HUMAN BIOLOGY
2024, VOL. 94, ISSUE 4
ORIGINAL ARTICLE

OPEN ACCESS

Bulk and Single-Cell Transcriptional Profiles Reveal Roles of Fibroblasts and Immunocytes in Pan-Cancer Progression

Yan Sun [1,2,3], *Bin Song* [1,2], *Qichao Yu* [1,2], *Huanming Yang* [1,2,3], *Wei Dong* [3,1,2,*]

ABSTRACT

Tumors carry various dysregulated genes, of which many are found to be related to the overall survival of patients. These dysregulated genes are usually identified by bulk transcriptional comparison between tumors and their matching non-tumor tissues. However, because tumor tissues usually contain stromal cells in addition to cancer cells, it remains unclear whether the stromal cells within tumors also carry dysregulated genes. Here, to address this question, we combine bulk and single-cell gene expression data of tumor, adjacent and non-tumor tissues from 7 organs to explore the molecular and cellular mechanism of cancer progression. We found that fibroblasts within tumors across 7 cancer types commonly carry multiple dysregulated genes related to the overall survival of patients. Cell-cell communication analysis revealed significant interactions between cytotoxic immune cells and cancer fibroblasts through the PARs pathway, and self-activation of cancer associated fibroblasts (CAFs) via the PERIOSTIN pathway in pan-cancer. We also identified Colon cancer specific cycling B cells, which influence patients' survival. Our study provides potential targets for pan-cancer therapy.

INTRODUCTION

Cancer is thought to be caused by mutated or dysregulated expression of oncogenes, tumor-suppressor genes or non-coding genes (Basu, 2018; Croce, 2008). Dysregulated expression of some genes represent general features in different cancer types Hufton et al. (1999), Kettunen et al. (2004), Xu et al. (2000), Zaravinos et al. (2011), Delakas et al. (2011), and help researchers to understand tumor biology and predict patients' survival (Rosario et al., 2018; Xue, Liu, Wan, & Zhu, 2020). The method to identify dysregulated or differentially expressed genes (DEGs) relies on comparison between non-tumor and tumor tissues, which is composed of multiple cell types including malignant cells, immune cells, stromal cells, and extra cellular matrix (ECM), by which cell-cell and cell-matrix communications are established (Dominiak, Chelstowska, Olejarz, & Nowicka, 2020; Garner & Visser, 2020; Schwager, Taufalele, & Reinhart-King, 2019). The existence of multiple cell types (malignant cells and stromal cells such as fibroblasts or immune cells) within cancer tissues makes it hard to tell whether the DEGs found in tumors are from malignant cells. Additionally, the change of cell ratio within the tissue also influences bulk gene expression levels. Single-cell technology has accelerated cancer research for its power to decipher the cellular and molecular landscape of tumor tissues.

A number of cancer single-cell atlases have been published to characterize the cellular heterogeneity Kumar et al. (2022), Wu et al. (2021), profile cancer immune microenvironments Binnewies et al. (2018), Leun et al. (2020), and unveil the mechanism of metastasis Lawson et al. (2018). However, current cancer single-cell studies focus on one cancer type or a limited number of patients, which ignores the diversity among cancers. Cancer is a heterogeneous disease Marusyk et al. (2010) with multiple subtypes based on the cell of origin, the expression of specific molecular markers, or the genetic aberrations Arora et al. (2019), Huvila et al. (2021), Kim et al. (2019), Marisa et al. (2013), Network et al. (2015), Parker et al. (2009), Prete et al. (2020), Rudin et al. (2019), Sia et al. (2017), Skibinski et al. (2015), West et al. (2012). Even tumors originate from the same organ and even if histologically they appear similar, their behavior and response to therapy can be different Cusnir et al. (2012). In a study on a cohort of 25 high-risk prostate tumors, researchers observed outlier transcripts in each tumor, which were associated with cell cycle, translational control or immune regulation Wyatt et al. (2014). Publicly available database such as The Cancer Genome Atlas (TCGA) program Network et al. (2013) has collected tens of thousands of bulk samples and adopts unified standards that ensure comparability between samples.

[1] College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.
[2] BGI Research, Shenzhen 518083, China.
[3] HIM-BGI Omics Center, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences (CAS), Hangzhou 310022, China

Correspondence to: Wei Dong, HIM-BGI Omics Center, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences (CAS), Hangzhou 310022, China. Email: dongw@bgi.com.

A total of 33 cancer types are included in this program, and researchers are able to rule out signatures caused by cancer heterogeneity.

Because no one single-cell study has covered so many cancer types and collecting so many samples with unified standards, if researchers want to gain a multidimensional understanding of the molecular signature of pan-cancer, it is necessary to combine bulk data covering plenty of heterogeneous patients and single-cell data that provide cellular-level insights. By integrating bulk and single-cell data, researchers have explored clonal architecture of tumors Malikic et al. (2019), and identified immune infiltration related genes in cholangiocarcinoma Chen et al. (2021).

It is recently recognized that non-tumor tissue adjacent to the tumor is not a good control for tumor study, because genes in the adjacent tissues could be activated by stimuli such as growth factors, hormones, or stress produced by tumors Dvir Aran et al. (2017). And researchers imported samples from The Genotype-Tissue Expression (GTEx) program Lonsdale et al. (2013) as control, which are collected from tumor-free individuals.

In this study, we collected TCGA bulk RNA samples of 7 cancer types, which contain > 20 adjacent samples and have clear tissue origins of tumor, to identify dysregulated genes in pan-cancer. In addition, we collected non-tumor tissues from GTEx of corresponding organs as extra controls to exclude tumor adjacent tissue specifically expressed genes. We examined cell origin of commonly dysregulated genes in single-cell data, and explored cell-cell communications. We applied weighted correlation network analysis (WGCNA) Langfelder et al. (2008) on bulk data to investigate correlations between dysregulated genes and cell-cell communication pathways. By performing survival analysis, we linked gene expression levels to cancer progression, and investigated the underlying mechanism that leads to opposite prognostic effects of genes in different cancer types.

MATERIALS AND METHODS

Data collection

We downloaded gene expression data of 3975 TCGA tumor samples, 408 TCGA adjacent samples and 1490 GTEx non-tumor samples from 7 organs (Breast, Colon, Liver, Lung, Prostate, Thyroid and Uterus), which has been re-quantified using identical analysis pipeline to remove batch effects caused by software Dvir Aran et al. (2017), Rahman et al. (2015) (see Data availability).

We collected 20 tumor single-cell samples of these 7 cancer types Dong et al. (2020), Luo et al. (2022), Luo et al. (2021), Ma et al. (2021), Pal et al. (2021), Zeng et al. (2022), 23 non-tumor tissue single- cell samples of six corresponding organs (Breast, Colon, Liver, Lung, Prostate and Uterus) Garcia-Alonso et al. (2021), Gray et al. (2022), Henry et al. (2018), MacParland et al. (2018), Madissoon et al. (2023), Smillie et al. (2019), Vilella et al. (2021), Wang et al. (2020) from tumor-free individuals and three Prostate non-tumor single-cell samples from Prostate cancer patients (Tuong et al., 2021) (see Data availability).

DEG analysis

DEG analysis was applied referring to Dvir Aran et al. (2017). The gene count data was used and we retained genes with count >= 10 in at least 2 samples. In the analysis of TCGA tumor tissues-vs-GTEx non-tumor tissues, we first employed upper-quantile correlation and then the RUVg method from the RUVSeq package Risso et al. (2014) (Version 1.16.1) to remove batch effects. The RUVg method corrected expressions based on a list of housekeeping genes (PSMB2, PSMB4, C1orf43, RAB7A, REEP5, VCP, VPS29, C15orf24, CHMP2A, SNRPD3), which were suggested by Eli Eisenberg et al. (2013). We then used the edgeR package (Robinson, McCarthy, & Smyth, 2010) (Version 3.24.3) to find DEGs with log2-fold change > 1 or < -1, log2-CPM > 3, FDR < 0.05.

In the comparison of TCGA tumor tissues-vs-TCGA adjacent tissues, edgeR was performed directly.

Gene functional enrichment

We performed enrichment analysis using the clusterProfiler package Yu et al. (2012), He et al. (2012) (Version 4.6.2) and the org.Hs.eg.db database Carlson et al. (2019), Li et al. (2019) (Version 3.16.0).

Protein-protein interaction analysis

PPI analysis was performed on the online database STRING (https://string-db.org) Snel et al. (2000), Huynen et al. (2000).

xCell

Cell type scores in bulk samples were calculated by the xCellAnalysis function from the xCell R package Aran et al. (2017) (Version 1.1.0) with default parameters, and the count matrix was used.

Single-cell clustering and visualization

The Seurat package Hao et al. (2021), Satija et al. (2015), Regev et al. (2015) (Version 4.3.0) was used to cluster and visualize single-cells, and to find marker genes for each cluster identified. In most of the single-cell datasets, quality control of cells has been finished (see Data availability). And we used the subset function to remove low-quality cells with number of expressed genes < 500 or > 3000, or percentage of mitochondrial reads > 20% in Breast cancer samples GSE161529_GSM4909283_TN_0106,

GSE161529_GSM4909306_ER_0029_9C, Prostate samples GSE120716_D17, GSE120716_D27, GSE120716_Pd, and low-quality cells with number of expressed genes < 500 or > 5000, or percentage of mitochondrial reads > 20% in Breast cancer sample GSE161529_GSM4909289_HER2_0308.

We log-normalized count data with scale.factor = 1e4, and chose 2000 highly variable genes with selection.method = "vst". PCA was performed with the 2000 variable genes and we retained top 10 PCs with the highest standard deviations for following analyses. We identified clusters with resolution = 0.5, and used UMAP (McInnes, Healy, & Melville, 2018) to visualize single-cells.

The FindAllMarkers function was used to find highly and specifically expressed genes of each identified cluster with parameters only.pos = TRUE, min.pct = 0.25 and logfc.threshold = 0.25.

We annotated cell types for clusters based on meta-data downloaded from papers where these single-cell datasets were published (see Data availability), and based on marker genes provided by CancerSCEM (https://ngdc.cncb.ac.cn/cancerscem/documents) Zeng et al. (2022).

Survival analysis

Survival analysis was performed based on gene expressions or GSVA scores of gene sets using the survival package Therneau et al. (2015), Lumley et al. (2015) (Version 3.5.5). We first classified patients into two groups, one group lowly expressed, and the other group highly express the target gene/gene set. To choose the best expression cut-offs for grouping the patients most significantly, all log2-CPM/GSVA values from the 25th to 75th percentiles were used to calculate a log-rank P value Nagy et al. (2021), Uhlen et al. (2017) with the survdiff function, and the percentile yielding the lowest P value was selected. Then we used the survfit function to perform survival analysis, which was visualized with the ggsurvplot function from the survminer package Kassambara et al. (2017), Fabian et al. (2017) (Version 0.4.9) with log.rank.weights = '1'.

The x-axis is days after diagnosed, the y-axis is the percentage of patients alive. The colored area around the curve indicates the confidence interval.

If there's a higher observed event than expected event in the group of patients with high expression of a selected gene, it is recognized as an unfavorable prognostic gene; otherwise, it is a favorable prognostic gene. Uhlen et al. (2017)

CellChat

We used the CellChat package Jin et al. (2021) (Version 1.6.1) to analyze cell communications in single-cell samples based on ligand-receptor interaction. We used log2-CPM as input. The analysis pipeline was the same as the tutorial 'Full tutorial for CellChat analysis of a single dataset with detailed explanation of each function' provided by the developers (https://github.com/sqjin/CellChat).

WGCNA

We applied the WGCNA package Langfelder et al. (2008), Horvath et al. (2008) (Version 1.69-81) to bulk gene expression data to detect gene co-expression modules. The count data was first normalized with the varianceStabilizingTransformation function from the DESeq2 package Love et al. (2014), Anders et al. (2014) (Version 1.22.2). Then we used a one-step network construction function blockwiseModules to detect modules with fixed parameters TOMType = "unsigned", minModuleSize = 5, reassignThreshold = 0, mergeCutHeight = 0.25. numericLabels = TRUE, pamRespectsDendro = FALSE, verbose = 3, maxBlockSize = 30000, and a data dependent parameter 'power', which were 14, 14, 14, 14, 14, 22, 14 for Breast, Colon, Liver, Lung, Prostate, Thyroid, Uterus cancers, respectively.

GSVA

We used the GSVA package Hanzelmann et al. (2013), Guinney et al. (2013), Subramanian et al. (2005) (Version 1.46.0) to evaluate the expression level of a gene set. Log2-CPMs were used as input. Then the gsva function was applied with the parameter kcdf="Gaussian", which returned the score.

Statistical test

Wilcoxon test was performed using the R command wilcox.test(), and the parameter alternative='greater' was set in the single-tailed test. Chi-squared test of pathway frequencies was performed using the R command chisq.test() with the parameter simulate.p.value = TRUE.

Analysis environment

Most of the analyses were performed in the R environment Team et al. (2013) (Version 4.2.0).

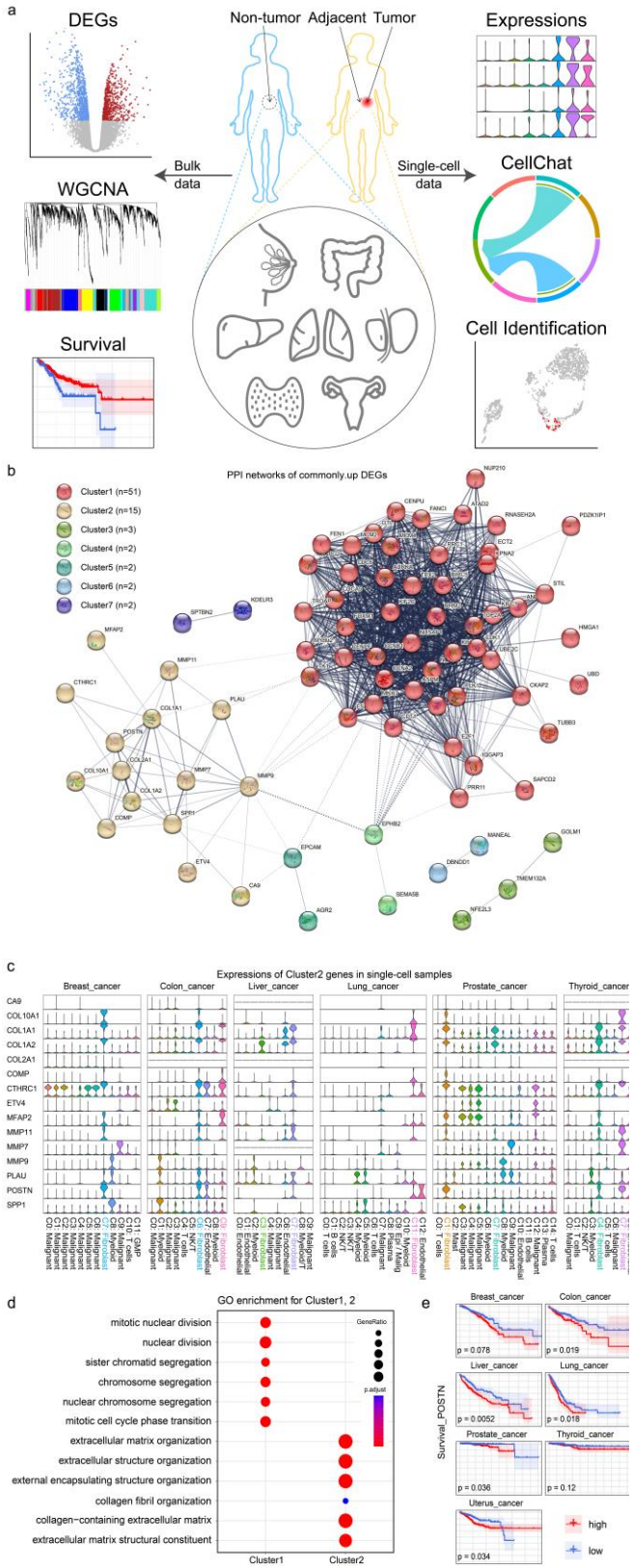
RESULTS

Dysregulated genes expressed by fibroblasts are common in pan-cancer

We used bulk samples to find DEGs, gene co-expression networks and to perform survival analysis,

which links gene expressions to cancer progressions. We used single-cell samples to investigate the contribution of each cell type to bulk gene expressions, to perform cell-cell communication analysis and to identify cell types (Fig.1a; Table S1).

Figure 1: Commonly up-regulated genes.



a: Workflow of this study. Three tissue types (non-tumor tissues from tumor free individuals, adjacent and tumor tissues from tumor patients) from 7 organs (Breast, Colon, Liver, Lung, Prostate, Thyroid, Uterus) are included. We collect both bulk and single-cell gene expression data, apply DEG analysis, co-expression network analysis (WGCNA), survival analysis on bulk data, and apply cell expression analysis, cell communication analysis (CellChat), cell type identification on single-cell data. b: PPI networks of commonly up-regulated genes. Seven clusters are identified by MCL clustering with inflation parameter = 2.5. Line thickness indicates the strength of data support. Nodes within a cluster are connected with solid lines, from different clusters with dashed lines. Disconnected proteins are hided. c: Expression levels of Cluster2 genes in six cancer single-cell samples: GSE161529_GSM4909306_ER_0029_9C (Breast cancer), CancerSCEM_CRC_016_08_1A (Colon cancer), GSE151530_H38 (Liver cancer), CancerSCEM_LUAD-003-11-1A (Lung cancer), GSE137829_P5 (Prostate cancer), ATC-WYF (Thyroid cancer). No fibroblast was identified in Uterus cancer samples. Fibroblasts are colored at the bottom. d: GO enrichment of Cluster1, 2. They are involved in cell cycle and ECM, respectively. e: Survival analysis by expressions of POSTN in TCGA cancer patients (see Methods). Prognostic effects are unfavorable in all the 7 cancer types.

Firstly, to investigate whether different cancer types share any feature on gene expression level, we performed differentially expressed gene (DEG) analysis Robinson et al. (2010) on tumor tissues from all the 7 organs against their corresponding adjacent and GTEx Lonsdale et al. (2013) tissues. To avoid identification of adjacent specifically activated genes immediate-early responding to stimuli such as growth factors, hormones, or stress produced by tumors Dvir Aran et al. (2017), we required DEGs to be up- or down- regulated in both tumor-vs-adjacent and tumor-vs-GTEx comparisons (log2-fold change > 1 or < -1, log2-CPM (count per million) > 3, FDR (false discovery rate) < 0.05). We identified 97 commonly up-regulated genes and 32 commonly down-regulated genes in >= 5 cancer types (which were just enough for functional enrichment, table S2.

According to protein-protein interaction (PPI) database Szklarczyk et al. (2019), GO enrichment Yu et al. (2012) and gene expressions in single-cell data, we found the majority (51 genes identified as Cluster1 by PPI, Fig.1b) of the commonly up-regulated genes regulated cell cycle (Fig. 1d) and were expressed by malignant cells (table S2). Besides, malignant cells also highly expressed MMP7 which is a member of matrix metalloproteinases (MMPs) and is involved in breakdown of ECM (Fig. 1c, table S2) Yokoyama et al. (2008), suggesting the ability of malignant cells to regulate extra cellular matrix directly.

The second largest group of the commonly up-regulated genes (15 genes identified as Cluster2 by PPI, including ECM associated genes COL10A1, COL1A1, COL1A2, COMP, CTHRC1, MFAP2, MMP11, POSTN, Fig.1b) are primarily expressed by fibroblasts (Fig.1c, table S2). Periostin encoded by POSTN is a ligand for ITGAV+ITGB3 and ITGAV+ITGB5 to support adhesion and migration of epithelial cells Gillan et al. (2002). Evidence shows POSTN can activate the TGF-β, PI3K/Akt, Wnt, RhoA/ROCK, NF-κB, MAPK and JAK pathways Wang et al. (2022), and play multiple functions in tumor development and progression, including activating invasion and metastasis, angiogenesis, resisting cell death, and avoiding immune destruction González-González et al. (2018), Alonso et al. (2018). Using survival analysis Therneau et al. (2015), Cynthia et al. (2015), we found the overexpression of POSTN produced unfavorable prognostic effects in all the 7 cancer types (though not significant in Breast and Thyroid cancers, Fig. 1e), suggesting the possibility of POSTN as a broad target for cancer therapy.

The commonly down-regulated genes were also primarily expressed by fibroblasts (table S2). The majority of them are involved in muscle activity and organization of ECM (Cluster1 and Cluster2, Fig. S1a, b), such as myosin genes MYH11 and MYL9 (Cluster1). Myosin is a structural component of muscle, while is recently recognized as a fundamental component during tumor genesis and progression Li et al. (2016), Yang et al. (2016), Ouderkirk et al. (2014), Krendel et al. (2014). Decreased expression levels of MYH11 in lung cancer patients were found to correlate with poor prognosis Nie et al. (2020). And MYL9 was reported to be low expressed in breast cancer, non-small cell lung cancer, and stomach adenocarcinoma, and to associate with immune infiltration and focal adhesion in these cancers Lv et al. (2022), Chen et al. (2022), Tan et al. (2014), Chen et al. (2014). This finding demonstrates fibroblasts are one of the origins of the dysregulation of muscle related components in cancers. We also found the down-regulated gene DCN (Cluster2) which is regarded as a tumor suppressor gene (Hu et al. (2021), Järvinen et al. (2015), Prince et al. (2015) expressed by fibroblasts (Table S2). Our results indicate fibroblasts play a fundamental role in the progression of multiple cancers.

DEGs found in tumors indicate cellular disorders or changes in cell ratios. To rule out tumor DEGs that were caused by the change in cell composition, we performed deconvolution that estimate levels of different cell types in bulk samples by calculating cell type scores in tumor, adjacent, non-tumor bulk samples by xCell Aran et al. (2017), which is an enrichment-based method incorporating marker gene signatures from multiple cancer types for pan-cancer deconvolution Tran et al. (2023). We found significantly higher epithelial scores and lower fibroblast scores in tumor tissues from most organs as compared no matter with adjacent or with non-tumor tissues (Fig. S1c). These results confirm the cancer commonly up-regulated genes expressed by fibroblasts are not caused by cellular disorders, while the other common DEGs might be a result of change in cell composition.

Activations of Cytotoxic immune cells produce effects on Fibroblasts which are correlated with dysregulated genes and unfavorable prognosis

Metastasis is the main cause of mortality in cancer patients Choi et al. (2018), Moon et al. (2018). While ECM is essential for tumor cell invasion and migration Brassart-Pasco et al. (2020), Stetler-Stevenson et al. (1993). The major source for ECM is fibroblasts no matter in normal or in cancer tissues Cusnir et al. (2012), Cavalcante et al. (2012), Xiong et al. (2016), Xu et al. (2016). We have demonstrated dysregulations of ECM components expressed by fibroblasts are common in different cancers. We wonder whether these alterations within fibroblasts are spontaneous or induced by other cells. Ligand-receptor interactions have been used to infer intercellular communication Armingol et al. (2021), Lewis et al. (2021). And here we applied R package CellChat Jin et al. (2021) on 16 tumor and 14 non-tumor single-cell samples in which fibroblasts were identified to explore ligand-receptor interactions. We identified 3-63 significant ($P < 0.05$) pathways in each single-cell sample (86 pathways in total) (table S3). And receptors of 26 pathways were primarily expressed by fibroblasts in at least 10% single-cell samples (table S3, Fig. 2a). These 26 pathways might associate with dysregulation of cancer fibroblasts. Additionally, four pathways (NOTCH, VCAM, PARs, PERIOSTIN) had > 40% higher detect rates ($p < 0.05$), four pathways (CLDN, CD46, PROS, PDGF) had 30%-40% higher detect rates ($p < 0.15$) in tumor samples as compared with non-tumor samples (table S3, Fig. 2a).

To evaluate which pathways were the most possible to correlate with commonly up-regulated genes in cancer fibroblasts, and considering the low quantity of single-cell samples Cusnir et al. (2012), Cavalcante et al. (2012), Marusyk et al. (2010), Polyak et al. (2010) and the dropouts in single-cell data Kharchenko et al. (2014), Scadden et al. (2014), Peng et al. (2020), we applied R package WGCNA (Weighted Correlation Network Analysis) Langfelder et al.(2008), Horvath et al. (2008) on 3795 TCGA bulk tumor samples. This software predicts gene modules based on gene co-expressions which have been frequently used to infer gene functions Tan et al. (2019), Wolfe et al. (2005), Butte et al. (2005). We identified 73, 66, 53, 86, 43, 26, 42 gene modules in Breast, Colon, Liver, Lung, Prostate, Thyroid, Uterus cancer bulk samples, respectively (for each cancer type,

modules were named as Module1, Module2, etc., in descending order of number of genes, table S4).

Except for in Uterus cancers, we found modules in all the other six cancer types (Breast cancer Module4, Colon cancer Module2, Liver cancer Module4, Lung cancer Module7, Prostate cancer Module13, Thyroid cancer Module8, table S4) including commonly up-regulated Cluster2 genes (primarily COL10A1, COL1A1, COL1A2, COMP, POSTN, table S4). And we found in addition to POSTN previously analyzed (Fig. 1e), high expressions of COL10A1, COL1A1, COMP also produced unfavorable prognostic effects in most of the 7 cancer types (Fig. S2a), which have been reported by other studies Kahlert et al. (2022), Liu et al. (2018), Ma et al. (2022), Zhang et al. (2020), Wu et al. (2020), Zhang et al. (2018), Yang et al. (2018).

In the six WGCNA modules, we noticed ligands/receptors from the aforementioned 26 pathways. And the most frequent ones (in at least four of the six modules) were POSTN (PERIOSTIN ligand, it is itself the commonly up-regulated DEG), F2R (PARs receptor), PDGFRA (PDGF receptor), PDGFRB (PDGF receptor) which belong to the top 8 differentially identified pathways (Fig.2a) and FGF7 (FGF ligand), CDH11 (ANGPTL receptor) (table S4). It is recognized that PDGF signaling promotes both the proliferation and differentiation of fibroblasts into cancer-associated fibroblasts (CAFs) Kalluri et al. (2016), Ren et al. (2021). Zhang et al. (2022), which express high levels of ECM proteins such as collagens and fibronectin Frangogiannis et al. (2020) and contribute to the growth, expansion and dissemination of malignant cells Aboussekhra et al. (2011). The identification of co-expression of PDGF receptors and up-regulated ECM genes in cancer fibroblasts is consistent with these studies and confirms the reliability of our analysis.

We found the PARs signaling were triggered by NK, T, Mast or Malignant cells in cancer single cell samples (Fig.2b). Protease-activated receptors (PARs) are a subfamily of related G protein-coupled receptors activated by cleavage of part of their extracellular domain Macfarlane et al. (2001), Plevin et al. (2001), and have been found to function in cell polarization Goldstein et al. (2007), Macara et al. (2007), inflammatory Heuberger et al. (2019) & Schuepbach et al. (2019). We noticed the main PARs ligands in cancers were GZMA expressed by NK (nature killer) or T cells (which highly express CD3D, CD3G, CD3E and KLRB1, KLRD1, NKG7, suggesting their cytotoxicity) and CTSG primarily expressed by mast cells (Fig.2c, table S3). GzmA encoded by GZMA is a tryptase isolated from cytotoxic T lymphocyte (CTL) granules and cleaves proteins best after arginine Daniéle et al. (1986)& Jürg et alo. (1986). In CTL-targeted cells, GzmA activates caspase-independent programmed cell death pathways Martinvalet et al. (2008), Lieberman et al. (2008), Martinvalet et al. (2009), Lieberman et al. (2009), Zhu et al. (2009).

In colorectal cancer, GZMA has been found to promote cancer development by enhancing gut inflammation Santiago et al. (2020). Cathepsin G encoded by CTSG is a member of the serine proteases family, which was first found in azurophilic granules of neutrophil granulocytes Starkey et al. (1976) , Barrett et al. (1976), then found in other myeloid cells including B cells, monocytes, dendritic cells Burster et al. (2018), Mellins et al. (2010), Gao et al. (2018) , Luo et al. (2018) and mast cells Caughey et al. (2007). In tumors, inhibition of cathepsin G was found to reduce tumor vascularity Gao et al. (2018), Wilson et al. (2010), Singh et al. (2010). These findings indicate cytotoxic immune cells (NK or T cells) and mast cells in cancers produce effect on fibroblasts through PARs pathway, and this effect is probably specifically activated in tumor microenvironment (Fig.2a), for there are low levels of immune cells in healthy tissues.

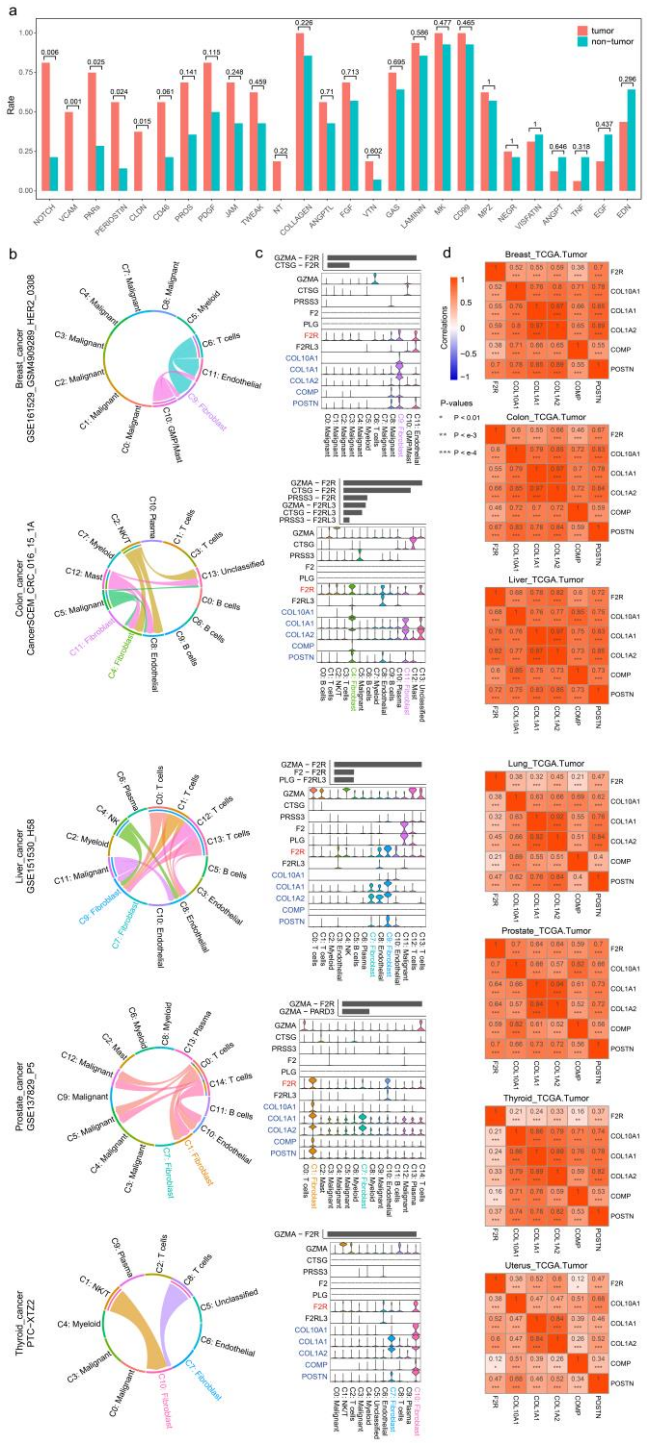
There are four members in PARs receptor family: PAR1 encoded by F2R, PAR2 by F2RL1, PAR3 by F2RL2 and PAR4 by F2RL3 Heuberger et al. (2019), Schuepbach et al. (2019). We found the primary PARs receptor expressed by cancer fibroblasts was F2R (Fig. 2c), which was in WGCNA modules of commonly up-regulated ECMs (COL10A1, COL1A1, COL1A2, COMP, POSTN) in four cancer types (Breast, Colon, Liver, and Prostate cancers, table S4). And although F2R were not in modules of these ECMs in the other three cancer types (Lung, Thyroid, Uterus cancers), the positive correlations between F2R and these ECMs were significant (Fig. 2d). Additionally, in single-cell samples, fibroblast clusters which express F2R and which express these ECMs were the same (Fig. 2c). These findings at bulk and single-cell levels prove the correlation between the F2R and commonly up-regulated ECM proteins, which suggest a potential mechanism to regulate fibroblasts through PARs in cancers.

We also examined correlations between F2R and COL10A1, COL1A1, COL1A2, COMP, POSTN in GTEx non-tumor and TCGA adjacent samples. We detected low correlations in non-tumor samples but moderate correlations in adjacent samples (Fig. S2b). We speculate this phenomenon might be caused by infiltration of immune cells in adjacent tissues which provided PARs ligands (GzmA or Cathepsin G). And we did observe a higher expression level of GZMA in adjacent tissues as compared with non-tumor tissues (Fig. S2c). This result indicates the gene networks in adjacent tissues are different from normal tissues and cannot be considered healthy.

POSTN is commonly up-regulated DEG (Fig.1c) and the only ligand of the PERIOSTIN pathway (Gillan et al., 2002), whose receptors are ITGAV+ITGB3 and ITGAV+ITGB5. Researchers have found the activation of ITGAV+ITGB5 on fibroblasts helps them to acquire

a myofibroblast phenotype (highly expressing alpha-smooth muscle actin encoded by ACTA2) Franco-Barraza et al. (2017), which in cancers is recognized as activated fibroblast and a major source of the CAFs Schmitt-Gräff et al. (1994), Gabbiani et al. (1994), Shiga et al. (2015), Xing et al. (2010), Watabe et al. (2010). We validated fibroblasts in cancers to express ligand POSTN and receptor ITGAV+ITGB5, simultaneously (table S3), and the PERIOSTIN pathway was more frequently activated in tumors (Fig. 2a).

Figure 2: Pathways identified in single-cell samples and PARs pathway in tumors.



a: Identification rates of pathways in tumor and non-tumor single-cell samples. The x-axis is pathways in descending order of rate difference between tumor and non-tumor samples. Chi-squared tests were performed on pathway frequencies between tumor and non-tumor samples, and p values were markerd on each pair of histograms. b: PARs signaling detected in five tumor single-cell samples by CellChat. GMP: granulocyte-monocyte progenitor. c: Upper-panel, relative contributions of ligand-receptor pairs to PARs signaling in corresponding samples. Lower-panel, expressions of PARs ligands and receptors and five correlated commonly up regulated genes in corresponding samples. d: Gene expression correlations of PARs receptor F2R and commonly up regulated genes in TCGA bulk tumor samples from 7 cancer types. Numbers in each cell are correlations, * labeled under the numbers are significance.

These findings indicate a self-activation mechanism of CAFs through the PERIOSTIN pathway in tumors. Previous study found activation of PAR1 (F2R) and PAR2 (F2RL1) promoted alpha-smooth muscle actin (ACTA2) expression in human lung fibroblasts Asokananthan et al. (2015). And in cancers, fibroblasts highly expressing ACTA2 are considered CAFs. Our findings suggest a possible underlying mechanism through PARs to PERIOSTIN which activate CAFs in cancers. And we propose that NK/T or mast cells which provide PARs ligands are the source of this signaling.

Immune activation is correlated with favorable prognosis in pan-cancer

Gene expression levels has been used to predict cancer patients' clinical outcomes Vijver et al. (2002). These genes are called prognostic genes which are potential therapy targets Mao et al. (2021) and may associate with cancer progression Tzanakis et al. (2006). In cells, prognostic genes may interact with other genes to form complex networks including interaction networks, regulatory networks, co-expression networks, signaling networks and metabolic networks. Among these networks, gene co-expression network has been used to investigate properties of prognostic genes in cancers for the following advantages: high coverage of the genome, little bias, and the ability to construct cancer-specific networks Yang et al. (2014).

Herein, to find out gene modules that predict patients' survival in pan-cancer, we summarized gene modules constructed by WGCNA and found 58 common modules, genes from which were co-expressed in all of the 7 cancer types. We named them as S01-58 in decreasing order of number of genes and categorize them into 13 classes according to functional enrichment and PPI networks (table S5). We applied the GSVA algorithm Hanzelmann et al. (2013), Subramanian et al. (2005) to evaluate expression levels of these gene sets in cancer tissues.

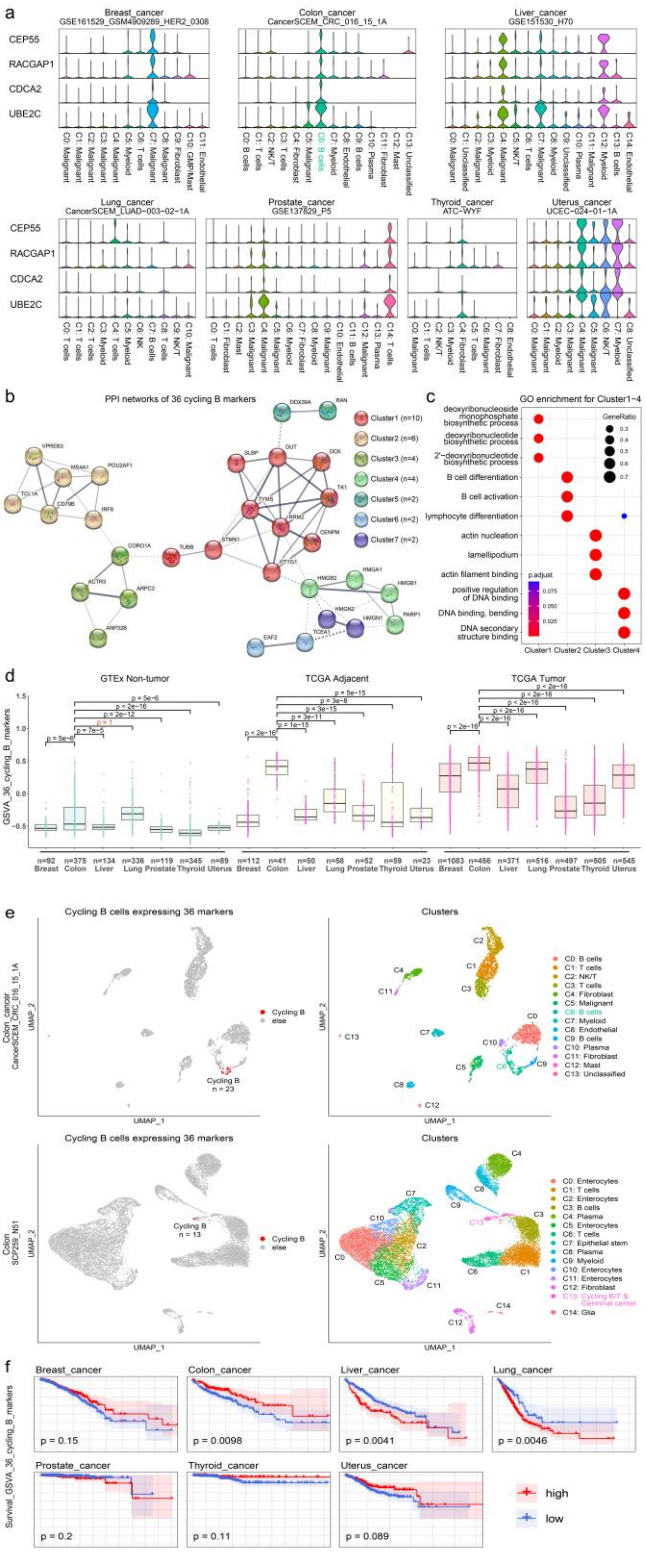
This algorithm ranks genes in order of expression level in each sample and then gives scores of gene sets based on the cumulative density function (CDF). Then we applied survival analysis on GSVA scores and found high expressions of immune activation related sets ('T cell activation' gene sets S01, S08, 'B cell activation' gene sets S09, S14, 'antigen processing MHCI' gene set S18 and 'lymphocyte differentiation' gene set S45) were associated with favorable prognosis in all the 7 cancer types (Fig. S3). This finding suggests the immune activity is a broad indicator for pan-cancer survival. And the co-expressions of these genes in all the 7 cancer types indicate immune activation mechanisms in different cancers are similar.

Among the 58 commonly co-expressed gene sets, we found 28 sets were not enriched on known functions. And genes from these sets were primarily expressed by malignant cells (table S5). This finding indicates unique gene networks in tumor cells, which do not perform known biology functions.

Cycling B cells are Colon specific and associate with cancer prognosis

Cancer tissue is occupied by malignant cells that proliferate out of control. And we found strong signals of cell division related gene sets ('cell division' gene sets S04, S10, S37 and 'mitotic nuclear division and protein polymerization' gene set S58) in cancer tissues (table S5, Fig. S4a). We noticed high expression levels of these gene sets were unfavorable to prognosis in Breast, Liver, Lung, Prostate, Thyroid, Uterus cancers but favorable to prognosis in Colon cancers (Fig. S4b). We checked 4 typical genes (CEP55, RACGAP1, CDCA2, UBE2C, according to gene functions and their positions in PPI networks, Fig. S4c) of these gene sets, and found similar prognostic properties (Fig. S4d). This phenomenon was also observed by other studies Uhlen et al. (2017). We speculated these genes might be expressed by different cell types in Colon cancers, and examined their expressions in single-cell samples. We found these genes were highly expressed by malignant cells, myeloid cells, fibroblasts or T cells in samples of other cancer types, but were expressed by B cells only in Colon cancers (Fig. 3a). In three Colon non-tumor single-cell samples, we noticed clusters annotated as cycling B cells Smillie et al. (2019) expressing these genes, as well. And we wondered whether the aforementioned B cells in Colon cancers were cycling B cells in these non-tumor single-cell samples. We summarized marker genes of Colon cancer B cell clusters and Colon non-tumor cycling B cell clusters identified by Seurat Hao et al. (2021), Satija et al. (2015) (expressed in at least 25% cells, log2-fold change > 1, adjusted P-value < 0.05) in two tumor and three non-tumor high-quality single-cell samples. We identified 182, 132 marker genes in two cancer samples, respectively; 93, 364, 181 marker genes in three non-tumor samples, respectively. We found 36 marker genes shared by all the five samples (table S6). PPI networks and GO enrichment revealed two main functional groups in the 36 markers which control deoxyribonucleoside biosynthesis (10 genes identified as Cluster1 by PPI, these genes are expressed by proliferating cells, Fig. 3b, c) and B cell differentiation (six genes identified as Cluster2, Fig. 3b, c), suggesting these cells were cycling B cells.

Figure 3: Cycling B cells in Colon tissues and cancers.



a: Expressions of CEP55, RACGAP1, CDCA2, UBE2C in cancer single-cell samples. b: PPI networks of the 36 cycling B cell markers, 7 clusters are identified by MCL clustering with inflation parameter = 2.5. Line thickness indicates the strength of data support. c: GO enrichment of Cluster1-4 identified by PPI d: GSVA scores of the 36 cycling B cell markers in GTEx non-tumor, TCGA adjacent and TCGA tumor samples. Scores in Colon tissues are significantly higher than samples from the other organs (except for Lung non-tumor samples).

P-values are calculated by one-tailed Wilcoxon test (null hypothesis: colon ≤ other organs). e: Visualizations of cells which simultaneously express above average levels of the 36 cycling B cell markers in colon tumor (CancerSCEM_CRC_016_15_1A, upper panel) and non-tumor (SCP259_N51, lower panel) single-cell samples (on the left). They belong to clusters C6: B cells and C13: Cycling B/T & Germinal center, respectively (on the right). f: Survival analysis on GSVA scores of the 36 cycling B cell markers in 7 cancer types.

Higher scores are significantly associated with better prognosis in Colon cancers.

To further verify whether these cycling B cells are Colon specific, we used GSVA to calculate scores of the 36 marker genes in tumor, adjacent and non-tumor bulk samples, and found significant higher levels in Colon tissues than tissues from other organs (one-tailed Wilcoxon test, $p < 0.0001$), except for Lung non-tumor tissues which expressed higher levels than Colon non-tumor tissues (Fig. 3d). We speculated the high GSVA scores observed in Lung non-tumor tissues might be contributed by two or more cell types rather than cycling B cells, and in order to verify this conjecture, we sought for cells simultaneously expressing the 36 marker genes at above average levels in single-cell samples from different organs and tissues. We identified 23 and 8 cells in the two Colon cancer samples, respectively; 4, 9 and 13 cells in the three Colon non-tumor samples, respectively. We found these cells were all B cells previously annotated.

We identified just one cell in a Lung cancer sample (which might be doublet that is an artifactual library generated from two cells), and none in the rest of the single-cell samples (Fig. 3e). This finding suggests these cycling B cells are Colon specific.

We performed survival analysis using GSVA scores of the 36 cycling B cell markers and found higher expression levels of these genes provide significantly better prognosis in Colon cancers but not in the other cancer types (Fig. 3f), which were similar to the cell cycle related genes (Fig. S4b, d). These findings indicate cycling B cells might be important to Colon cancer progression and survival though their amount is limited (Fig. 3e).

DISCUSSION

Cancer heterogeneity limits the efficiency of cancer studies with small sample sizes. While conducting in-depth and detailed studies with large sample sizes is expensive. Our findings demonstrate systematic combination of publicly available bulk and single-cell "big data" resources an effective approach to dissect the cancer microenvironment. The data shows that fibroblasts were dysregulated in most cancer types and associated with immune cells.

Despite considerable advances in the development of targeted therapies, no significant improvements have been made in the overall survival of patients with malignant tumors. One factor is that these therapies primarily target the fast-growing tumor but largely ignore the tumor microenvironment. Groot et al. (2017), Amend et al. (2017), Liu et al. (2019), Zhang et al. (2019) Tumor microenvironment includes ECM and the surrounding stromal cells such as immune cells and fibroblasts. CAFs are a major component of the cancer stroma Cirri et al. (2011), Chiarugi et al. (2011), Shiga et al. (2015), and secret the majority of ECM Kendall et al. (2014) which contribute to the growth, expansion and dissemination of malignant cells Aboussekhra et al. (2011). Our findings demonstrate most of the dysregulated genes expressed by fibroblasts were ECM components, such as COL10A1, COL1A1, COL1A2, COMP and POSTN. In oesophageal adenocarcinomas, CAFs release Periostin (POSTN) and promote tumor cell growth through paracrine signaling. Underwood et al. (2015) We observed the over expression of POSTN in pan-cancer and found its high expressions were identically associated with unfavorable prognosis in pan-cancer. We also propose POSTN may induce self-activation of CAFs. Current cancer drugs are effective only in a subgroup of cancers because of the heterogeneity of different cancer types Brennan et al. (2010), Gallagher et al. (2010). And these commonly dysregulated genes provide basis for more broader therapeutic approach.

CAFs can break down proteins in the ECM leading to disruption of the normal structure allowing cancer cells to escape from their primary region. MMP proteins are key to this process. Shiga et al. (2015) We observed many MMPs over expressed by fibroblasts in cancers except for MMP-7. This MMP has been reported to be expressed by malignant cells in Pancreas Crawford et al. (2000), Leach et al. (2002) and Gastro/Esophagus Adachi et al. (1998) cancers, and we observed its expressions by malignant cells in Breast, Colon, Lung, Prostate, Uterus cancers (Fig. 1c, table S2), which indicates direct regulations to ECM by malignant cells in pan-cancer.

Significantly higher plasma MMP-7 levels and serum MMP-7 levels were detected in Pancreas cancers and Colon cancers, respectively Liao et al. (2021), Zhang et al. (2021), which suggest the possibility to predict or diagnose cancers in non-invasive manners.

We found pathways targeting fibroblasts in pan-cancer single-cell samples. Among these pathways, PARs, PERIOSTIN, PDGF were frequently activated in tumors and associated with dysregulated ECM genes which were correlated with worse survival (Fig. 1e, Fig. S2a). Especially, we validated the correlation between PARs pathway and these ECM DEGs at both bulk and single-cell levels (Fig. 2c, d). And we first report the ligands of PARs are primarily provided by cytotoxic immune cells (NK or T cells) in most cancers (Fig. 2b, c). On the other hand, we observed the activation of immune related genes improved cancer patients' survival (Fig. S3). These findings illustrate the multiple effects of the immune activity in cancers. Immunotherapy has been proved an effective strategy to cure cancers Riley et al. (2019), Mitchell et al. (2019). One classic case of immunotherapy is checkpoint inhibitors targeting programmed cell death or cytotoxic lymphocyte associated proteins, while only a subset of patients responds to these inhibitors, and a substantial proportion of initial responders ultimately relapse with lethal, drug-resistant disease months or years later Syn et al. (2017), Soo et al. (2017). Our findings reveal factors that might lead to the failure of cytotoxic lymphocyte associated protein inhibitors.

Pro-inflammatory immediate-early response genes have been found to be activated in tumor adjacent tissues Dvir Aran et al. (2017). Though the batch effect should be considered, we observed higher immune activity signatures in adjacent tissues (Fig. S2c), together with different gene co-expression networks (Fig. S2b). And gene co-expressions are not affected by batch effects, because the correlations were calculated within a single batch. These findings illustrate that adjacent tissues are influenced by tumors more or less. And this effect should be considered in cancer studies using adjacent tissues as control.

Opposite prognostic effects of genes have been observed in different cancer types, but the underlying mechanism remains poorly understood Uhlen et al. (2017). By identifying cycling B cells in Colon cancers, we demonstrate gene prognostic effects are associated with specific cell types. And we provide 36 marker genes for Colon cancer cycling B cells. Germinal center (GC) is a transiently formed structure (persisting for weeks to months) in lymph nodes or the spleen. In GCs, B cells are activated, proliferate, differentiate, and mutate their antibody genes during normal immune response. Chronic GCs of longer duration are found in intestinal Peyer's patches, with B cells in these sites undergoing antibody selection in response to persistent exposure to gut microbiota Chen et al. (2020), Nowosad et al. (2020), Young et al. (2021), Brink et al. (2021). The longer duration may make GC and cycling B cells detectable in Colon samples. And in non-tumor Colon single-cell samples, researchers identified cycling B cells around GC cells after dimensionality reduction Smillie et al. (2019) (Fig. 3e), which suggest that the origin of cycling B cells is GC. Even if the ratio of cycling B cells were limited in Colon samples (Fig. 3e), the high expression levels of their marker genes were significantly correlated with better survival in Colon cancers (Fig. 3f), and their existing may reverse the prognostic effects of many other cell cycle related genes in Colon cancers (Fig. S4). These findings all suggest the specific microenvironment in Colon cancers, and indicate that cycling B cells are important in Colon cancer progression.

CONCLUSION

In cancers, fibroblast cells express numerous dysregulated genes, and are associated with patients' overall survival. GzmA expressed by NK or T cells is significantly correlated with the dysregulation of fibroblasts in cancers. There is a higher amount of cycling B cells in Colon cancers, which is correlated with Colon cancers' clinical outcomes.

DECLARATIONS

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of interest statement

The authors declare no competing interests.

Data availability statement

TCGA cancer and GTEx non-tumor bulk expression data are from GSE62944 Rahman et al. (2015) and GSE86354 Dvir Aran et al. (2017), respectively. Researchers can download tumor single-cell data of Breast cancer from GSE161529 Pal et al. (2021), of Liver cancer from GSE151530 Ma et al. (2021), of Prostate cancer from GSE137829 Dong et al. (2020), of Thyroid cancer from GSE210347 Luo et al. (2022), Luo et al. (2021), of Colon cancer, Lung cancer, Uterus cancer from CancerSCEM (https://ngdc.cncb.ac.cn/cancerscem) Zeng et al. (2022), non-tumor tissue single-cell data of Breast from Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) SCP1731 Gray et al. (2022) (tumor-free individuals), of Colon from Single Cell Portal SCP259 Smillie et al. (2019) (tumor-free individuals), of Liver from GSE115469 MacParland et al. (2018) (tumor-free individuals), of Lung from ERP136992 Madissoon et al. (2023) (tumor-free individuals),

of Prostate from GSE120716 Henry et al. (2018) (tumor-free individuals) and from Prostate Cell Atlas (https://www.prostatecellatlas.org/) Tuong et al. (2021) (cancer patients), of Uterus from E-MTAB-10287 Garcia-Alonso et al. (2021) (tumor-free individuals) and from GSE111976 Vilella et al. (2021), Wang et al. (2020) (tumor-free individuals).

Acknowledgements

We thank Kun Ma and Siqi Liu for their academic guidance. We thank Longqi Liu, Shiping Liu, Liang Wu, Ying Lei, Zhenkun Zhuang for providing technical support. This research is supported by China National GeneBank.

Author contribution

All authors contributed to the study. Specifically, Yan Sun collected the data, wrote the main manuscript text and prepared all the figures and supplementary material. Bin Song and Qichao Yu performed differentially expressed gene (DEG) analysis and polished the manuscript. Huanming Yang contributed to the final version of the manuscript. Wei Dong supervised the project. All authors reviewed the manuscript.

Ethics declarations

This is an observational study based entirely on publicly available data. The Institute of Review Board of Bioethics and Biosafety (BGI-IRB) has confirmed that no ethical approval is required.

REFERENCES

1.Aboussekhra A. 2011. Role of cancer-associated fibroblasts in breast cancer development and prognosis. Int J Dev Biol. 55(7-8-9): 841-849.

2.Adachi Y, Itoh F, Yamamoto H, et al. 1998 Nov. Matrix metalloproteinase matrilysin (MMP-7) participates in the progression of human gastric and esophageal cancers. Int J Oncol. 13(5): 1031-1035.

3.Aran D, Camarda R, Odegaard J, et al. 2017 Oct 20. Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat Commun. 8(1): 1077.

4.Aran D, Hu Z, Butte AJ. 2017 Nov 15. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 18(1): 220.

5.Armingol E, Officer A, Harismendy O. 2021 Feb. Deciphering cell-cell interactions and communication from gene expression. Nat Rev Genet. 22(2):71-88.

6Arora K, Barbieri CE. 2018 Jun 1. Molecular subtypes of prostate cancer. Curr Oncol Rep. 20(8):58.

7.Asokananthan N, Lan RS, Graham PT, et al. 2015 Feb 6. Activation of protease-activated receptors (PARs)-1 and -2 promotes alpha-smooth muscle actin expression and release of cytokines from human lung fibroblasts. Physiol Rep. 3(2):e12295.

8.Basu AK. 2018 Mar 23. DNA damage, mutagenesis and cancer. Int J Mol Sci. 19(4):970.

9.Binnewies M, Roberts EW, Kersten K, et al. 2018 May. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 24(5): 541-550.

10.Brassart-Pasco S, Brézillon S, Brassart B, et al. 2020 Apr 15. Tumor microenvironment: extracellular matrix alterations influence tumor progression. Front Oncol. 10:397.

11.Brennan DJ, O'Connor DP, Rexhepaj E, et al. 2022 Jun. Antibody-based proteomics: fast-tracking molecular diagnostics in oncology. Nat Rev Cancer. 22(6):373.

12.Burster T, Macmillan H, Hou T, et al. 2010 Jan. Cathepsin G: roles in antigen presentation and beyond. Mol Immunol. 47(4):658-65.

13.Carlson M, Falcon S, Pages H, et al. 2019. org. Hs. eg. db: Genome wide annotation for Human. R package version. 3(2):3.

14.Caughey GH. 2007 Jun. Mast cell tryptases and chymases in inflammation and host defense. Immunol Rev. 217:141-54.

15.Chen H, Zhang Y, Ye A Y, et al. 2020 Jun. BCR selection and affinity maturation in Peyer's patch germinal centres. Nature. 582(7812):421-425.

16.Chen Z, Yu M, Yan J, et al. 2021 Mar 24. PNOC Expressed by B Cells in Cholangiocarcinoma Was Survival Related and LAIR2 Could Be a T Cell Exhaustion Biomarker in Tumor Microenvironment: Characterization of Immune Microenvironment Combining Single-Cell and Bulk Sequencing Technology. Front Immunol. 12:647209.

17.Choi H, Moon A. 2018 Jul. Crosstalk between cancer cells and endothelial cells: implications for tumor progression and intervention. Arch Pharm Res. 41(7):711-724.

18.Cirri P, Chiarugi P. 2011. Cancer associated fibroblasts: the dark side of the coin. Am J Cancer Res. 1(4):482-97.

19.Crawford HC, Scoggins CR, Washington MK, et al. 2002 Jun. Matrix metalloproteinase-7 is expressed by pancreatic cancer precursors and regulates acinar-to-ductal metaplasia in exocrine pancreas. J Clin Invest. 109(11):1437-44.

20.Croce CM. 2008 Jan 31. Oncogenes and cancer. N Engl J Med. 358(5):502-11.

21.Cusnir M, Cavalcante L. 2012 Aug. Inter-tumor heterogeneity. Hum Vaccin Immunother. 8(8):1143-5.

22.Cusnir M, Cavalcante L. 2012 Aug. Inter-tumor heterogeneity. Hum Vaccin Immunother. 8(8):1143-5.

23.Daniéle M, Moreno Z, Jürg T. 1986 Nov 10. Identification of granzyme A isolated from cytotoxic T-lymphocyte-granules as one of the proteases encoded by CTL-specific genes. FEBS Lett. 208(1):84-8.

24.Dominiak A, Chełstowska B, Olejarz W, et al. 2020 May 14. Cancers (Basel). 12(5):1232.

25.Dong B, Miao J, Wang Y, et al. 2020 Dec 16. Single-cell analysis supports a luminal-neuroendocrine transdifferentiation in human prostate cancer. Commun Biol. 3(1):778.

26.Eisenberg E, Levanon EY. 2013 Oct. Human housekeeping genes, revisited. Trends Genet. 29(10):569-74.

27.Fontana E, Eason K, Cervantes A, et al 2019 Apr 1. Context matters-consensus molecular subtypes of colorectal cancer as biomarkers for clinical trials. Ann Oncol. 30(4):520-527.

28.Franco-Barraza J, Francescone R, Luong T, et al. 2017 Jan 31. Matrix-regulated integrin α(v)β (5) maintains α (5)β(1)-dependent desmoplastic traits prognostic of neoplastic recurrence. Elife. 6:e20600.

29.Frangogiannis N. 2020 Feb 13. Transforming growth factor-β in tissue fibrosis. J Exp Med. 217(3):e20190103.

30.Gao S, Zhu H, Zuo X, et al. 2018 Jan 22. Cathepsin G and its role in inflammation and autoimmune diseases. Arch Rheumatol. 33(4):498-504.

31.Garcia-Alonso L, Handfield L-F, Roberts K, et al. 2021 Dec. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. Nat Genet. 53(12): 1698-1711.

32.Garner H, Visser KEd. 2020 Aug. Immune crosstalk in cancer progression and metastatic spread: a complex conversation. Nat Rev Immunol . 20(8):483-497.

33.Gillan L, Matei D, Fishman DA., Gerbin, et al. 2002 Sep 15. Periostin secreted by epithelial ovarian carcinoma is a ligand for αVβ3 and αVβ5 integrins and promotes cell motility. Cancer Res. 62(18):5358-64.

34.Goldstein B, Macara IG. 2007 Nov. The PAR proteins: fundamental players in animal cell polarization. Dev Cell. 13(5):609-622.

35.González-González L, Alonso J. 2018 Jun 12. Periostin: a matricellular protein with multiple functions in cancer development and progression. Front Oncol. 8:225.

36.Gray GK, Li C M-C, Rosenbluth J, et al. 2022 Jun 6. A human breast atlas integrating single-cell proteomics and transcriptomics. Dev Cell. 57(11):1400-1420.e7.

37.Groot AEd, Roy S, Brown JS, et al. 2017 Apr. Revisiting Seed and Soil: Examining the Primary Tumor and Cancer Cell Foraging in Metastasis. Mol Cancer Res. 15(4):361-370.

38.Hanzelmann S, Castelo R, Guinney J. 2013 Jan 16. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 14:7.

39.Hao Y, Hao S, Andersen-Nissen E, et al. 2021 Jun 24. Integrated analysis of multimodal single-cell data. Cell. 184(13):3573-3587.

40.Henry GH, Malewska A, Joseph DB, et al. 2018 Dec 18. A cellular anatomy of the normal adult human prostate and prostatic urethra. Cell Rep. 25(12):3530-3542.e5.

41.Heuberger DM, Schuepbach RA. 2019 Mar 29. Protease-activated receptors (PARs): mechanisms of action and potential therapeutic modulators in PAR-driven inflammatory diseases. Thromb J. 17:4.

42.Hu X, Villodre ES, Larson R, et al. 2021 Jan 15. Decorin-mediated suppression of tumorigenesis, invasion, and metastasis in inflammatory breast cancer. Commun Biol. 4(1):72.

43.Hufton SE, Moerkerk PT, Brandwijk R, et al. 1999 Dec 10. A profile of differentially expressed genes in primary colorectal cancer using suppression subtractive hybridization. FEBS Lett. 463(1-2):77-82.

44.Huvila J, Pors J, Thompson EF, et al. 2021 Apr. Endometrial carcinoma: molecular subtypes, precursors and the role of pathology in early diagnosis. J Pathol. 253(4):355-365.

45.Järvinen TAH, Prince S. 2015. Decorin: a growth factor antagonist for tumor growth inhibition. Biomed Res Int. 2015:654765.

46.Jin S, Guerrero-Juarez CF, Zhang L, et al. 2021 Feb 17. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 12(1):1088.

47.Kahlert UD, Shi W, Strecker M, et al. 2022 Oct 4. COL10A1 allows stratification of invasiveness of colon cancer and associates to extracellular matrix and immune cell enrichment in the tumor parenchyma. Front Oncol. 12:1007514.

48.Kalluri R. 2016 Aug 23. The biology and function of fibroblasts in cancer. Nat Rev Cancer. 16(9):582-98.

49.Kassambara A, Kosinski M, Biecek P, et al. 2017. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.3. 1.

50.Kendall RT, Feghali-Bostwick CA. 2014 May 27. Fibroblasts in fibrosis: novel roles and mediators. Front Pharmacol. 5:123.

51.Kettunen E, Anttila S, Seppänen JK, et al. 2004 Mar. Differentially expressed genes in non small cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. Cancer Genet Cytogenet. 149(2):98-106.

52.Kharchenko PV, Silberstein L, Scadden DT. 2014 Jul. Bayesian approach to single-cell differential expression analysis. Nat Methods. 11(7):740-742.

53.Kim HK, Park KH, Kim Y, et al. 2019 Apr. Discordance of the PAM50 intrinsic subtypes compared with immunohistochemistry-based surrogate in breast cancer patients: potential implication of genomic alterations of discordance. Cancer Res Treat. 51(2):737-747.

54.Kumar V, Ramnarayanan K, Sundar R, et al. 2022 Mar 1. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. Cancer Discov. 12(3):670-691.

55.Langfelder P, Horvath S. 2008 Dec 29. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 9:559.

56.Lawson DA, Kessenbrock K, Davis RT, et al. 2018 Dec. Tumour heterogeneity and metastasis at single-cell resolution. Nat Cell Biol. 20(12):1349-1360.

57.Leun AM vd, Thommen DS, Schumacher TN, 2020 Apr. CD8+ T cell states in human cancer: insights from single-cell analysis. Nat Rev Cancer. 20(4):218-232.

58.Li Y-R, Yang W-X. 2016 Jul 19. Myosins as fundamental components during tumorigenesis: diverse and indispensable. Oncotarget. 7(29):46785-46812.

59.Liao H-Y, Da C-M, Liao B, et al. 2021 Jun. Roles of matrix metalloproteinase-7 (MMP-7) in cancer. Clin Biochem. 92:9-18.

60.Liu J, Shen J-X, Wu H-T, et al. 2018 May. Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target. Discov Med. 25(139):211-223.

61.Liu T, Zhou L, Li D, et al. 2019 Apr 24. Cancer-Associated Fibroblasts Build and Secure the Tumor Microenvironment. Front Cell Dev Biol. 7:60.

62.Lonsdale J, Thomas J, Salvatore M, et al. 2013 Jun. The genotype-tissue expression (GTEx) project. Nat Genet. 45(6):580-585.

63.Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550.

64.Luo H, Xia X, Huang L-B, et al. 2022 Nov 4. Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment. Nat Commun. 13(1):6619.

65.Luo H, Xia X, Kim GD, et al. 2021 Jul 28. Characterizing dedifferentiation of thyroid cancer by integrated analysis. Sci Adv. 7(31):eabf3657.

66.Lv M, Luo L, Chen X. 2022 Feb. The landscape of prognostic and immunological role of myosin light chain 9 (MYL9) in human tumors. Immun Inflamm Dis. 10(2):241-254.

67.Ma H, Qiu Q, Tan D, et al. 2022 Dec 28. The Cancer-Associated Fibroblasts-Related Gene COMP Is a Novel Predictor for Prognosis and Immunotherapy Efficacy and Is Correlated with M2 Macrophage Infiltration in Colon Cancer. Biomolecules. 13(1):62

68.Ma L, Wang L, Khatib SA, et al. 2021 Dec. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. J Hepatol. 75(6):1397-1408

69.Macfarlane SR, Seatter MJ, Kanke T, et al. 2001 Jun. Proteinase-activated receptors. Pharmacol Rev. 53(2):245-82.

70.MacParland SA, Liu JC, Ma X-Z, et al. 2018 Oct 22. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat Commun. 9(1):4383.

71.Madissoon, E, Oliver AJ, Kleshchevnikov V, et al. 2023 Jan. A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. Nat Genet. 55(1):66-77.

72.Malikic S, Jahn K, Kuipers J, et al. 2019 Jun 21. Integrative inference of subclonal tumour evolution from single-cell and bulk sequencing data. Nat Commun. 10(1):2750.

73.Mao W, Wang K, Xu B, et al. 2021 Nov 5. ciRS-7 is a prognostic biomarker and potential gene therapy target for renal cell carcinoma. Mol Cancer. 20(1):142.

74.Marisa L, Reyniès Ad, Duval A, et al. 2013. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 10(5):e1001453.

75.Martinvalet D, Dykxhoorn DM, Ferrini R, et al. 2008 May 16. Granzyme A cleaves a mitochondrial complex I protein to initiate caspase-independent cell death. Cell. 133(4):681-92.

76.Martinvalet D, Walch M, Jensen D K, et al. 2009. Response: Granzyme A: cell death–inducing protease, proinflammatory agent, or both? Blood. 114(18):3969-3970.

77.Marusyk A, Polyak K. 2010 Jan. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 1805(1):105-17.

78.McInnes L, Healy J, Melville J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv.

79.Nagy Á, Munkácsy G, Győrffy B. 2021 Mar 15. Pancancer survival analysis of cancer hallmark genes. Sci Rep. 11(1):6047.

80.Network CGAR. 2015 Nov 5. The molecular taxonomy of primary prostate cancer. Cell. 163(4):1011-25.

81.Network CGAR, Weinstein JN, Collisson EA, et al. 2013 Oct. The cancer genome atlas pan-cancer analysis project. Nat Genet. 45(10):1113-20.

82.Nie M-J, Pan X-T, Tao H-Y, et al. 2020 Jun. Clinical and prognostic significance of MYH11 in lung cancer. Oncol Lett. 19(6):3899-3906.

83.Nowosad CR, Mesin L, Castro TBR, et al. 2020 Dec. Tunable dynamics of B cell selection in gut germinal centres. Nature. 588(7837):321-326.

84.Ouderkirk JL, Krendel M. 2014 Aug. Non-muscle myosins in tumor progression, cancer cell invasion, and metastasis. Cytoskeleton (Hoboken). 71(8):447-63.

85.Pal B, Chen Y, Vaillant F, et al. 2021 Jun 1. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. EMBO J. 40(11):e107333.

86.Parker JS, Mullins M, Cheang MCU, et al. 2009 Mar 10. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 27(8):1160-7

87.Peng Q. 2020 Mar 3. Embracing the dropouts in single-cell RNA-seq analysis. Nat Commun. 11(1):1169

88.Prete A, Souza PBd, Censi S, et al. 2020 Mar 13. Update on fundamental mechanisms of thyroid cancer. Front Endocrinol (Lausanne). 11:102

89.Rahman M, Jackson LK, Johnson WE, et al. 2015 Nov 15. Alternative preprocessing of RNA-Sequencing data in The Cancer Genome Atlas leads to improved analysis results. Bioinformatics. 31(22):3666-72.

90.Ren X, Li L, Wu J, et al. 2021 Jul. PDGF-BB regulates the transformation of fibroblasts into cancer-associated fibroblasts via the lncRNA LURAP1L-AS1/LURAP1L/IKK/IκB/NF-κB signaling pathway. Oncol Lett. 22(1):537.

91.Riley RS, June CH, Langer R, et al. 2019 Mar. Delivery technologies for cancer immunotherapy. Nat Rev Drug Discov. 18(3):175-196

92.Risso D, Ngai J, Speed TP, et al. 2014 Sep. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 32(9):896-902.

93.Robinson MD, McCarthy DJ, Smyth GK. 2010 Jan 1. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26(1):139-40.

94.Rosario SR, Long MD, Affronti HC, et al. 2018 Dec 14. Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. Nat Commun. 9(1):5330.

95.Rudin CM, Poirier JT, Byers LA, et al. 2019 May. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat Rev Cancer. 19(5):289-297.

96.Santiago L, Castro M, Sanz-Pamplona R, et al. 2020 Jul 7. Extracellular granzyme A promotes colorectal cancer development by enhancing gut inflammation. Cell Rep. 32(1):107847.

97.Satija R, Farrell JA, Gennert D, et al. 2015 May. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 33(5):495-502.

98.Schmitt-Gräff A, Desmoulière A, Gabbiani G. 1994. Heterogeneity of myofibroblast phenotypic features: an example of fibroblastic cell plasticity. Virchows Arch. 425(1):3-24.

99.Schwager SC, Taufalele PV, Reinhart-King CA. 2019 Feb. Cell-Cell Mechanical Communication in Cancer. Cell Mol Bioeng. 12(1):1-14.

100.Shiga K, Hara M, Nagasaki T, et al. 2015 Dec 11. Cancer-Associated Fibroblasts: Their Characteristics and Their Roles in Tumor Growth. Cancers (Basel). 7(4):2443-58.

101.Sia D, Villanueva A, Friedman S. L, et al. 2017 Mar. Liver cancer cell of origin, molecular class, and effects on patient prognosis. Gastroenterology. 152(4):745-761.

102.Skibinski A, Kuperwasser C. 2015 Oct 16. The origin of breast tumor heterogeneity. Oncogene. 34(42):5309-16.

103.Smillie C. S, Biton M, Ordovas-Montanes J, et al. 2019 Jul 25. Intra-and inter-cellular rewiring of the human colon during ulcerative colitis. Cell. 714-730.e22.

104.Snel B, Lehmann G, Bork P, et al. 2000 Sep 15. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 28(18):3442-4.

105.Starkey P. M,Barrett A. J. 1976 May 1. Human cathepsin G. Catalytic and immunological properties. Biochem J. 155(2):273-8.

106.Stetler-Stevenson W. G, Liotta L. A, Kleiner DE Jr. 1993 Dec. Extracellular matrix 6: role of matrix metalloproteinases in tumor invasion and metastasis. FASEB J. 7(15):1434-41.

107.Subramanian A, Tamayo P, Mootha V. K, et al. 2005 Oct 25. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 102(43):15545-50.

108.Syn NL, Teng MWL, Mok TSK, et al. 2017 Dec. De-novo and acquired resistance to immune checkpoint targeting. Lancet Oncol. 18(12): e731-e741.

109.Szklarczyk D, Gable AL, Lyon D, et al. 2019 Jan 8. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47(D1): D607-D613.

110.Tan H, Huang S, Zhang Z, et al. 2019 May. Pan-cancer analysis on microRNA-associated gene activation. EBioMedicine. 43:82-97.

111.Tan X, Chen M. 2014 Dec. MYLK and MYL9 expression in non-small cell lung cancer identified by bioinformatics analysis of public expression data. Tumour Biol. 35(12):12189-200.

112.Team RC. (2013). R: A language and environment for statistical computing.

113.Therneau TM, Lumley T. (2015). Package 'survival'. R Top Doc. 128(10): 28-33.

114.Therneau T. M, Lumley T, Elizabeth A, et al. (2015). Package 'survival'. R Top Doc. 128(10): 28-33.

115.Tran KA, Addala V, Johnston RL, et al. 2023 Sep 16. Performance of tumour microenvironment deconvolution methods in breast cancer using single-cell simulated bulk mixtures. Nat Commun. 14(1):5758.

116.Tuong ZK, Loudon KW, Berry B, et al. 2021 Dec 21. Resolving the immune landscape of human prostate at a single-cell level in health and cancer. Cell Rep. 37(12):110132.

117.Tzanakis N, Gazouli M, Rallis G, et al. 2006 Dec 1 .Vascular endothelial growth factor polymorphisms in gastric cancer development, prognosis, and survival. J Surg Oncol. 94(7):624-30.

118.Uhlen M, Zhang C, Lee S, et al. 2017 Aug 18. A pathology atlas of the human cancer transcriptome. Science. 357(6352):eaan2507.

119.Underwood TJ, Hayden AL, Derouet M, et al. 2015 Feb. Cancer-associated fibroblasts predict poor outcome and promote periostin-dependent invasion in oesophageal adenocarcinoma. J Pathol. 235(3):466-77.

120.Vijver MJvd, He YD, Veer LJvt, et al. 2002 Dec 19. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 347(25): 1999-2009.

121.Vilella F, Wang W, Moreno I, et al. 2021 Sep 18. Single-cell RNA sequencing of SARS-CoV-2 cell entry factors in the preconceptional human endometrium. Hum Reprod. 36(10):2709-2719.

122.Wang W, Vilella F, Alama P, et al. 2020 Oct. Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. Nat Med. 26(10): 1644-1653.

123.Wang Z, An J, Zhu D, et al. (2022). Periostin: An emerging activator of multiple signaling pathways. J Cell Commun Signal.16(4):515-530.

124.West L, Vidwans S J, Campbell NP, et al. (2012). A novel classification of lung cancer into molecular subtypes. PLoS One. 7(2):e31906.

125.Wilson TJ, Nannuru KC, Futakuchi M, et al. 2010 Feb 28. Cathepsin G-mediated enhanced TGF-β signaling promotes angiogenesis via upregulation of VEGF and MCP-1. Cancer Lett. 288(2):162-9.

126.Wolfe CJ, Kohane IS, Butte AJ. 2005 Sep 14. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics. 6:227.

127.Wu SZ, Al-Eryani G, Roden DL, et al. 2021 Sep. A single-cell and spatially resolved atlas of human breast cancers. Nat Genet. 53(9):1334-1347.

128.Wyatt AW, Mo F, Wang K, et al. 2014 Aug 26. Heterogeneity in the inter-tumor transcriptome of high risk prostate cancer. Genome Biol. 15(8):426.

129.Xing F, Saidou J, Watabe K. 2010 Jan 1. Cancer associated fibroblasts (CAFs) in tumor microenvironment. Front Biosci (Landmark Ed). 15(1):166-79.

130.Xiong G-F, Xu R. 2016. Function of cancer cell-derived extracellular matrix in tumor progression. J Cancer Metastasis Treat. 2: 357-364.

131.Xu J, Stolk JA, Zhang X, et al. 2000 Mar 15. Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. Cancer Res. 60(6):1677-82.

132.Xue JM, Liu Y, Wan LH, et al. 2020 Feb 8. Comprehensive analysis of differential gene expression to identify common gene signatures in multiple cancers. Med Sci Monit. 26:e919953.

133.Yang Y, Han L, Yuan Y, et al. 2014. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat Commun. 5:3231.

134.Yokoyama Y, Grünebach F, Schmidt SM, et al. (2008). Matrilysin (MMP-7) is a novel broadly expressed tumor antigen recognized by antigen-specific T cells. Clin Cancer Res. 14(17):5503-11.

135.Young C, Brink R. 2021 Aug 10. The unique biology of germinal center B cells. Immunity. 54(8):1652-1664.

136.Yu G, Wang L-G, Han Y, et al. 2012 May. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS. 16(5):284-7.

137.Delakas, A, Lambrou GI, Boulalas I, et al. 2011 Apr 4. Identification of common differentially expressed genes in urinary bladder cancer. PLoS One. 6(4):e18135.

138.Zeng J, Zhang Y, Shang Y, et al. 2022 Jan 7. CancerSCEM: a database of single-cell expression map across various human cancers. Nucleic Acids Res. 50(D1):D1147-D1155.

139.Zhang M, Chen H, Wang M, et al. 2020 Feb 28. Bioinformatics analysis of prognostic significance of COL10A1 in breast cancer. Biosci Rep. 40(2):BSR20193286.

140.Zhang Y, Manouchehri Doulabi E, Herre M, et al. 2022 Apr 12. Platelet-Derived PDGFB Promotes Recruitment of Cancer-Associated Fibroblasts, Deposition of Extracellular Matrix and Tgfβ Signaling in the Tumor Microenvironment. Cancers (Basel). 14(8):1947.

141.Zhang Z, Wang Y, Zhang J, et al. 2018 Apr. COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. Mol Med Rep. 17(4):5037-5042.

142.Zhu P, Martinvalet D, Chowdhury D, et al. 2009 Aug 6. The cytotoxic T lymphocyte protease granzyme A cleaves and inactivates poly (adenosine 5'-diphosphate-ribose) polymerase-1. Blood. 114(6):1205-16.

