atlas Research Network, M.S. Lawrence, A. Godzik, N. Lopez-Bigas, J. Stuart, D. Wheeler, G. Getz, K. Chen, A.J. Lazar, G. B. Mills, R. Karchin, L. Ding, Comprehensive Characterization of Cancer Driver Genes and Mutations, *Cell* 173 (2018) 371–385.e18.
- [45] C. Cava, G. Bertoli, A. Colaprico, C. Olsen, G. Bontempi, I. Castiglioni, Integration of multiple networks and pathways identifies cancer driver genes in pan-cancer analysis, *BMC Genomics* 19 25-017-4423-x, 2018.
- [46] H. Kim, Y.M. Kim, Pan-cancer analysis of somatic mutations and transcriptomes reveals common functional gene clusters shared by multiple cancer types, *Sci. Rep.* 8 (2018), 6041-018-24379-y.
- [47] M. Uroz, S. Wistorf, X. Serra-Picamal, V. Conte, M. Sales-Pardo, P. Roca-Cusachs, R. Guimera, X. Trepas, Regulation of cell cycle progression by cell-cell and cell-matrix forces, *Nat. Cell Biol.* 20 (2018) 646–654.

- [48] C. Gerard, A. Goldbeter, Dynamics of the mammalian cell cycle in physiological and pathological conditions, *Wiley Interdiscip. Rev. Syst. Biol. Med.* 8 (2016) 140–156.
- [49] B. Baum, M. Georgiou, Dynamics of adherens junctions in epithelial establishment, maintenance, and remodeling, *J. Cell Biol.* 192 (2011) 907–917.
- [50] N. Defamie, A. Chepied, M. Mesnil, Connexins, gap junctions and tissue invasion, *FEBS Lett.* 588 (2014) 1331–1338.
- [51] M. Hornsveld, T.B. Dansen, P.W. Derksen, B.M.T. Burgering, Re-evaluating the role of FOXOs in cancer, *Semin. Cancer Biol.* 50 (2018) 90–100.
- [52] J.H. Kim, S.Y. Park, Y. Jun, J.Y. Kim, J.S. Nam, Roles of Wnt Target Genes in the Journey of Cancer Stem Cells, *Int. J. Mol. Sci.* 18 (2017) <https://doi.org/10.3390/ijms18081604>.
- [53] J. Du, Y. Zu, J. Li, S. Du, Y. Xu, L. Zhang, L. Jiang, Z. Wang, S. Chien, C. Yang, Extracellular matrix stiffness dictates Wnt expression through integrin pathway, *Sci. Rep.* 6 (2016) 20395.
- [54] Z.J. Rutnam, T.N. Wight, B.B. Yang, miRNAs regulate expression and function of extracellular matrix molecules, *Matrix Biol.* 32 (2013) 74–85.
- [55] J. Chou, P. Shahi, Z. Werb, microRNA-mediated regulation of the tumor microenvironment, *Cell Cycle* 12 (2013) 3262–3271.
- [56] M.W. Pickup, J.K. Mouw, V.M. Weaver, The extracellular matrix modulates the hallmarks of cancer, *EMBO Rep.* 15 (2014) 1243–1253.
- [57] E. Powell, D. Piwnica-Worms, H. Piwnica-Worms, Contribution of p53 to metastasis, *Cancer. Discov.* 4 (2014) 405–414.
- [58] E. Brown, S. Hua, P.S. Tanwar, Extracellular matrix-mediated regulation of cancer stem cells and chemoresistance, *Int. J. Biochem. Cell Biol.* 109 (2019) 90–104.
- [59] M. Brahmi, O. Bally, L. Eberst, P. Cassier, Therapeutic targeting of Notch signaling in cancer, *Bull. Cancer* 104 (2017) 883–891.
- [60] L. Li, P. Tang, S. Li, X. Qin, H. Yang, C. Wu, Y. Liu, Notch signaling pathway networks in cancer metastasis: a new target for cancer therapy, *Med. Oncol.* 34 (2017), 180-017-1039-6.
- [61] C. Bellomo, L. Caja, A. Moustakas, Transforming growth factor beta as regulator of cancer stemness and metastasis, *Br. J. Cancer* 115 (2016) 761–769.
- [62] M.A. Cichon, D.C. Radisky, Extracellular matrix as a contextual determinant of transforming growth factor-beta signaling in epithelial-mesenchymal transition and in cancer, *Cell Adhes. Migr.* 8 (2014) 588–594.
- [63] G.L. Semenza, The hypoxic tumor microenvironment: A driving force for breast cancer progression, *Biochim. Biophys. Acta* 1863 (2016) 382–391.
- [64] S. Kaushik, M.W. Pickup, V.M. Weaver, From transformation to metastasis: deconstructing the extracellular matrix in breast cancer, *Cancer Metastasis Rev.* 35 (2016) 655–667.
- [65] V. Ramovs, L. Te Molder, A. Sonnenberg, The opposing roles of laminin-binding integrins in cancer, *Matrix Biol.* 57-58 (2017) 213–243.
- [66] K.C. Cotto, A.H. Wagner, Y.Y. Feng, S. Kiwala, A.C. Coffman, G. Spies, A. Wollam, N.C. Spies, O.L. Griffith, M. Griffith, DGIdb 3.0: a redesign and expansion of the drug-gene interaction database, *Nucleic Acids Res.* 46 (2018) D1068–D1073.
- [67] D.F. Quail, R.L. Bowman, L. Akkari, M.L. Quick, A.J. Schuhmacher, J.T. Huse, E.C. Holland, J.C. Sutton, J.A. Joyce, The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas, *Science* 352 (2016) aad3018.
- [68] J. Menke, J. Kriegsmann, C.C. Schimanski, M.M. Schwartz, A. Schwarting, V.R. Kelley, Autocrine CSF-1 and CSF-1 receptor coexpression promotes renal cell carcinoma growth, *Cancer Res.* 72 (2012) 187–200.
- [69] V. Thorsson, D.L. Gibbs, S.D. Brown, D. Wolf, D.S. Bortone, T.H. Ou Yang, E. Porta-Pardo, G.F. Gao, C.L. Plaisier, J.A. Eddy, E. Ziv, A.C. Culhane, E.O. Paull, I.K.A. Sivakumar, A.J. Gentles, R. Malhotra, F. Farshidfar, A. Colaprico, J.S. Parker, L.E. Mose, N.S. Vo, J. Liu, Y. Liu, J. Rader, V. Dhankani, S.M. Reynolds, R. Bowlby, A. Califano, A.D. Cherniack, D. Anastassiou, D. Bedognetti, A. Rao, K. Chen, A. Krasnitz, H. Hu, T.M. Malta, H. Noushmehr, C.S. Pedamallu, S. Bullman, A.I. Ojesina, A. Lamb, W. Zhou, H. Shen, T.K. Choueiri, J.N. Weinstein, J. Guinney, J. Saltz, R.A. Holt, C.E. Rabkin, Cancer Genome Atlas Research Network, A.J. Lazar, J.S. Serody, E.G. Demicco, M.L. Disis, B.G. Vincent, L. Shmulevich, The Immune Landscape of Cancer, *Immunity* 48 (2018) 812–830.e14.
- [70] M. Wang, J. Zhao, L. Zhang, F. Wei, Y. Lian, Y. Wu, Z. Gong, S. Zhang, J. Zhou, K. Cao, X. Li, W. Xiong, G. Li, Z. Zeng, C. Guo, Role of tumor microenvironment in tumorigenesis, *J. Cancer* 8 (2017) 761–773.
- [71] J. Vivian, A.A. Rao, F.A. Nothhaft, C. Ketchum, J. Armstrong, A. Novak, J. Pfeil, J. Narkizian, A.D. Deran, A. Musselman-Brown, H. Schmidt, P. Amstutz, B. Craft, M. Goldman, K. Rosenbloom, M. Cline, B. O'Connor, M. Hanna, C. Birger, W. J. Kent, D.A. Patterson, A.D. Joseph, J. Zhu, S. Zaranek, G. Getz, D. Haussler, B. Paten, Toil enables reproducible, open source, big biomedical data analyses, *Nat. Biotechnol.* 35 (2017) 314–316.
- [72] H. Han, J.W. Cho, S. Lee, A. Yun, H. Kim, D. Bae, S. Yang, C. Y. Kim, M. Lee, E. Kim, S. Lee, B. Kang, D. Jeong, Y. Kim, H. N. Jeon, H. Jung, S. Nam, M. Chung, J.H. Kim, I. Lee, TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions, *Nucleic Acids Res.* 46 (2018) D380–D386.
- [73] ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, *Nature* 489 (2012) 57–74.
- [74] D. Marbach, D. Lamparter, G. Quon, M. Kellis, Z. Kutalik, S. Bergmann, Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases, *Nat. Methods* 13 (2016) 366–370.
- [75] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B. L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 15545–15550.
- [76] FANTOM Consortium and the RIKEN PMI and CLST (DGT), A.R. Forrest, H. Kawaji, M. Rehli, J.K. Baillie, M.J. de Hoon, V. Haberle, T. Lassmann, I.V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C.J. Mungall, T.F. Meehan, S. Schmeier, N. Bertin, M. Jorgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y.A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. Semple, Y. Ishizu, R.S. Young, M. Francescato, I. Alam, D. Albanese, G.M. Altschuler, T. Arakawa, J.A. Archer, P. Arner, M. Babina, S. Rennie, P.J. Balwiercz, A.G. Beckhouse, S. Pradhan-Bhatt, J.A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A.M. Burroughs, A. Califano, C.V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H.C. Clevers, E. Dalla, C.A. Davis, M.

- Detmar, A.D. Diehl, T. Dohi, F. Drablos, A.S. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M.C. Farach-Carson, G.J. Faulkner, A.V. Favorov, M.E. Fisher, M.C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furino, J. Furusawa, T.B. Geijtenbeek, A.P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T.J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K.J. Hitchens, S.J. Ho Sui, O.M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A.S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y.I. Kawamura, T. Kawashima, J.S. Kempfle, T.J. Kenna, J. Kere, L.M. Khachigian, T. Kitamura, S.P. Klinken, A.J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A.T. Kwon, J.F. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-Sim, R. Manabe, J.C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D.A. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C.L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, S. Noma, T. Nozaki, S. Ogishima, N. Ohkura, H. Ohimiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D.A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J.G. Prendergast, O.J. Rackham, J.A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M.B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, S. Savvi, A. Saxena, C. Schneider, E.A. Schultes, G.G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J.W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R.K. Swoboda, P.A. Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyodo, T. Toyoda, E. Valen, M. van de Wetering, L.M. van den Berg, R. Verado, D. Vijayan, I.E. Vorontsov, W.W. Wasserman, S. Watanabe, C.A. Wells, L.N. Winteringham, E. Wolvetang, E.J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S.E. Zabierowski, P.G. Zhang, X. Zhao, S. Zucchelli, K.M. Summers, H. Suzuki, C.O. Daub, J. Kawai, P. Heutink, W. Hide, T.C. Freeman, B. Lenhard, V.B. Bajic, M.S. Taylor, V.J. Makeev, A. Sandelin, D.A. Hume, P. Carninci, Y. Hayashizaki, A promoter-level mammalian expression atlas, *Nature* 507 (2014) 462–470.
- [77] I. Abugessaisa, H. Shimoji, S. Sahin, A. Kondo, J. Harshbarger, M. Lizio, Y. Hayashizaki, P. Carninci, FANTOM consortium, A. Forrest, T. Kasukawa, H. Kawaji, FANTOM5 transcriptome catalog of cellular states based on Semantic MediaWiki, Database (Oxford) 2016 10.1093/database/baw105. Print 2016, 2016.
- [78] A. Chatr-Aryamontri, R. Oughtred, L. Boucher, J. Rust, C. Chang, N.K. Kolas, L. O'Donnell, S. Oster, C. Theesfeld, A. Sellam, C. Stark, B.J. Breitkreutz, K. Dolinski, M. Tyers, The BioGRID interaction database: 2017 update, *Nucleic Acids Res.* 45 (2017) D369–D379.
- [79] M. Uhlen, L. Fagerberg, B.M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigartyo, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwaan, G. von Heijne, J. Nielsen, F. Ponten, Proteomics. Tissue-based map of the human proteome, *Science* 347 (2015) 1260419.
- [80] J. Sun, Q. Wei, Y. Zhou, J. Wang, Q. Liu, H. Xu, A systematic analysis of FDA-approved anticancer drugs, *BMC Syst. Biol.* (2017) (11 87-017-0464-7).
- [81] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, *Nucleic Acids Res.* 45 (2017) D362–D368.
- [82] D.R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, A.M. Chinnaiyan, ONCOMINE: a cancer microarray database and integrated data-mining platform, *Neoplasia* 6 (2004) 1–6.
- [83] M. Greiner, D. Pfeiffer, R.D. Smith, Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests, *Prev. Vet. Med.* 45 (2000) 23–41.
- [84] H. Zou, The Adaptive Lasso and Its Oracle Properties, *J. Am. Stat. Assoc.* 101 (2006) 1418–1429.
- [85] B. Aragam, J. Gu, Q. Zhou, Learning Large-Scale Bayesian Networks with the sparsebn Package, arXiv (2018) 1703.04025.
- [86] J.M.B. Haslbeck, L.J. Waldorp, MGM: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data, arXiv (2018) 1510.06871.