

## ARTICLE OPEN



# CAFrgDB: a database for cancer-associated fibroblasts related genes and their functions in cancer

Qiang Yuan<sup>1,3</sup>, Yi Chu<sup>1,3</sup>, Xiaoyu Li<sup>1</sup>, Yunshu Shi<sup>1</sup>, Yingying Chen<sup>1</sup>, Jimin Zhao<sup>1</sup>, Jing Lu<sup>1</sup>, Kangdong Liu<sup>1,2</sup>✉ and Yaping Guo<sup>1</sup>✉

© The Author(s) 2023

As one of the most essential components of the tumor microenvironment (TME), cancer-associated fibroblasts (CAFs) interact extensively with cancer cells and other stromal cells to remodel TME and participate in the pathogenesis of cancer, which earmarked themselves as new promising targets for cancer therapy. Numerous studies have highlighted the heterogeneity and versatility of CAFs in most cancer types. Thus, the identification and appropriate use of CAF-related genes (CAFGenes) in the context of specific cancer types will provide critical insights into disease mechanisms and CAF-related therapeutic targets. In this study, we collected and curated 5421 CAFGenes identified from small- or large-scale experiments, encompassing 4982 responders that directly or indirectly participate in cancer malignant behaviors managed by CAFs, 1069 secretions that are secreted by CAFs and 281 regulators that contribute in modulating CAFs in human and mouse, which covered 24 cancer types. For these human CAFGenes, we performed gene expression and prognostic marker-based analyses across 24 cancer types using TCGA data. Furthermore, we provided annotations for CAF-associated proteins by integrating the knowledge of protein-protein interaction(s), drug-target relations and basic annotations, from 9 public databases. CAFrgDB (CAF related Gene DataBase) is free for academic research at <http://caf.zbiolab.cn> and we anticipate CAFrgDB can be a useful resource for further study of CAFs.

*Cancer Gene Therapy* (2023) 30:917–925; <https://doi.org/10.1038/s41417-023-00603-4>

## INTRODUCTION

Mounting evidence shows that traditional treatment paradigms which predominantly target the tumor cells are usually not sufficient to root out their malignancy, as the tumor stroma or tumor microenvironment (TME) may hamper the efficacy of therapies and thus lead to therapeutic failure in clinical practice [1]. TME, as a complex system between tumor and their supporting stromal, including macrophages, endothelial cells, immune cell, vessels, extracellular matrix and fibroblasts, is depicted as the soil to participate in tumor initiation and progression and has drawn extensive attention in cancer research and therapy [2]. As a central component of the TME, cancer-associated fibroblasts (CAFs) can not only interact productively with cancer cells but also make a profound impact on the other components of the TME. Thus, CAFs have cemented themselves as a prominent functional role in modulating the efficacy of therapies and has evolved into a novel kind of therapy target.

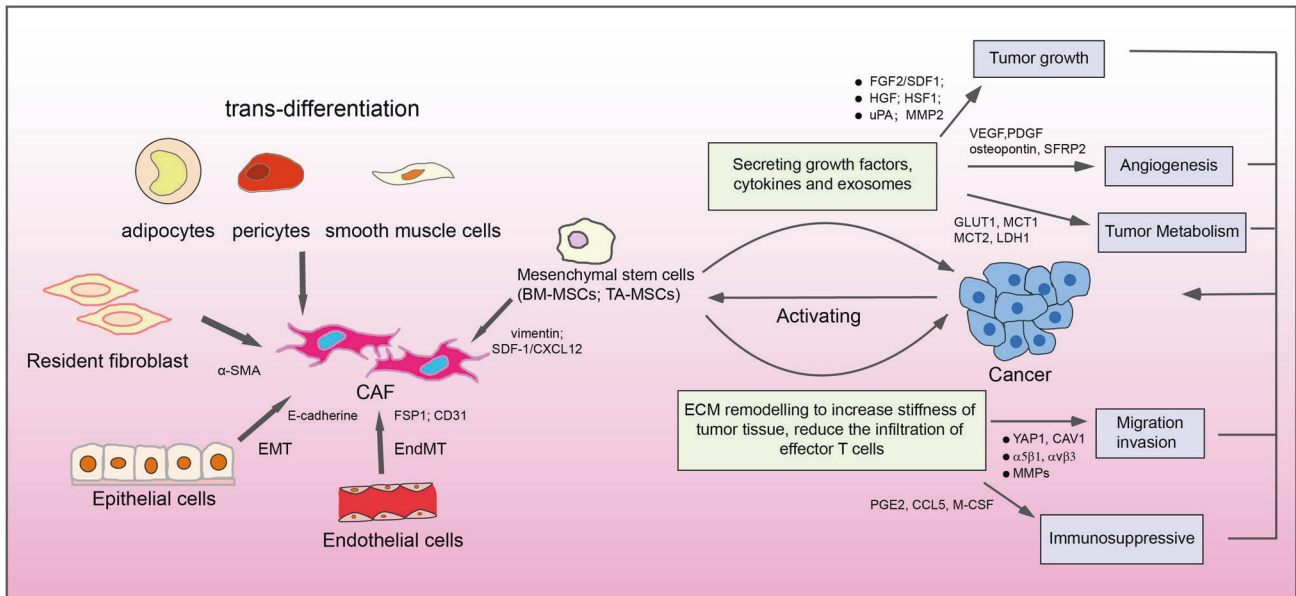
Emerging advances indicated that CAFs are complex populations with property of heterogeneity. On the one hand, the spectrum of origins contributes to explain the heterogeneity of CAFs. Resident fibroblast acquires a myofibroblast-like phenotype via transforming growth factor  $\beta$  (TGF $\beta$ ) activation, accompanying with  $\alpha$ -SMA expression [3]. Mesenchymal Stem Cells are one of the major sources of CAFs. Bone marrow-derived mesenchymal stem cells (BM-MSCs) can differentiate into CAFs, which feature a higher

SDF1 expression [4]. In addition, tumor-associated MSCs (TA-MSCs) also act as potential CAF precursor cells with vimentin expression [5]. Furthermore, there is evidence that epithelial cells in the contest of cancer go through epithelial-to-mesenchymal transition (EMT) described as lower level of E-cadherin, and can contribute to become CAF [6]. Another source of CAF is endothelial cells, which can acquire CAF phenotypes via the endothelial-to-mesenchymal transition (EndMT), upregulating FAP1 and downregulating CD31 [7]. Finally, a small portion from trans-differentiation of cells can also convert into CAFs, further share fibroblast-like properties, including pericytes, adipocytes, smooth muscle cells [8, 9] (Fig. 1).

Reactive fibroblast shares the particular features, such as enhanced capacity of protein synthesis, proliferation and contraction [10]. In terms of structure and function, CAFs make crucial contributions to the biological behavior of tumor cells via versatile mechanisms. For instance, CAFs serve as synthetic machines that release large quantities of factors to involve in tumor proliferation, angiogenesis, tumor metastasis. TGF $\beta$ , Fibroblast growth factor 2 (FGF2), matrix metalloproteases 2 (MMP2) and hepatocyte growth factor (HGF) derived from CAF endow tumor cells with stronger proliferative behavior [1, 11–13]. Many other factors from CAFs, such as VEGF, PDGF SFRP2, osteopontin, can drive tumor angiogenesis [1, 14]. CAFs share a metabolic symbiosis with tumor cells to favor malignancy progression. CAFs utilize metabolic reprogramming through “Warburg Effect” to promote malignant phenotype of cancer cells [15].

<sup>1</sup>Department of Pathophysiology, State Key Laboratory of Esophageal Cancer Prevention and Treatment, School of Basic Medical Sciences, Zhengzhou University, Zhengzhou 450001, China. <sup>2</sup>China-US (Henan) Hormel Cancer Institute, Zhengzhou, Henan 450001, China. <sup>3</sup>These authors contributed equally: Qiang Yuan, Yi Chu. ✉email: [kdliu@zzu.edu.cn](mailto:kdliu@zzu.edu.cn); [guoyaping@zzu.edu.cn](mailto:guoyaping@zzu.edu.cn)

Received: 11 August 2022 Revised: 3 February 2023 Accepted: 23 February 2023  
Published online: 15 March 2023



**Fig. 1 Origin of CAFs and summary CAF-mediated regulation of cancer biological behavior.** Potential cellular sources of CAF including resident fibroblast, bone marrow-derived mesenchymal stem cells and tumor-associated MSCs, epithelial cells, endothelial cells, and trans-differentiation of cells (pericytes, adipocytes, smooth muscle cell). CAFs contribute to tumor progression through multiple mechanisms. CAFs induce EMT remodeling, and produce a plethora of pro-tumorigenic factors to orchestrate the development of cancer through proliferation, angiogenesis, metastasis, changing metabolic state of the tumor and immunosuppression.

MCT1, MCT2, LDH1 and so on are found as higher expression in CAFs [16, 17]. Meanwhile, CAFs also remodel the extracellular matrix (ECM) structure to guide cancer cell to invade. CAFs secrete matrix MMPs to affect ECM stiffness, and YAP1 or CAV1 overexpression in CAFs can highly induce directional migration and invasiveness of carcinoma cells [12, 18, 19]. CAFs also build up ECM structure through  $\alpha$ v $\beta$ 3 integrin and  $\alpha$ 5 $\beta$ 1 integrin to support cancer cell [20]. On the other hand, ECM remodeling by CAFs forms a physical barrier against the immune system. CAFs produce some factors, such as CCL5, PGE2, M-CSF, to have an important role in immune evasion [1, 5, 21] (Fig. 1). Although increasing evidences suggested CAFs have negative regulation in cancer, the exact mechanism needs to illustrate.

Diversity of cellular sources manifests the heterogeneity of CAF. Such heterogeneity also is ascribed to the distribution of biological markers and cancer types, which is described as a strategy to at least partially contribute to distinct CAF subpopulations [22].  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA), fibroblast activation protein (FAP), fibroblast specific protein 1 (FSP1, also termed as S100A4) and platelet-derived growth factor receptor- $\alpha$ / $\beta$  (PDGFR $\alpha$ / $\beta$ ) are considered as marker to define CAFs [23]. However, the identification of these candidate markers has been challenging due to the fact that other stromal cell populations also shared [24]. Hence, to understand their biological roles, CAFs subpopulation is divided into different subtypes according to the type of cancer and the distribution of markers [25]. Although CAFs subpopulation improved understanding of CAFs heterogeneity and functional diversity, targeting heterogeneity therapies has remained many challenges. Thus, there is an urgent need to understand functional heterogeneity of CAFs in depth.

In this study, we collected and curated 5421 CAFGenes. According to their function and distribution, we annotated them in detail. Owing to different roles in CAFs, these proteins are divided into three types, respectively named regulator, responder and secretion. In addition, we provided rich annotations for CAFs-related proteins by integrating the knowledge from 9 additional resources that covered many aspects including the knowledge of cancer-associated information, protein-protein interaction(s),

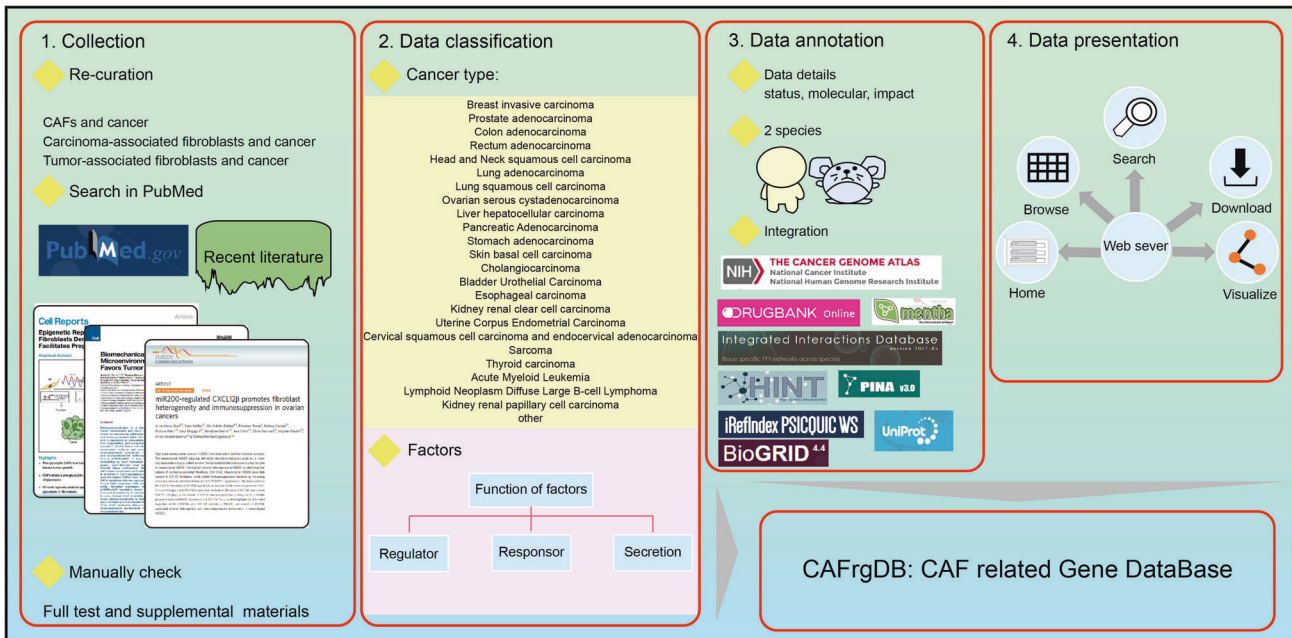
drug-target relations and basic annotations. The CAFrgDB database will be continuously updated and could also serve as a valuable tool to understand CAFs function.

## METHODS

### Data collection and curation, classification

In CAFrgDB 1.0, we manually collected 5421 CAFGenes together with their exact functions from the literature. To obtain known CAFGenes from the literature, we manually searched the PubMed database by a set of topically relevant keywords, such as “cancer-associated fibroblasts and cancer”, “CAF and cancer” and “tumor-associated fibroblasts and cancer”. Totally, 13,621 entries are recorded involved in influencing the function of CAFs from 24 cancer type. For each literature, the information regarding samples was extracted. In addition, we directly retrieved other pertinent information such as cancer type, expression of CAF-related factors, experimental conditions along with the regulation mode of each factor from the research articles. Labels ‘increase’ or ‘decrease’ were used to indicate up-regulation or down-regulation of the regulation factors that enhance CAFs function (Fig. 2).

Notably, all collected genes were classified into secretion, regulator or responder. Secretion was defined as the proteins which were secreted by CAFs. For example, FGF2, a well-characterized CAF-secreted protein contributes to tumor cell growth [26]. Thus, some factors or cytokines deriving from CAFs are categorized into “secretion”. In addition, it has been discovered that various tumor-derived stimuli signals can activate quiescent fibroblasts to become CAFs, which include soluble factors produced by the tumor and immune infiltrate such as TGF $\beta$  family ligands, lysophosphatidic acid (LPA), FGF, platelet-derived PDGF and interleukin-1 (IL-1), and here we classified the above of these proteins as regulator. Furthermore, a large number proteins were identified to be participated in CAF-mediated cancer progression. CAFs act as highly heterogeneous stromal cells, the interaction between CAFs and cancer cell supports tumorigenesis in various ways, especially signaling pathways. Meanwhile the crosstalk also mediates the biological behaviors of CAFs [27]. CAFs promote cancer progression through activation signal cascades including PI3K/AKT/mTOR signaling pathway, TGF- $\beta$  signaling pathway, MAPK signaling pathway, etc [28]. Several major signal cascades involving in CAFs tumor-promoting or tumor-restraining phenotype, named as responder. Responder is recognized as a range of genes executing CAFs phenotype. These three components are not independent of others, and there is some overlap between them. Of course,



**Fig. 2 Overview of CAFrgDB.** Data were collected from published literature in PubMed through manually check. We classified all collected proteins into 3 categories including responzor, regulator and secretion from 24 cancer type. A total of 5421 CAF-associated proteins are annotated in detail from human and mouse. Further, we further integrated annotations in 9 public data resources covering 5 aspects: protein function, interacting proteins, FDA-approved drugs, expression and prognosis in TCGA. Eventually, all entry were obtained with reliable evidence and CAFrgDB database was constructed. Users can access these different categorized annotations by clicking the picture.

these molecules might play entirely different roles in CAF from different cancer, which also reveals heterogeneity of CAFs.

**Experimental conditions**

Experimental conditions include cancer types or cell lines, eukaryotic species, which are extracted from the literature. In CAFrgDB database, diverse cell lines from the same tissue also are recognized as parallel samples. To explore the relationship between CAFs and cancer, we also integrated transcriptome, proteomics and secretome from literature into our dataset. Meanwhile, we recorded the expression signature corresponding to their exact effect such as promotion or inhibition in CAFs.

**Protein basic and annotation information**

After filtering for duplicates and reclassifying cancer subtypes, CAFrgDB encompassed 13,621 entries involving 24 cancer subtypes. For each study, we recorded how the CAFGenes regulate or effect the function of CAFs or act as secretor derived from CAFs in different cancer type. To enrich the data and facilitate users, we also integrated UniProt [29] and The Cancer Genome Atlas (TCGA) database [30] to annotate the CAFGenes. The basic information contains gene name, protein name, organism, UniProt ID, protein sequence, and function. According to the role in CAFs, we described these items as different CAFs type. More importantly, we provided detailed evidence for these factors in CAF including cancer type and regulation mode. Further, we used UniProt database to present annotation of the sequence and function of factors [31]. TCGA is currently the largest available dataset for deep sequencing for cancer patients [32]. The CAFrgDB exhibits various genome-wide data including gene expression data and clinical data from TCGA database, which provides clinical value for cancer research. We also depict the interaction partners in CAFs by compiling and integrating protein-protein Interaction database including IID [33], iRefIndex [34], PINA [35], HINT [36], Mentha [37] and BIOGRID [38]. Because targeting CAF heterogeneity has been expected to be a novel therapy for cancer, we integrated the information of Drugbank [39] (Fig. 2).

**Enrichment analysis based on CAFGenes**

Two-sided hypergeometric test was adopted for enrichment analysis of each type of CAFGene based on the annotations of KEGG. For each type, we defined the following:

$N$  = number of human proteins annotated by at least one term  
 $n$  = number of proteins annotated by term  $t$   
 $M$  = number of CAFGenes in each type annotated by at least one term  
 $m$  = number of CAFGenes in each type annotated by term  $t$   
 Then, the enrichment ratio was computed, and the  $p$ -value was calculated based on the hypergeometric distribution as below:

$$Enrichment\ ratio = \frac{m}{n} \cdot \frac{N}{M}$$

$$p - value = \sum_{m'=m}^n \frac{\binom{M}{m'} \binom{N-M}{n-m'}}{\binom{N}{n}} \quad (Enrichment\ ratio \geq 1)$$

or

$$p - value = \sum_{m'=0}^m \frac{\binom{M}{m'} \binom{N-M}{n-m'}}{\binom{N}{n}} \quad (Enrichment\ ratio < 1)$$

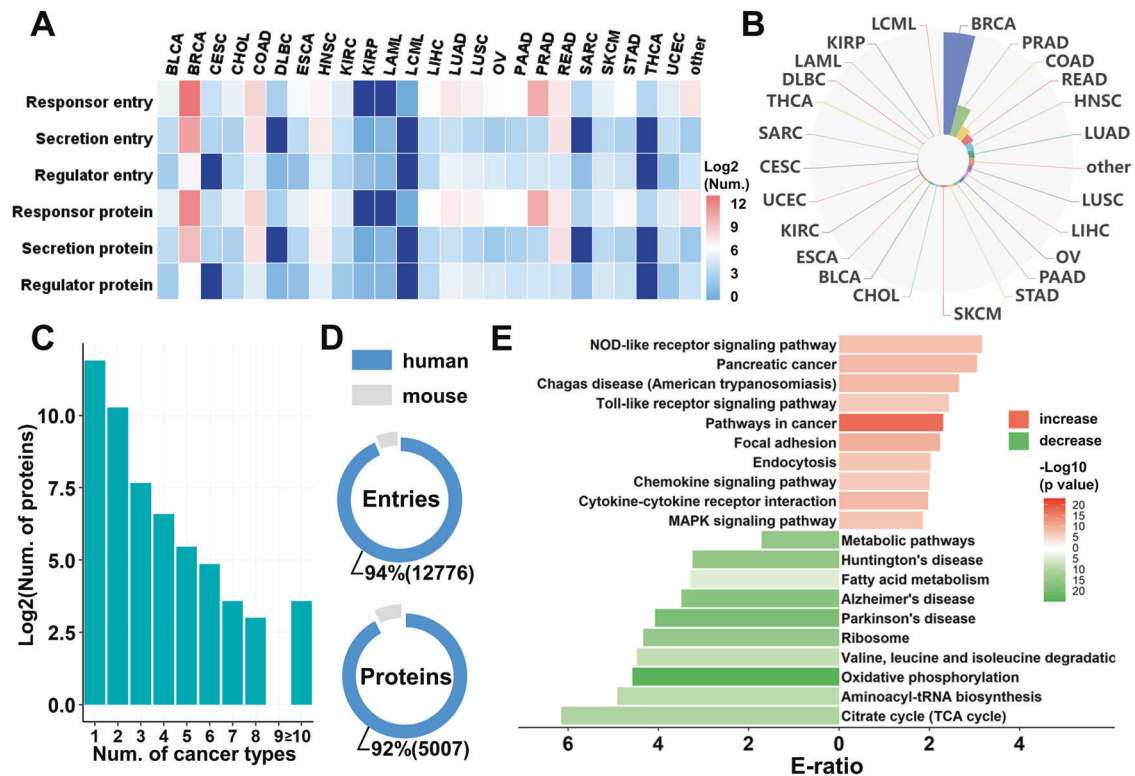
**Web interface implementation**

All the analysis results and entry information were curated using MySQL tables (<https://www.mysql.com/cn/>). The web interfaces of CAFrgDB were implemented in PHP + MySQL + JavaScript. Furthermore, we inlayed multiple retrieval mode and allowed the users to search for and download the corresponding information. All datasets and annotations can be downloaded at: <http://caf.zbiolab.cn/download.php>.

**RESULTS**

**Overview of CAFrgDB**

CAFrgDB serves as the first literature-based database of the relations between CAFs and cancer. To establish CAFrgDB, we first collected all reported CAFGenes and their functions from the



**Fig. 3 The statistics and analysis of the dataset. A** Numbers of responder, regulator and secretor in different cancer. **B** Numbers of CAF-related factors in different cancer type are summarized. **C** The distribution of numbers of CAF-entry in different cancer. **D** Comparison the number of entries from human and mouse. **E** The KEGG-based enrichment analysis of the CAF-related proteins.

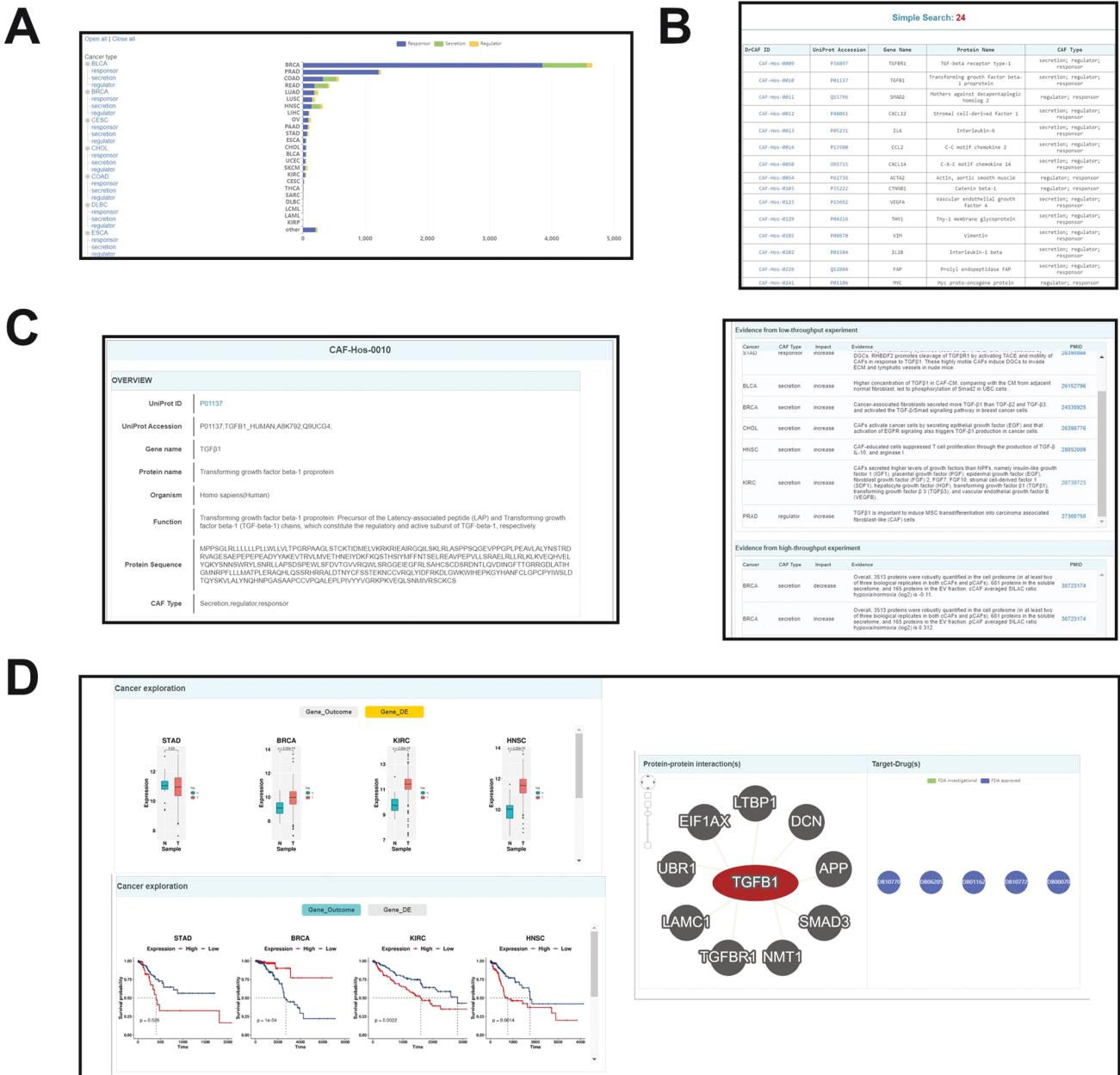
literature. After manual check of the full text and supplementary materials, 13621 entries involving 24 cancer types were obtained. Detailed information including cancer type or cell lines used in experiments, factors, the status of factors in CAFs and evidence of the relations were provided in the Webserver. In the current version of database, these entries were divided into 3 groups following the classification scheme from CAFrgDB. More specifically, we divide 13621 entries into three categories including 4982 responders that participate in the function of CAFs to cancers, 1069 secretions that are secreted by CAFs and 281 regulators that contribute in modulating CAFs in human and mouse (Fig. 2). As shown in Fig. 3A, the distribution of three CAF type in different cancer is described. For example, totally 9352 entries have been implicated in breast cancer, all of which 114 are secretions, 1723 entries are regulators, and up to 7515 proteins acting as responder that execute the function of CAFs. Presently, the largest number of CAF entries observed is for breast cancer, which encompasses 69% of all entries in our dataset, followed by prostate cancer (9%), colon cancer (4%) (Fig. 3B) (Table S1). In Fig. 3C, we summarized each of CAF-related entries involved in diverse cancer progression. The results indicated that only a small portion of protein factors are prone to involve in multiple cancers. However, notably, these proteins are more specific, and we found that more than 50% of known entries were only detected in one cancer type, suggesting they might be candidate CAF proteins for explanation the heterogeneity of CAFs. Only 11 proteins participate more than 10 types of cancer, including IL6, MET, IL8, CCL2, VEGFA, AKT1, AKT2, STAT3, SDF1, ACTA, SEPR (Table S2). In our dataset, known CAFGenes were collected from 2 eukaryotic organisms, including *Homo sapiens* and *Mus musculus*. In total, there were 12776 (92%) entries of 5007 (94%) proteins in human, and the result indicated that the most of CAFGenes-related experiments were conducted in human (Fig. 3D).

### Functions of CAFGenes

The status of factors in CAFs includes increase and decrease, and an enrichment analysis was performed based on the information. The top 10 mostly enriched KEGG pathways were chosen and visualized. These increasing-factors were enriched in NOD-like receptor signaling pathway (hsa04621), Toll like receptor signaling pathway (hsa04620), and Pathways in cancer (hsa05200). Furthermore, decreasing-factors were enriched in metabolic pathways (Fig. 3E).

Using the 5007 human CAFGenes belonged to 3 categories, we performed an enrichment analysis for each category of CAFGenes based on Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations with the hypergeometric test [40, 41]. The mostly enriched KEGG pathways were chosen and visualized for each of these three categories. For the responders, we observed that they were significantly enriched in cancer metabolism-associated processes, such as Spliceosome (hsa03040), Ribosome (hsa03010) and Oxidative phosphorylation (hsa00190). In addition, cancer-associated pathways including Pathways in cancer (hsa05200) and Citrate cycle (TCA cycle) (hsa00020) were also significantly enriched in responders. The results were highly consistent with experimental studies [42, 43]. For the secretions, such as Focal adhesion (hsa04510), ECM-receptor interaction (hsa04512), Regulation of actin cytoskeleton (hsa04810), Antigen processing and presentation (hsa04612) and Complement and coagulation cascades (hsa04610) were enriched and demonstrated that these CAFGenes were highly involved in ECM remodeling, which underlines the impact that CAFs present the ability to alter the surrounding environment via allowing the tumor matrix to be remodeled [44]. Further analysis of regulators indicated that they were highly involved in regulating various types of signal transduction associated processes including TGF-beta signaling pathway (hsa04350), Cytokine-cytokine receptor interaction





**Fig. 4 Overview of the CAFrgDB web interface. A** The interface for data browsing. **B** Browse page for CAF type. **C** Basic annotation page of human TGFβ1 in CAFrgDB database. **D** Detailed information regarding TGFβ1 in the other database.

(hsa04060) together with NF-kappa B signaling pathway (hsa04064) and immune-associated processes such as Toll-like receptor signaling pathway (hsa04620), NOD-like receptor signaling pathway (hsa04621) and Chemokine signaling pathway (hsa04062) (Table S3). Notably, these results were not only highly consistent with previous studies [45, 46], but also provided valuable information for further deciphering versatile roles of CAFGenes. It further provided clear evidence that CAFGenes play key roles in the response to cancer development, which indicated that CAFs have great potential to become target for cancer therapy.

**A comprehensive annotation of CAFGenes**

We constructed CAFrgDB as a gene-centered database, and a variety of basic annotations, such as protein/gene names/aliases, UniProt accession numbers, functional descriptions and protein sequences were obtained from UniProt databases. For each

known CAFrgDB-related protein, brief descriptions on its roles in CAF were present, and we termed “CAF type” including regulator, responder and secretion. In addition, CAFrgDB presented two aspect evidences derived from low-through experiment and high-through experiment. In particularly, for each known CAFGenes, descriptions on available assays of sets of experiments were also summarized. Corresponding cancer type or cell lines for experimental analyses were supported, as well as users can acquire raw resource via PMIDs of primary references.

Moreover, importantly, by integrating the knowledge of 9 additional databases, we further annotated CAFGenes from 5 aspects. CAFrgDB database also provided specific annotations, for instance mRNA expression profiles in TCGA and prognosis of patients was also evaluated from TCGA in different cancer. We presented the analysis of expression level of 4974 CAFGenes transcripts from TCGA database and found total more than 30% genes ( $|\log_2(\text{Foldchange})| > 1$  and  $p\text{-value} < 0.05$ ) showed

significantly differential expression between cancer samples and adjacent normal samples at least in one cancer type. Furthermore, we also provided clinical association analysis about relationship between gene expression and patients survival, which accounts for more than 90% (Logrank test,  $p < 0.05$ ) CAFGenes expression could indicated the prognosis of the patients at least in one cancer type. Herein, we developed comprehensive database, which included information on numerous PPIs in the BIOGRID, PINA, HINT, IID, MENTHA and iRefIndex database [46], corresponding to CAFGenes interactions, thus providing extensive information regarding CAF function. In addition, more than 90% CAFGenes have specific regulator network. In order to improve clinical utility, we also integrated the information of DrugBank and 1280 CAFGenes have FDA-approved drugs for further research reference.

### Web design and interface

CAFRgDB website feature a freely available and user-friendly interface for users to explore CAFs data within different cancer. The website comprises four sections: 'Home', 'Browse', 'Search', and 'Download'. On the 'Home' page, users can acquire an overview of CAFs, including the total number of CAFs-related factors in different cancer from human and mouse. Based on the function, we classified factors into three categories including responder, regulator and secretor, termed as CAF-related genes. For browsing the data in CAFrgDB, we implemented two options, including 'Browse by CAF types', and 'Browse by cancer type' (Fig. 4A). By clicking a specific cancer type, information of CAF-related proteins is shown in a table (Fig. 4B). Each entry in CAFrgDB database has a unique identification, named CAFrgDB ID, and we provided gene name, protein name, description from Uniprot and CAF type. Here, we selected the human TGF $\beta$ 1 protein as an example to depict the usage of CAFrgDB. For browsing the data in CAFrgDB, we implemented two options, including 'Browse by cancer types' and 'Browse by CAF type'. In the option of 'Browse by cancer types', users can click the cancer type of 'ESCA' under the cancer types super-class to browse all known CAFGenes involved in ESCA (Fig. 4A). Since human TGF $\beta$ 1 is a known regulator for CAFs, users can also directly click 'regulator' in the option of 'Browse by CAF types', and then the result will be displayed in a table of the returned page for users to view all human CAF regulators (Fig. 4B). By clicking "BLCA", "regulator" and "CAF-Hos-0010", the final page of human TGF $\beta$ 1 will be shown (Fig. 4C). In the gene page, basic annotations such as protein/gene names, Uniprot gene/protein IDs, and CAF type can be viewed (Fig. 4C left). The two aspect evidences derived from low-through experiment and high-through experiment of CAFGenes and its primary references PMIDs were also be linked (Fig. 4C right).

As for additional fundamental annotations, we compiled and integrated the knowledge from 9 additionally public resources. For additional annotations, user will obtain three modules including "cancer exploration", "protein-protein interaction" and "target-drug". In "cancer exploration" section, users can click the "GeneDE" button to obtain multiple barcharts displaying the gene expression levels among different tumor tissues (Fig. 4D left). Besides, by clicking the 'Gene Outcome' tag, all survival analysis of this CAFGenes based on Kaplan–Meier method will also be displayed (Fig. 4D left). We also designed a "protein-protein interaction" module to construct synergistic regulatory network online by integrating three protein-protein interaction databases including IID, PINA and HINT. Furthermore, the database provided some FDA-approved drugs targeting the protein in "target-drug" section (Fig. 4D right). Specially, the annotation datasets can be freely downloaded at <http://caf.zbiolab.cn/download.php>. These interesting findings will benefit users that are devoted to investigating the regulatory mechanisms of CAFs. The three options for browsing the database, we also provided several

options, including 'Simple Search', 'Batch Search' and 'Advance Search', for searching the data in CAFrgDB (<http://caf.zbiolab.cn>).

### DISCUSSION

Malignant behavior of tumor cells is partially determined by TME and CAF is a cell type of paramount importance for TME with diverse functions, including secreting growth factors or cytokines and remodeling the tumor stroma to impact tumor malignancy [27, 47]. Thus, numerous evidence proposed a new cancer treatment paradigm, targeting CAFs [48, 49]. TME elicit a wide spectrum of dynamic alterations in cancer development, and CAFs have important roles in cancer pathogenesis within TME. Based on complex of TME and partially lack of specific CAF cell surface marker, precisely targeting CAFs has obtain some therapy effect along with side-effect [50–52]. Several major signal pathways affect not only the biological functions of CAFs but also the crosstalk between CAFs and cancer cells, even they also participate in malignant behaviors of cancers. Thus, it seems more feasible to target activation signaling and downstream effectors of CAFs. Owing to the complex inter-cellular interactions and regulator systems involving CAFs in the tumor stroma, it is hard to precisely target CAFs without influencing other cell populations and it is reasonable to be forced on cancer cells. Therefore, in this study, we summarized CAF-related gene and evaluated their expression profile in different cancer types via TCGA database, and tried to find these factors sharing tumorigenic phenotype and function in CAFs. We considered that targeting these factors not only might inhibit the cancer cell, more importantly convert TME into tumor-suppressive microenvironment. In order to speed up the development to clinic, we also integrated the DrugBank information, which provided rich annotations for therapy strategy.

Distinct sets of biomarkers and different classification strategies have been used to paraphrase diverse CAF subpopulations in different cancer types [53]. Particularly, emerging evidence provided a detailed characterization of CAFs in PDAC. There is lower expression of  $\alpha$ -SMA and higher expression of inflammatory markers IL-6 and LIF, which are termed inflammatory CAFs (iCAFs) [3, 54]. Although iCAFs are more distal to cancer cells, they share immunomodulating secretome to involve in tumorigenesis. Another CAF populations are myofibroblasts (myCAFs), which is defined higher-SMA levels and exhibits myofibroblastic phenotypes. Compared with iCAFs, myCAFs are adjacent to tumor foci [55, 56]. In addition, CAFs subpopulation is possibly caused by marker expression or gene profiling. In colorectal cancer, according to gene expression and marker profiling, CAFs is identified with two distinct subpopulations, CAF-A and CAF-B. ACTA2, TAGLN and PDGFA serve as markers for CAF-B, whereas CAF-A cells specially express MMP2, DCN and COL1A2 [22]. Functionally, CAF-A cells are confirmed to express rebuilding ECM-related proteins, while CAF-B cells express cytoskeleton-related genes [57]. The heterogeneity of CAFs also present heterogeneous functions on cancer or produce a range of tumor-promoting (pCAFs) or tumor-restraining (rCAFs) effects in diverse cancers [58]. Although CAFs subpopulation highlights the requirement for studying CAFs in physiologically function, CAFs exhibit molecular and functional heterogeneity in different cancers and even at different stages of the same cancer. Therefore, any therapeutic strategies for targeting CAFs should exploit the specificity and diversity of CAFs.

With the development of single-cell RNA sequencing (scRNA-seq), scRNA-seq analysis has provided a deeper understanding of CAF heterogeneity with diverse putative functions in different cancer types, which has been performed to define different potential subpopulations of CAFs. For example, in lung cancer, clusters 2 highly expressing ACTA2 involved in angiogenesis; cluster 1 are enriched in tumor with strongly ECM proteins;

clusters 4 are enriched at the leading edge of tumor mass; clusters 5 and clusters 7 possess similar high mTOR gene signature, but they are different from location distribution [53, 59]. Similarly, four CAFs subpopulations are identified in breast cancer [60–62]. Furthermore, scRNA-seq has been conducted to reveal heterogeneous alterations including tumor-promoting phenotype from non-invasive intraductal papillary mucinous neoplasms (IPMNs) to PDAC. During the progression of PDAC, CAFs subpopulation can shift between myCAF and iCAF. During non-invasive dysplasia to invasive cancer, myCAF proportion is increasing from LGD-IPMNs to HGD-IPMNs, but in invasive cancer iCAF is in a dominant position [63].

Recently, mounting studies for targeting CAFs have been explored in preclinical as below. Firstly, Directing CAF depletion via cell surface markers. FAP and  $\alpha$ SMA act as marker of CAFs, have also been used to pharmacologically deplete CAFs. In PDAC, selective depletion of the  $\alpha$ -SMA + myofibroblasts suppressed cancer metastasis [64]. A number of preclinical studies have been reported that specific depletion of FAP-expressing CAFs or pharmacological inhibition of FAP also has shown promising antitumor activities [65].  $\alpha$ FAP-PE38 as FAP-targeting immunotoxin presents attractive tumor suppression in breast cancer mouse model [50]. Secondly, reprogramming CAFs to quiescent state. CAFs simultaneously possess tumor-promoting and tumor-restraining function, the transfer function depending on TME. Therefore, strategy would be designed to tumor-promoting CAFs population or inducing tumor-promoting CAFs into quiescent state or even tumor-restraining phenotype. One of the examples of such an approach is supported by providing all-trans retinoic acid (ATRA) in pancreatic cancer. ATRA reset the fibroblasts to an inactive state and leads to reduced tumor growth via inhibiting Wnt–Catenin Signaling [66]. Targeting ECM proteins. Several clinical trials have been evaluated chemotherapeutic agents targeting Hyaluronan (HA) derived from CAFs in PDAC and gastric cancer. Currently, pathway inhibitors including JAK inhibitors [67], PDGFR inhibitors [68] and so on have conducted in clinical trials. In this study, we summarized all proteins involving in CAFs function and evaluated their expression profile in different cancer types via TCGA database, which might help optimize combination therapy strategy for target CAFs.

In summary, CAFrgDB hosts 13,621 entries of CAFGenes, which covered 24 cancer types. Furthermore, we provided annotations for CAF-associated proteins by integrating the knowledge of basic information, cancer-associated information, protein-protein interaction(s), drug-target relations and basic annotations, from 9 public databases. The database serves as a useful resource to help understanding the heterogeneity of CAF and such an integrative annotation supports valuable information for researchers. In the future, we will attempt to continuously maintain and update CAFrgDB by collecting and annotating newly identified CAFGenes and their exact functions in cancers. It should be noted that new functions might be reported for existing proteins in CAFrgDB, and their CAF types and the classification information will also be refined. We anticipate CAFrgDB can serve as a useful resource for further study of CAFGenes.

## DATA AVAILABILITY

All data generated and analyzed during this study can be download at <http://caf.zbiolab.cn/download.php>.

## REFERENCES

- Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. *Nat Rev Drug Disco*. 2019;18:99–115.
- Negrete-Garcia MC, Ramirez-Rodriguez SL, Rangel-Escareno C, Munoz-Montero S, Kelly-Garcia J, Vazquez-Manriquez ME, et al. Deregulated microRNAs in cancer-associated fibroblasts from front tumor tissues of lung adenocarcinoma as potential predictors of tumor promotion. *Tohoku J Exp Med*. 2018;246:107–20.
- Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, et al. IL1-induced JAK/STAT signaling is antagonized by TGFbeta to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Disco*. 2019;9:282–301.
- Mishra PJ, Mishra PJ, Humeniuk R, Medina DJ, Alexe G, Mesirov JP, et al. Carcinoma-associated fibroblast-like differentiation of human mesenchymal stem cells. *Cancer Res*. 2008;68:4331–9.
- Liu T, Han C, Wang S, Fang P, Ma Z, Xu L, et al. Cancer-associated fibroblasts: an emerging target of anti-cancer immunotherapy. *J Hematol Oncol*. 2019;12:86.
- Huang L, Xu AM, Liu S, Liu W, Li TJ. Cancer-associated fibroblasts in digestive tumors. *World J Gastroenterol*. 2014;20:17804–18.
- Ganguly D, Chandra R, Karalis J, Teke M, Aguilera T, Maddipati R, et al. Cancer-associated fibroblasts: versatile players in the tumor microenvironment. *Cancers (Basel)*. 2020;12:2652.
- Roy A, Bera S. CAF cellular glycolysis: linking cancer cells with the micro-environment. *Tumor Biol*. 2016;37:8503–14.
- Miyazaki Y, Oda T, Mori N, Kida YS. Adipose-derived mesenchymal stem cells differentiate into pancreatic cancer-associated fibroblasts in vitro. *FEBS Open Bio*. 2020;10:2268–81.
- Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer*. 2016;16:582–98.
- Scherz-Shouval R, Santagata S, Mendillo ML, Sholl LM, Ben-Aharon I, Beck AH, et al. The reprogramming of tumor stroma by HSF1 is a potent enabler of malignancy. *Cell*. 2014;158:564–78.
- Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer*. 2020;20:174–86.
- Deying W, Feng G, Shumei L, Hui Z, Ming L, Hongqing W. CAF-derived HGF promotes cell proliferation and drug resistance by up-regulating the c-Met/P13K/Akt and GRP78 signalling in ovarian cancer cells. *Biosci Rep*. 2017;37:BSR20160470.
- Kubo N, Araki K, Kuwano H, Shirabe K. Cancer-associated fibroblasts in hepatocellular carcinoma. *World J Gastroenterol*. 2016;22:6841–50.
- Fiaschi T, Marini A, Giannoni E, Taddei ML, Gandellini P, De Donatis A, et al. Reciprocal metabolic reprogramming through lactate shuttle coordinately influences tumor-stroma interplay. *Cancer Res*. 2012;72:5130–40.
- Becker LM, O'Connell JT, Vo AP, Cain MP, Tampe D, Bizarro L, et al. Epigenetic reprogramming of cancer-associated fibroblasts deregulates glucose metabolism and facilitates progression of breast cancer. *Cell Rep*. 2020;31:107701.
- Ippolito L, Morandi A, Taddei ML, Parri M, Comito G, Iscaro A, et al. Cancer-associated fibroblasts promote prostate cancer malignancy via metabolic rewiring and mitochondrial transfer. *Oncogene*. 2019;38:5339–55.
- Goetz JG, Minguet S, Navarro-Lerida I, Lazzano JJ, Samaniego R, Calvo E, et al. Biomechanical remodeling of the microenvironment by stromal caveolin-1 favors tumor invasion and metastasis. *Cell*. 2011;146:148–63.
- Hassona Y, Cirillo N, Heesom K, Parkinson EK, Prime SS. Senescent cancer-associated fibroblasts secrete active MMP-2 that promotes keratinocyte dis-cohesion and invasion. *Brit J Cancer*. 2014;111:1230–7.
- Kechagia JZ, Ivaska J, Roca-Cusachs P. Integrins as biomechanical sensors of the microenvironment. *Nat Rev Mol Cell Bio*. 2019;20:457–73.
- Barrett R, Pure E. Cancer-associated fibroblasts: key determinants of tumor immunity and immunotherapy. *Curr Opin Immunol*. 2020;64:80–87.
- Nurmik M, Ullmann P, Rodriguez F, Haan S, Letellier E. In search of definitions: cancer-associated fibroblasts and their markers. *Int J Cancer*. 2020;146:895–905.
- Bu L, Baba H, Yoshida N, Miyake K, Yasuda T, Uchihara T, et al. Biological heterogeneity and versatility of cancer-associated fibroblasts in the tumor micro-environment. *Oncogene*. 2019;38:4887–901.
- Zhai J, Shen J, Xie G, Wu J, He M, Gao L, et al. Cancer-associated fibroblasts-derived IL-8 mediates resistance to cisplatin in human gastric cancer. *Cancer Lett*. 2019;454:37–43.
- Park D, Sahai E, Rullan A. SnapShot: Cancer-associated fibroblasts. *Cell*. 2020;181:486–486.e1.
- Hegab AE, Ozaki M, Kameyama N, Gao J, Kagawa S, Yasuda H, et al. Effect of FGF/FGFR pathway blocking on lung adenocarcinoma and its cancer-associated fibroblasts. *J Pathol*. 2019;249:193–205.
- Wu F, Yang J, Liu J, Wang Y, Mu J, Zeng Q, et al. Signaling pathways in cancer-associated fibroblasts and targeted therapy for cancer. *Signal Transduct Target Ther*. 2021;6:218.
- Teng F, Tian WY, Wang YM, Zhang YF, Guo F, Zhao J, et al. Cancer-associated fibroblasts promote the progression of endometrial cancer via the SDF-1/CXCR4 axis. *J Hematol Oncol*. 2016;9:8.
- UniProt C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480–9.



30. Blum A, Wang P, Zenklusen JC. SnapShot: TCGA-Analyzed tumors. *Cell*. 2018;173:530.
31. Kulandaiamy A, Sakthivel R, Gromiha MM. MPTherm: database for membrane protein thermodynamics for understanding folding and stability. *Brief Bioinform*. 2021;22:2119–25.
32. Lee M, Lee K, Yu N, Jang I, Choi I, Kim P, et al. ChimerDB 3.0: an enhanced database for fusion genes from cancer transcriptome and literature data mining. *Nucleic Acids Res*. 2017;45:D784–D789.
33. Kotlyar M, Pastrello C, Malik Z, Jurisica I. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res*. 2019;47:D581–D589.
34. Razick S, Magklaras G, Donaldson IM. iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinforma*. 2008;9:405.
35. Cowley MJ, Pines M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM, et al. PINA v2.0: mining interactome modules. *Nucleic Acids Res*. 2012;40:D862–5.
36. Das J, Yu H. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol*. 2012;6:92.
37. Calderone A, Castagnoli L, Cesareni G. mentha: a resource for browsing integrated protein-interaction networks. *Nat Methods*. 2013;10:690–1.
38. Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*. 2021;30:187–200.
39. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46:D1074–D1082.
40. The Gene Ontology C. Expansion of the gene ontology knowledgebase and resources. *Nucleic acids Res*. 2017;45:D331–D338.
41. Ning W, Lin S, Zhou J, Guo Y, Zhang Y, Peng D, et al. WocEA: The visualization of functional enrichment results in word clouds. *J Genet Genomics*. 2018;45:415–7.
42. Gong J, Lin Y, Zhang H, Liu C, Cheng Z, Yang X, et al. Reprogramming of lipid metabolism in cancer-associated fibroblasts potentiates migration of colorectal cancer cells. *Cell Death Dis*. 2020;11:267.
43. Yaffe MB. Phosphotyrosine-binding domains in signal transduction. *Nat Rev Mol Cell Biol*. 2002;3:177–86.
44. Kugeratski FG, Atkinson SJ, Neilson LJ, Lilla S, Knight JRP, Serneels J, et al. Hypoxic cancer-associated fibroblasts increase NCBP2-AS2/HIAR to promote endothelial sprouting through enhanced VEGF signaling. *Sci Signal*. 2019;12:eaan8247.
45. Woods NT, Mesquita RD, Sweet M, Carvalho MA, Li X, Liu Y, et al. Charting the landscape of tandem BRCT domain-mediated protein interactions. *Sci Signal*. 2012;5:rs6.
46. Reinhardt HC, Yaffe MB. Phospho-Ser/Thr-binding domains: navigating the cell cycle and DNA damage response. *Nat Rev Mol Cell Biol*. 2013;14:563–80.
47. Pape J, Magdeldin T, Stamati K, Nyga A, Loizidou M, Emberton M, et al. Cancer-associated fibroblasts mediate cancer progression and remodel the tumouroid stroma. *Br J Cancer*. 2020;123:1178–90.
48. De P, Aske J, Dey N. Cancer-associated fibroblast functions as a road-block in cancer therapy. *Cancers (Basel)*. 2021;13:5246.
49. von Ahrens D, Bhagat TD, Nagrath D, Maitra A, Verma A. The role of stromal cancer-associated fibroblasts in pancreatic cancer. *J Hematol Oncol*. 2017;10:76.
50. Fang J, Xiao L, Joo KI, Liu Y, Zhang C, Liu S, et al. A potent immunotoxin targeting fibroblast activation protein for treatment of breast cancer in mice. *Int J Cancer*. 2016;138:1013–23.
51. Ozdemir BC, Pentcheva-Hoang T, Carstens JL, Zheng X, Wu CC, Simpson TR, et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell*. 2014;25:719–34.
52. Lo A, Li CP, Buza EL, Blomberg R, Govindaraju P, Avery D, et al. Fibroblast activation protein augments progression and metastasis of pancreatic ductal adenocarcinoma. *JCI Insight*. 2017;2:e92232.
53. Chen Y, McAndrews KM, Kalluri R. Clinical and therapeutic relevance of cancer-associated fibroblasts. *Nat Rev Clin Oncol*. 2021;18:792–804.
54. Affo S, Nair A, Brundu F, Ravichandra A, Bhattacharjee S, Matsuda M, et al. Promotion of cholangiocarcinoma growth by diverse cancer-associated fibroblast subpopulations. *Cancer Cell*. 2021;39:866–82.e11.
55. Ohlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvisé M, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med*. 2017;214:579–96.
56. Pereira BA, Vennin C, Papanicolaou M, Chambers CR, Herrmann D, Morton JP, et al. CAF subpopulations: a new reservoir of stromal targets in pancreatic cancer. *Trends Cancer*. 2019;5:724–41.
57. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JLL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet*. 2017;49:708–18.
58. Iida T, Mizutani Y, Esaki N, Ponik SM, Burkel BM, Weng L, et al. Pharmacologic conversion of cancer-associated fibroblasts from a protumor phenotype to an antitumor phenotype improves the sensitivity of pancreatic cancer to chemotherapeutics. *Oncogene*. 2022;41:2764–77.
59. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med*. 2018;24:1277–89.
60. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast heterogeneity and immunosuppressive environment in human breast cancer. *Cancer Cell*. 2018;33:463–79.e10.
61. Kieffer Y, Hocine HR, Gentric G, Pelon F, Bernard C, Bourachot B, et al. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Disco*. 2020;10:1330–51.
62. Pelon F, Bourachot B, Kieffer Y, Magagna I, Mermet-Meillon F, Bonnet I, et al. Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. *Nat Commun*. 2020;11:404.
63. Bernard V, Semaan A, Huang J, San Lucas FA, Mulu FC, Stephens BM, et al. Single-cell transcriptomics of pancreatic cancer precursors demonstrates epithelial and microenvironmental heterogeneity as an early event in neoplastic progression. *Clin Cancer Res*. 2019;25:2194–205.
64. Murakami M, Ernsting MJ, Undzys E, Holwell N, Foltz WD, Li SD. Docetaxel conjugate nanoparticles that target alpha-smooth muscle actin-expressing stromal cells suppress breast cancer metastasis. *Cancer Res*. 2013;73:4862–71.
65. Brunker P, Wartha K, Friess T, Grau-Richards S, Waldhauer I, Koller CF, et al. RG7386, a novel tetravalent FAP-DR5 antibody, effectively triggers FAP-dependent, avidity-driven DR5 hyperclustering and tumor cell apoptosis. *Mol Cancer Ther*. 2016;15:946–57.
66. Froeling FE, Feig C, Chelala C, Dobson R, Mein CE, Tuveson DA, et al. Retinoic acid-induced pancreatic stellate cell quiescence reduces paracrine Wnt-beta-catenin signaling to slow tumor progression. *Gastroenterology*. 2011;141:1486–97. 1497 e1–14
67. Hurwitz H, Uppal N, Wagner SA, Bendell JC, Beck JT, Wade SM 3rd, et al. Randomized, double-blind, phase II study of ruxolitinib or placebo in combination with capecitabine in patients with metastatic pancreatic cancer for whom therapy with gemcitabine has failed. *J Clin Oncol*. 2015;33:4039–47.
68. Pietras K, Pahler J, Bergers G, Hanahan D. Functions of paracrine PDGF signaling in the proangiogenic tumor stroma revealed by pharmacological targeting. *PLoS Med*. 2008;5:e19.

## AUTHOR CONTRIBUTIONS

QY: collect data and edit. YChu: draw the graph. XL: check the data. YS: check the data. YChe: draw the graph. JZ: discussion. JL: discussion. KL: concept and supervise. YG: design and revise.

## FUNDING

This work was supported by Natural Science Foundation of China (32100532), China Postdoctoral Science Foundation (2021M692936), the Central Plains Science and Technology Innovation Leading Talents (224200510015), and the Supporting Plan of Zhengzhou University Youth Innovation Team (32320442).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41417-023-00603-4>.

**Correspondence** and requests for materials should be addressed to Kangdong Liu or Yaping Guo.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023