



OPEN

Novel CAF-identifiers via transcriptomic and protein level analysis in HNSC patients

Nehanjali Dwivedi^{1,2}, Nidhi Shukla¹, K. M. Prathima³, Manjula Das¹ & Sujan K. Dhar⁴✉

Cancer-associated fibroblasts (CAFs), a prominent component of the tumor microenvironment, play an important role in tumor development, invasion, and drug resistance. The expression of distinct “CAF-markers” which separates CAFs from normal fibroblasts and epithelial cells, have traditionally been used to identify them. These commonly used CAF-markers have been reported to differ greatly across different CAF subpopulations, even within a cancer type. Using an unbiased -omic approach from public data and in-house RNAseq data from patient derived novel CAF cells, TIMP-1, SPARC, COL1A2, COL3A1 and COL1A1 were identified as potential CAF-markers by differential gene expression analysis using publicly available single cell sequencing data and in-house RNAseq data to distinguish CAF populations from tumor epithelia and normal oral fibroblasts. Experimental validation using qPCR and immunofluorescence revealed CAF-specific higher expression of TIMP-1 and COL1A2 as compared to other markers in 5 novel CAF cells, derived from patients of diverse gender, habits and different locations of head and neck squamous cell carcinoma (HNSC). Upon immunohistochemical (IHC) analysis of FFPE blocks however, COL1A2 showed better differential staining between tumor epithelia and tumor stroma. Similar data science driven approach utilizing single cell sequencing and RNAseq data from stabilized CAFs can be employed to identify CAF-markers in various cancers.

Abbreviations

ACTA2	Actin Alpha 2
AKT	Alpha serine/threonine-protein kinase
BSA	Bovine serum albumin
CAF	Cancer associated fibroblast
CAV1	Caveolin 1
cDNA	Complementary deoxyribonucleic acid
COL11A	Collagen type XI Alpha 1 chain
COL1A1	Collagen type I Alpha 1 chain
COL1A2	Collagen type I Alpha 2 chain
COL3A1	Collagen type III Alpha 1 chain
CRC	Colorectal cancer
CXCR4	C-X-C Motif chemokine receptor 4
DAPI	4',6-Diamidino-2-phenylindole
ddCT	Delta-delta Ct
DGE	Differential gene expression
DIO2	Iodothyronine Deiodinase 2
ECM	Extracellular matrix
EMT	Epithelial to mesenchymal transition
EpCAM	Epithelial cellular adhesion molecule
EREG	Epiregulin
FACS	Flourescence associated cell sorting
FAP	Fibroblast-activation protein
FBS	Fetal Bovine serum
FGF7	Fibroblast growth factor 7

¹Molecular Immunology, Mazumdar Shaw Medical Foundation, Narayana Health City, Bommasandra, Bangalore, Karnataka 560099, India. ²MAHE, Manipal 576104, India. ³Manipal Hospital, Miller's Road, Bangalore, Karnataka 560052, India. ⁴Computational Biology, Mazumdar Shaw Medical Foundation, Narayana Health City, Bommasandra, Bangalore, Karnataka 560099, India. ✉email: sujan.dhar@ms-mf.org

FSP1	Fibroblast-specific protein-1
HGF	Hepatocyte growth factor
HNSC	Head and neck squamous cell carcinoma
iCAFs	Inflammatory CAFs
IL-1 β	Interleukin 1 Beta
ITGA11	Integrin subunit Alpha 11
MAP1B	Microtubule associated protein 1B
MFAP5	Microfibril associated protein 5
MMP11	Matrix Metalloproteinase 11
myCAFs	Myofibroblastic CAFs
NCBI	National Center for Biotechnology Information.
NOX4	NADPH oxidase 4
OSCC	Oral squamous cell carcinoma
PBS	Phosphate buffer saline
PDAC	Pancreatic ductal adenocarcinoma
PDGFR α	Platelet derived growth factor receptor alpha
PDGFR β	Platelet Derived growth factor receptor beta
PDPN	Podoplanin
PI3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PLEKHA3	Pleckstrin homology domain containing A3
POSTN	Periostin
qPCR	Quantitative polymerase chain reaction
RIC8B	Resistance to inhibitors of cholinesterase 8 Homolog B
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SDF-1	Stromal cell-derived factor-1
SLCC	Stem like cancer cells
SMA	Smooth muscle actin
SMTN	Smoothelin
SPARC	Secreted protein acidic and cysteine rich
STRING	Search tool for the retrieval of interacting genes/proteins
TAGLN	Transgelin
TCGA	The cancer genome atlas
TGF	Transforming growth factor
TIMP-1	Tissue inhibitor of metalloproteinases 1
TME	Tumor Microenvironment
TYW5	TRNA-YW synthesizing protein 5
VIM	Vimentin

Tumor microenvironment (TME) consists of dynamic cell populations such as fibroblasts, immune cells, inflammatory cells, adipocytes, endothelial cells, mesenchymal cells, and extracellular matrix (ECM)^{1–3}. Cross-talks between CAF and TME signal the tumor to survive and advance^{4,5}. During cancer development, there is a prominent expansion of quiescent fibroblasts residing in the host tissue in response to the developing neoplasm⁶. CAFs can be assertively defined as cells with an elongated morphology and negative for epithelial (EpCAM), endothelial (CD31) and leukocyte (CD45) markers^{7,8}.

Multiple markers have been described to define CAFs for years. For example, Fibroblast Activation Protein (FAP) and Smooth Muscle Actin (alpha SMA/ACTA2) have been employed as a marker of activated CAFs in colorectal and breast cancer respectively^{8–10}, because of their high expression in the tumor stroma. However, a number of investigations have revealed that epithelial cells undergoing epithelial-mesenchymal transition (EMT) also exhibit higher levels of FAP¹¹, whereas, alpha SMA displays fluctuating expression amongst various CAF subtypes^{12,13}. Although not significantly upregulated in CAF populations, due to a more stable expression, not sensitive to environmental variables like hypoxia, PDGFR α and PDGFR β have been considered as CAF-markers in breast cancer¹⁴. Expression of FSP1¹⁵, Transgelin (TAGLN) and Periostin (POSTN)¹⁶ in CAFs vary among CAF-subtypes as reported in colorectal cancer¹³. Podoplanin (PDPN), a membrane-bound marker that has been utilised to identify pro-tumorigenic fibroblast subpopulations lacks specificity since it is also expressed in epithelial tumor cells and inflammatory macrophages^{17,18}. Integrin 11 (ITGA11) has been identified to be upregulated in CAFs from non-small cell lung cancer^{19,20}, though demonstrated to be present in a variety of tumor cell lines and sensitive to hypoxia²¹ and TGF- signalling²², playing a dual role of an inducer and antagonist of specific CAF subtypes²³. Though described a few years back, Microfibril Associated Protein 5 (MFAP5) and Collagen Type XI Alpha I Chain (COL11A1) have not yet found popularity among the users. Further research is therefore necessary to characterise these markers' activity and expression in various CAF subtypes^{24,25}.

Not only in identifying them, CAF-markers can be utilized to classify CAFs into various subtypes in different cancers^{26,27}, giving insight about their functional diversity^{28–30}. For example, Interleukin-6 (IL6) has been used in pancreatic ductal adenocarcinoma (PDAC) to describe inflammatory CAFs (iCAFs). Similarly, in breast cancer, podoplanin and S100A³¹, in colorectal cancer (CRC), ACTA2-TAGLN-PDGFA, and DCN-MMP2-COL1A2¹³ have been used.

In HNSC specifically, Vimentin and ACTA2 are the most frequently used CAF-markers. Vimentin's ubiquitous expression in the whole fibroblast population as well as multiple other cell types, including macrophages,

adipocytes, and the cells undergoing EMT severely limits its utility as a CAF-specific marker^{32–34}. Similarly, ACTA2 has inconsistent expression across different subtypes of CAFs^{12,13}. Fibronectin 1 and THY1 have been known to recognize myofibroblasts and fibroblasts of the reticular lineage respectively³⁵. Two other markers—SHH and GLI1 have also been reported, however, without any exclusivity to CAFs. Many authors used highly sensitive single cell sequencing approach to classify CAFs based on differential expression of several markers in HNSC (supplementary table I). Puram et.al, identified immediate early response genes (e.g. JUN, FOS), mesenchymal markers (e.g. VIM, THY1), ligands and receptors (e.g. FGF7, TGFBR2/3), and ECM proteins (e.g. MMP11, CAV1)³⁶ to classify CAFs into CAF1 and CAF2 subtypes. Wang et.al reported FTH1, TM4SF1, SLC16A3 and IER3 as markers for prognosis-related subtype of CAFs³⁷. However, their functional utility and confirmation at the protein expression level have not been explored yet^{36,37}.

Since many positive CAF-markers utilised lack of specificity, negative selection is commonly used to filter out the cell types that are frequently present in tumor microenvironment. In order to distinguish between epithelial and smooth muscle cells, markers like epithelium cell adhesion molecule (EpCAM) and Smoothelin (SMTN) are frequently utilised^{33,38,39}. Leukocytes and endothelial cells are typically excluded using CD45, CD34, and CD11b⁴⁰.

Though not directly for the clinic, CAF-markers are very important for research, to differentiate them from the tumor cells to study the mechanism of action of the microenvironment in cancer progression. Multiple methods, as in Supplementary Table II, have been employed to arrive at a set of CAF-markers to distinguish them from normal fibroblast and epithelial cells. Since the development of sequencing technologies, availability of multi-omics data has transformed medicine and biology by allowing integrated system-level techniques to be developed⁴¹. Researchers may compare, compute, and analyse the expression patterns of multiple genes across samples and cell types using gene expression analyses methods such as microarray, mRNA sequencing, and single cell sequencing that are more robust as compared to the gold standard immunohistochemistry (IHC) methodology. However, experimental validation of the identified markers on pure primary cultures becomes indispensable to support the omics data in localization studies. Thus, it is important to consider all these aspects to analyse the markers before reporting.

The aim of the study was therefore, to explore reliable CAF-marker(s) via multi-omics analysis followed by experimental validation in HNSC, the sixth most prevalent cancer in the world, affecting the mucosal epithelium of the larynx, oral cavity, and pharynx⁴².

The present study employed an integrative analysis of publicly available single-cell, bulk gene expression data and in-house RNAseq data to identify the CAF-markers, which were then validated on patient derived cells by qPCR, immunofluorescence and on HNSC-FFPE blocks by IHC.

Results

Identification of differentially expressed genes between CAF and epithelial cells. Differential gene expression analysis between the established CAFs MhCT012-F or MhCT08-F and epithelial MhCT08-E and MhCT12-E revealed 313 significantly upregulated genes (Fig. 1a), whereas single-cell sequencing data

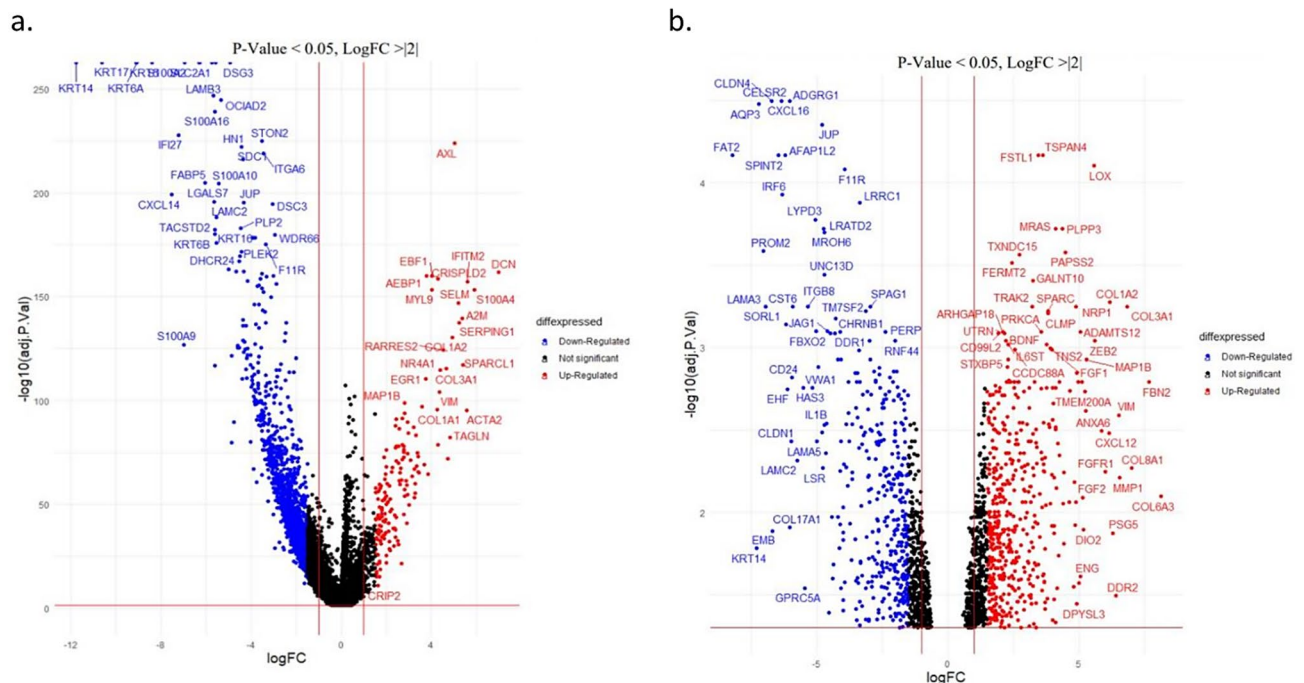


Figure 1. Differentially expressed genes between CAF and epithelial cells. (a) from in-house RNAseq data from patient derived CAF (MhCT08-F and MhCT12-F) and epithelial cell lines (MhCT08-E and MhCT12-E) (b) Single cell sequencing data from GSE10332237³⁶. Overexpressed genes ($\log\text{FoldChange} > |2|$) and adjusted p -value < 0.05 in fibroblasts and epithelial cells are shown as red and blue dots respectively.

from non-lymphatic fibroblast and epithelial cells obtained from GSE103322 [21] revealed 97 genes that were up-regulated in CAFs as compared to epithelial cells (Fig. 1b). Twenty two genes were concordant between the two differential expression subsets (Supplementary Table III) that shows eight genes having significant p-values between CAFs vs normal fibroblasts.

Selection of putative CAF-markers. Student's t-test with RNAseq counts from dataset GSE135975 indicated only seven (DIO2, MAP1B, COL1A1, SPARC, COL1A2, COL3A1 and TIMP-1) out of the 22 concordant genes to be upregulated in CAFs over normal fibroblasts. One gene IFITM2 was found to be downregulated (Data not shown). Further using the criterion that a CAF-marker should be over-expressed in tumor samples, by analysis of TCGA HNSC dataset, we selected five genes viz. COL1A1, SPARC, COL1A2, COL3A1 and TIMP-1, where DIO2 and IFITM2 were downregulated and MAP1B did not show much difference (Fig. 2a).

Validation of putative CAF-marker transcripts. In qPCR analysis of the potential markers, TIMP-1 and COL1A2 were significantly overexpressed ($p < 0.05$) in 5 fibroblasts as compared to epithelial cultures, (Fig. 2b). SPARC, followed by COL1A1, had the lowest mRNA expression among the 5 CAF-markers shortlisted. COL3A1 was found to be significantly overexpressed in four out of five fibroblasts; however in MhCL03-F it was expressed meekly. TIMP-1 and COL1A2 were therefore chosen for further analysis and validation.

Protein Expression of CAF-markers in cell lines. Protein expression of TIMP-1 and COL1A2 was detected in the fibroblasts but absent in the corresponding epithelial cultures (Fig. 2c). The cells were also qualified by staining with FSP-1, an established fibroblast marker as positive control along with two epithelial specific markers—PanCK and EpCAM as negative control for the CAFs. The fibroblast cells stained only with FSP-1, TIMP-1 and COL1A2 while the epithelial cells stained only with the epithelial specific markers—PanCK and EpCAM (Fig. 2c). Compared to FSP-1, TIMP-1 and COL1A2 displayed uniform staining across the 5 CAFs.

COL1A2 is the best CAF-marker. Analysis of slides from paraffin embedded oral squamous cell carcinoma blocks using COL1A2 and TIMP-1 staining revealed that TIMP-1, stained meekly (Fig. 2d) as opposed to COL1A2 (Fig. 2d), which showed a strong and significant differential staining in the stromal compartment of oral cancer patient sample, thus winning the race over TIMP-1.

Discussion

In this study initiating from a data science driven unbiased genome-wide approach, from ~400 genes, COL1A2 and TIMP-1 were experimentally validated at the protein level to be the general CAF-markers in HNSC.

Earlier used methods for identification of CAF-markers majorly constituted of immunofluorescence, immunohistochemistry or flow cytometry based individual pre-selected-gene-centric techniques. More recently, single cell sequencing based analyses were performed to identify potential CAF signature as a biomarker for the prognosis of HNSC patients, or for the classification of different CAF subtypes; however in these studies, protein level validation of the identified markers was not performed^{36,37}. The novel approach described in this study for the identification of CAF-markers is unbiased towards any particular gene, dealing with a large number of in-house and public domain data obtained from completely independent experiments, finally converging at validation at the protein level. This kind of analysis was possible due to the availability of a set of novel CAF cells^{43,44}. To our knowledge, the study is the first report of its kind to utilize an integrative way of identifying CAF-markers at mRNA and protein level. The present study identified three collagen markers along with TIMP-1 and SPARC, compared among themselves and with the existing marker, FSP-1 to finally choose COL1A2 followed by TIMP-1.

Collagens are a protein family that helps to build and support numerous structural tissues, including cartilage, bone, tendon, skin, and the white area of the eye (the sclera)⁴⁵. In a study by Misawa et al.⁴⁶, hypermethylation of the COL1A2 promoter occurred with a high frequency and the expression levels of COL1A2 was downregulated in HNSC cell lines, which are epithelial in nature. Collagen has been reported to induce a more aggressive phenotype in HNSC via DDR1⁴⁷. COL1A2 has also been reported extensively to be downregulated in melanoma^{48,49} and bladder cancer⁵⁰, while being upregulated in colorectal cancer⁵¹, gastric cancer^{52,53}, breast cancer⁵⁴ and medulloblastoma⁵⁵. COL1A2 gene codes for a protein that is part of a larger molecule known as type I collagen. The ECM's primary structural component is type I collagen, whose modification has been linked to stromal invasion and epithelial carcinogenesis in several cancers. Mesenchymal fibroblast cells have been characterized by the presence of COL1A2 promoter gene⁵⁶. COL1A1 has been identified as a potential CAF-marker in murine cells⁵⁷. In humans, COL1A2 along with MMP2 and FAP as CAF identifier has been reported in colorectal and pancreatic cancer^{13,58,59}. The present study identified three collagen family markers, viz—COL1A2, COL3A1 and COL1A1 as putative CAF identifiers, however, COL1A1 and COL3A1 displayed irregular expression in the five CAFs under study. COL1A2 on the other hand showed a uniform expression among CAFs originating from diverse sources like larynx (MhCL03-F), buccal mucosa (MhCB05-F), upper alveolus (MhCA04-F) and tongue (MhCT08-F and MhCT12-F)⁴⁴ and hence was taken forward among others. Its function as a CAF factor is yet to be ascertained.

TIMP-1 has been associated with the tumor aggressive phenotype in various cancers^{60–62}. Compared to FSP-1, the popular CAF-identifier, both TIMP-1 and COL1A2 displayed uniform and brighter staining across the 5 CAFs tested. However, TIMP-1 did not stain well in IHC on FFPE block, may be due to the difference in TIMP-1 expression across different stages and location of oral cancers⁶².

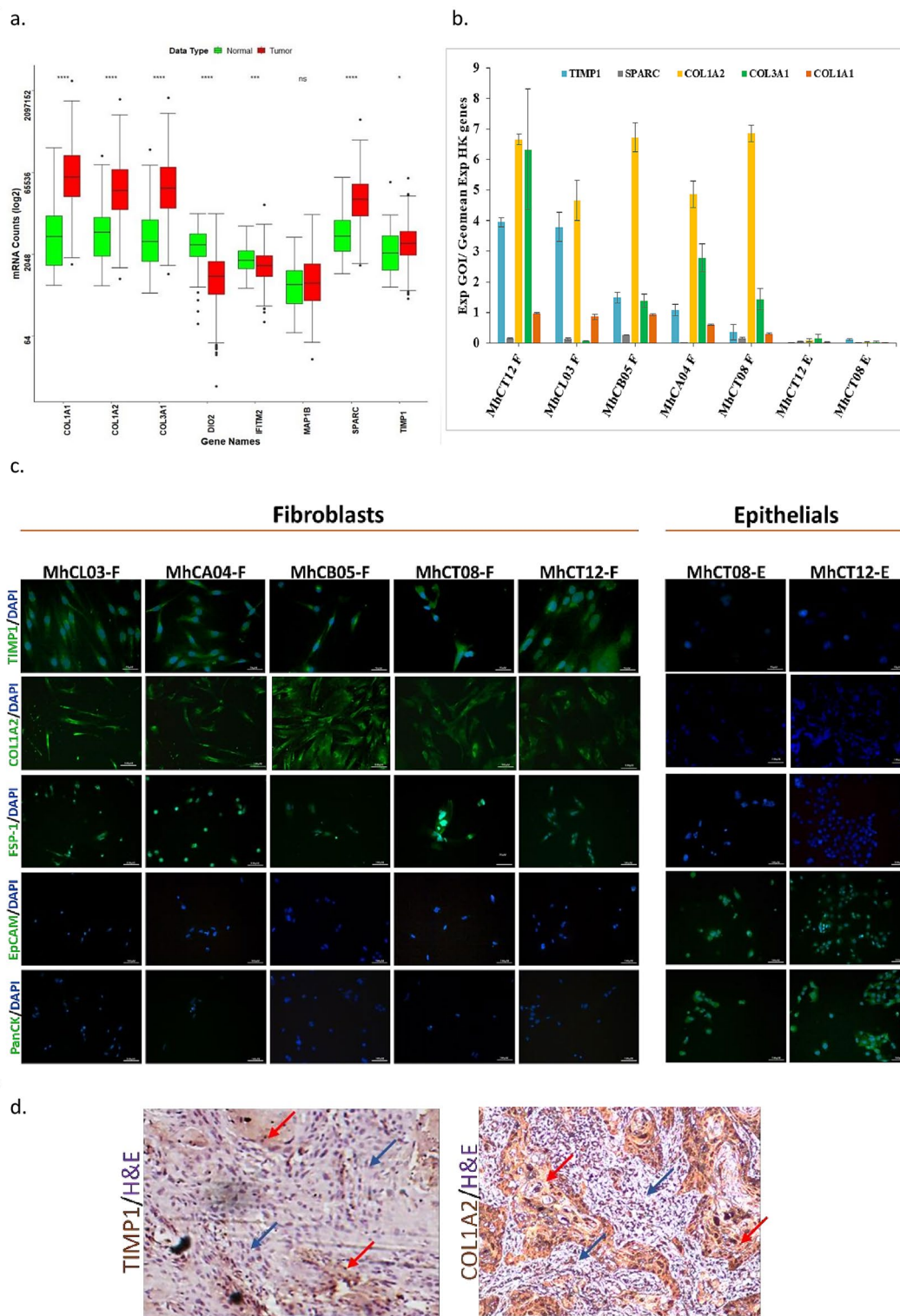


Figure 2. Validation of expression of the candidate CAF markers. (a) in TCGA HNSC dataset (**p*-value < 0.05; ns—non significant), (b) on CAF cells by qPCR; (c) by immunofluorescence (40X images) showing staining of TIMP1, COL1A2, FSP-1, EpCAM and PanCK on the indicated cell types; (d) Immuno-histochemical validation of TIMP1 and COL1A2 in oral cancer FFPE blocks. Red arrows in the representative image indicate the stromal compartment, while the blue arrows indicate the epithelial compartment of the tissue (images taken at 10X).

Though majority of solid tumors, including HNSC, exhibit SPARC expression^{63,64}, is linked to EMT signalling⁶⁵ and unfavourable clinico-pathological characteristics⁶⁶, it did not qualify as a dependable marker in the present study, due to its overall lower expression across varieties of CAFs tested.

The study is limited by the fact that all the molecules shortlisted are secretory in nature, hence cannot be used for sorting CAFs from tumors. However, it can be used for their extensive characterization, even by IHC. In the present study we have used golgi-plug to arrest the secretion in the established CAF culture to detect the presence of the putative CAF-markers in cells via immunofluorescence. Curiously, the entire list of the markers identified by the data science driven approach presented here are secretory in nature indicating that the over expressing CAF-factors reach out more effectively to the tumor as well as the other cells of the microenvironment by spreading through the matrix. The CAF-markers open the door to future selective and therapeutically viable targeting of the tumor cells, enabling the establishment of a successful oral cancer therapy regimen.

Conclusions

The current study shows that COL1A2, followed by TIMP-1, are the most potent CAF-markers found and verified at the transcriptome and protein level across multiple CAF types, differentiating CAF from tumour epithelia and normal fibroblast in HNSC patient samples.

Methods

The overall study workflow (Fig. 3) begins with patient-derived cell culture maintenance, sequencing, data collection, and ends with final candidate markers using various levels of filters like (i) comparison of CAF and normal fibroblast transcriptome (ii) normal versus tumor transcriptome from a public database, and a final filter of (iii) validation in patient derived samples and FFPE blocks. All methods were performed in accordance with the relevant guidelines and regulations.

Cell culture. MhCT12-E, MhCT12-F, MhCT08-F, MhCT08-E^{43,44} MhCL03-F, MhCB05-F and MhCA04-F⁴⁴ were maintained RPMI- 1640 media (#AT222A; HiMedia) with 20% Fetal Bovine Serum (FBS; 10,270–106; Gibco), 1X penicillin- streptomycin (#15,140,122; Gibco) and subcultured every 48 h.

mRNA sequencing of patient derived cell lines. Sequencing of mRNA from the autologous pair—MhCT08-E, MhCT08-F, MhCT12-E and MhCT12-F was performed using Illumina HiSeq series X sequencer using 150 bp paired-end chemistry. Reads obtained from the sequencer were quality-checked and filtered using fastp⁶⁷ and subsequently aligned using STAR⁶⁸ pipeline.

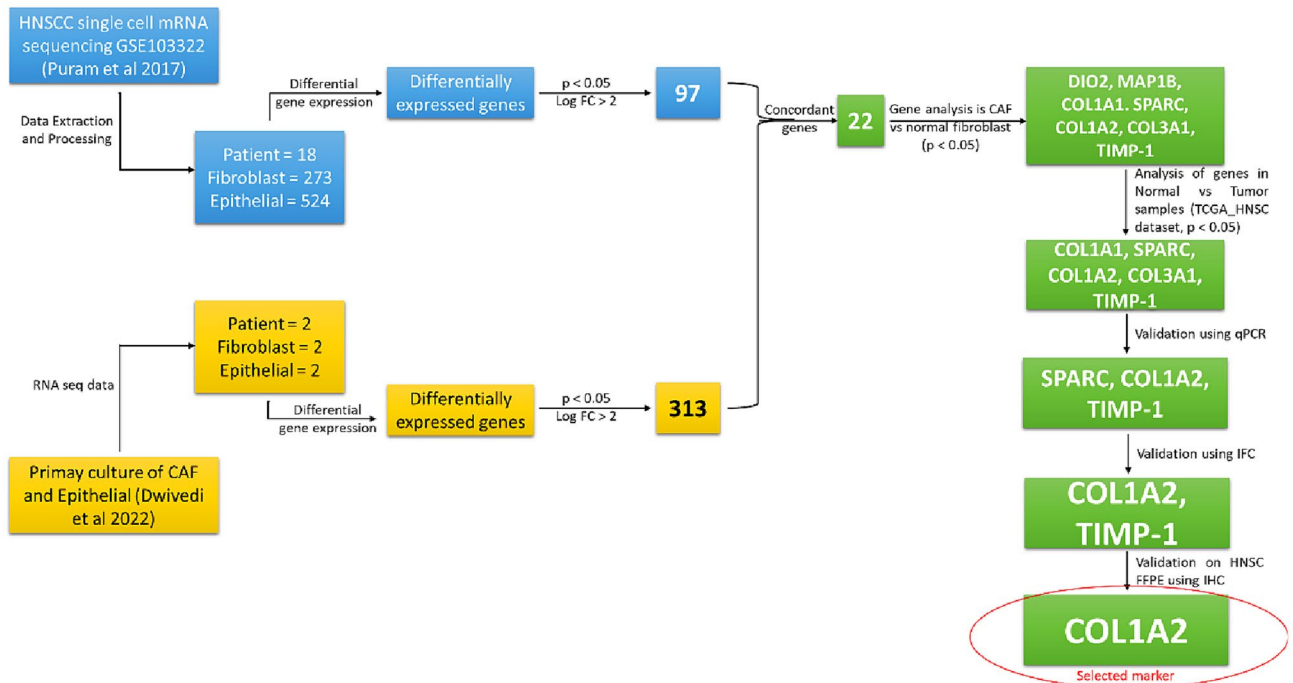


Figure 3. Workflow for identification of putative CAF marker candidates. Data analysis was performed from public databases and in-house RNA-seq data to arrive at a concordant set of differentially expressed genes between epithelial and CAFs which were further selected over normal fibroblasts (Lai et al. 2019), and adjacent normal in TCGA HNSC dataset. The chosen markers were then validated in patient derived CAF cells using qPCR and at protein level by fluorescence microscopy (IFC) and immunohistochemistry (IHC).

Acquisition of single cell sequencing data. Search with the terms (((Head and Neck Squamous Cell Carcinoma OR "ORAL CANCER" OR HNSC) AND (single-cell OR "single cell"))) with organism filter for homo sapiens in Gene Expression Omnibus (GEO)⁶⁹ yielded three series with accession codes GSE103322, GSE135975 and GSE163872 respectively. GSE103322³⁶, single-cell gene expression processed data and GSE135975⁴⁷ RNAseq FPKM (fragments per kilobase of transcript per million mapped reads) values were retrieved from GEO portal. Data from GSE163872 was not considered since the clinical details of the samples were not available.

Data extraction and processing. For single cell data twenty human gene expression datasets were discovered using primary search terms from the GEO database. GSE103322 was selected appropriately after pruning the data which consisted of sequenced transcriptomes of roughly 6,000 single cells from 18 HNSC patients containing 273 and 524 non-lymphatic fibroblast and epithelial cells. In parallel, RNAseq data from two novel autologous pairs of CAFs and epithelial cells were also analysed.

Differential gene expression (DGE) analysis of these two datasets yielded a set of CAF indicators that were over-expressed in CAFs relative to epithelial cells, which were further assessed for expression in normal fibroblasts and CAFs using RNAseq data from GSE135975⁴⁷, containing 8 patient CAFs and 3 normal fibroblast samples.

Selection of putative CAF-markers. Over-expressed genes in fibroblasts compared to epithelial cells in dataset GSE103322 were enumerated using the limma R package. Overexpressed genes with Log Fold Change value ≥ 2 and Benjamini–Hochberg adjusted p -value ≤ 0.05 in fibroblasts from the in-house cell line sequencing data and from dataset GSE103322 were combined to identify genes concordant in both differential expression sets. The concordant genes were further filtered for differential expression between cancer-associated and normal fibroblasts (in dataset GSE135975) using t-test to shortlist genes with over-expression in cancer-associated fibroblasts. Expression of these shortlisted genes were assessed in the TCGA HNSC dataset using RNAseq expression values acquired from Broad Institute Firehose portal⁷⁰ to obtain set of genes overexpressed in tumor compared to its normal counterpart.

RNA isolation, cDNA conversion and qPCR. One million cells were re-suspended and incubated for 10 min on ice in 1 mL of TRIzol⁷¹ (#15,596,018; Invitrogen) followed by addition of 200 μ L of chloroform (#496,189; Sigma) and centrifugation at 13,000 g for 15 min at 4 °C. The aqueous layer was transferred into a fresh tube containing equal volume of 100% isopropanol (#DB4DF64078; Merck) without disturbing either the middle interface or the lower organic phase. The tube was gently invert-mixed and incubated at room temperature for 10 min followed by centrifugation at 13,000 g for 10 min at 4 °C. The supernatant was discarded and the pellet was washed with 75% ice cold ethanol (#MB228; HiMedia) followed by centrifugation at 7500 g for 5 min at 4 °C. Air-dried pellet was then dissolved in nuclease free water for further processing. For cDNA conversion, 1 μ g of total RNA was used in a 20 μ L reaction volume with AMV Reverse Transcriptase enzyme (#M0277S; NEB). With a set of three reference genes — TYW5, RIC8B and PLEKHA3⁷², qPCR was performed in triplicates on a Roche LightCycler 480 II instrument using KAPA SyBr green Universal kit (#KK4601; Kapa) in a total of 5 μ L reaction volume with 1:5 diluted cDNA and primers (100 nM each). The reaction mix was pre-incubated at 95 °C for 10 s followed by 45 cycles of amplification (95 °C — 1 s; 95 °C — 10 s; 60 °C — 15 s; and 72 °C — 15 s).

RNA was quantified using Qubit RNA Assay BR (#Q10210; Invitrogen) for sequencing on Illumina HiSeq X instrument to generate 30 M, 150 bp paired end reads.

Design of primers. Primers, with amplicon sizes of 100–150 base pairs and melting temperatures of 60–65 °C were designed using Primer Bank Harvard and synthesized by Eurofins, Bangalore, India (primers used in the study are listed below).

Primers used for qPCR.

S. No	Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
1	TYW5	CAGCATCAAGAGCTGCACAAA	TGTGTAGGACCATTTCGTCGTG
2	PLEKHA3	ACTGTGACCTCTTAATGCAGC	CTCAAGCGTTGTGATGAATGTG
3	RIC8B	ATAGTGTTC AACAGTCAGATGGC	GCAAGCGCAAGTCAAAGCA
4	TIMP1	AGAGTGTCTGCGGATACTTCC	CCAACAGTGTAGGCTCTGGTGTG
5	SPARC	CCCATTGGCGAGTTTGAGAAG	CAAGGCCCGATGTAGTCCA
6	COL1A2	GGCCCTCAAGGTTTCCAAGG	CACCCTGTGGTCCAACAACCTC
7	COL3A1	GCCAAATATGTGTCTGTGACTCA	GGGCGAGTAGGAGCAGTTG
8	COL1A1	GTGCGATGACGTGATCTGTGA	CGGTGGTTTCTTGGTCGGT

qPCR data analysis. For efficiency check a two-fold five-point dilution of Cal27-Parental cDNA was used as template. Thermo primer efficiency calculator⁷³ was used to calculate the efficiency of primers as $E = 10^{-1/\text{slope}}$. For data analysis, Quantification cycle (Cq) values thus obtained were subtracted by geometric mean of house-keeping genes (HK genes) to obtain $\Delta Cq = Cq_{\text{sample}} - \text{Geom mean } Cq_{\text{HK genes}}$ from which the relative expression was calculated as $E^{-\Delta Cq}$ for each replicate.

Immunofluorescence. Immunofluorescence was performed as detailed previously^{43,44} and in accordance to the protocol by EuroMabNet⁷⁴. Briefly, for analysis by fluorescence, 20,000 cells in 100 μ L of medium were grown on coverslips and allowed to attach overnight followed by fixing with 100% ice cold methanol (#AS059; HiMedia) for 15 min at -20°C and blocking in 1% Bovine Serum Albumin (BSA; # TC194; HiMedia) and 0.3% triton-X 100 (#10,655; Fischer Scientific) in Phosphate Buffer Saline (PBS; #10,010–023; Gibco) for 1 h at room temperature. Cells were then incubated with anti-TIMP-1 (MAB970-SP; R&D systems, used at 0.5 mg/mL) and anti-COL1A2 (#GTX102996; GeneTex; used at 0.5 mg/mL) along with FSP-1 (#F4771; Sigma; used at 2 mg/mL) and epithelial specific markers PanCK (#4545; CST; used at 1:200 dilution) and EpCAM (#AN820; BioGenex; working solution) as primary antibodies for 3 h at room temperature, followed by treatment with anti-mouse Alexa488-conjugated secondary antibody (#A11029; Invitrogen; 1:200) for 1 h at room temperature in dark. The processed coverslips were mounted on slides with 4',6-diamidino-2-phenylindole (DAPI) (#F6057; HiMedia) histology mount and visualized under Zeiss Scope A1 fluorescent microscope. All the intermittent washes were given thrice with 1X PBS.

Immunohistochemistry. FFPE tumor blocks (n=10) containing both the tumor stroma and tumor epithelia from oral cancer patients with lesion of buccal mucosa were used for immunohistochemical staining of COL1A2 and TIMP-1. Briefly, slides were deparaffinized for 30 min at 70 °C and passed through xylene, graded alcohol, and deionized water. Slides were then rinsed in low pH antigen retrieval buffer, and microwaved at 60% power. After cooling, the slides were incubated in 3% hydrogen peroxide (#1,072,090,500; Supelco) for 10 min followed by rinsing with TBST and blocking for 30 min using 3% BSA and incubated with respective antibodies at room temperature for 60 min followed by rinsing with TBST and incubation with DAKO kit (#K5007; Dako; Agilent Technologies) for 30 min. Signal was detected using DAB followed by counterstain with hematoxylin solution. Slides were washed with xylene before being mounted using DPX mountant (#DAL1025; Qualigen Fine Chemicals TM; ThermoFisher Scientific) and observed under Nikon bright field microscope.

Ethical approval. The present study was approved [NHH/MEC-CL-EA-2015-405(A)] by the ethics committee of Narayana Health (Bangalore, India). Informed consent for the study was obtained from all patients.

Data availability

The datasets generated and/or analysed during the current study are available in GEO repository with the accession number—GSE233043. All the materials used have been enlisted in supplementary table IV.

Received: 2 May 2023; Accepted: 18 August 2023

Published online: 25 August 2023

References

- Meurette, O. & Mehlen, P. Notch signaling in the tumor microenvironment. *Cancer Cell* **34**, 536–548 (2018).
- Vuong, L., Kotecha, R. R., Voss, M. H. & Hakimi, A. A. Tumor microenvironment dynamics in clear-cell renal cell carcinoma. *Cancer Discov.* **9**, 1349–1357 (2019).
- Vitale, I., Manic, G., Coussens, L. M., Kroemer, G. & Galluzzi, L. Macrophages and metabolism in the tumor microenvironment. *Cell Metab.* **30**, 36–50 (2019).
- Pienta, K. J., McGregor, N., Axelrod, R. & Axelrod, D. E. Ecological therapy for cancer: Defining tumors using an ecosystem paradigm suggests new opportunities for novel cancer treatments. *Transl. Oncol.* **1**, 158–164 (2008).
- Kwa, M. Q., Herum, K. M. & Brakebusch, C. Cancer-associated fibroblasts: How do they contribute to metastasis?. *Clin. Exp. Metastasis* **36**, 71–86 (2019).
- Raz, Y. *et al.* Bone marrow-derived fibroblasts are a functionally distinct stromal cell population in breast cancer. *J. Exp. Med.* **215**, 3075–3093 (2018).
- Sahai, E. *et al.* A framework for advancing our understanding of cancer-associated fibroblasts. *Nat. Rev. Cancer* **20**, 174–186 (2020).
- Han, C., Liu, T. & Yin, R. Biomarkers for cancer-associated fibroblasts. *Biomark. Res.* **8**, 1–8 (2020).
- Nishishita, R. *et al.* Expression of cancer-associated fibroblast markers in advanced colorectal cancer. *Oncol. Lett.* **15**, 6195 (2018).
- Muchlińska, A. *et al.* Alpha-smooth muscle actin-positive cancer-associated fibroblasts secreting osteopontin promote growth of luminal breast cancer. *Cell. Mol. Biol. Lett.* **27**, 1–14 (2022).
- Kahounová, Z. *et al.* The fibroblast surface markers FAP, anti-fibroblast, and FSP are expressed by cells of epithelial origin and may be altered during epithelial-to-mesenchymal transition. *Cytometry. A* **93**, 941–951 (2018).
- Latif, N., Sarathchandra, P., Chester, A. H. & Yacoub, M. H. Expression of smooth muscle cell markers and co-activators in calcified aortic valves. *Eur. Heart J.* **36**, 1335–1345 (2015).
- Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).
- Madsen, C. D. *et al.* Hypoxia and loss of PHD2 inactivate stromal fibroblasts to decrease tumour stiffness and metastasis. *EMBO Rep.* **16**, 1394–1408 (2015).
- Zeisberg, E. M., Potenta, S., Xie, L., Zeisberg, M. & Kalluri, R. Discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts. *Cancer Res.* **67**, 10123–10128 (2007).
- Planche, A. *et al.* Identification of prognostic molecular features in the reactive stroma of human breast and prostate cancer. *PLoS ONE* **6**, e18640 (2011).
- Kerrigan, A. M. *et al.* Podoplanin-expressing inflammatory macrophages activate murine platelets via CLEC-2. *J. Thromb. Haemost.* **10**, 484–486 (2012).
- Atsumi, N. *et al.* Podoplanin, a novel marker of tumor-initiating cells in human squamous cell carcinoma A431. *Biochem. Biophys. Res. Commun.* **373**, 36–41 (2008).
- Navab, R. *et al.* Integrin α 11 β 1 regulates cancer stromal stiffness and promotes tumorigenicity and metastasis in non-small cell lung cancer. *Oncogene* **35**, 1899–1908 (2016).
- Zeltz, C., Lu, N. & Gullberg, D. Integrin α 11 β 1: A major collagen receptor on fibroblastic cells. *Adv. Exp. Med. Biol.* **819**, 73–83 (2014).

21. Ju, J. A. *et al.* Hypoxia selectively enhances integrin $\alpha 5 \beta 1$ receptor expression in breast cancer to promote metastasis. *Mol. Cancer Res.* **15**, 723–734 (2017).
22. Grella, A., Kole, D., Holmes, W. & Dominko, T. FGF2 overrides TGF β 1-driven integrin ITGA11 expression in human dermal fibroblasts. *J. Cell. Biochem.* **117**, 1000–1008 (2016).
23. Biffi, G. *et al.* IL1-induced JAK/STAT signaling is antagonized by TGF β to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Discov.* **9**, 282–301 (2019).
24. Leung, C. S. *et al.* Calcium-dependent FAK/CREB/TNNC1 signalling mediates the effect of stromal MFAP5 on ovarian cancer metastatic potential. *Nat. Commun.* **5**, 5092 (2014).
25. Jia, D. *et al.* A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets. *Cancer Lett.* **382**, 203–214 (2016).
26. Knops, A. M. *et al.* Cancer-associated fibroblast density, prognostic characteristics, and recurrence in head and neck squamous cell carcinoma: A meta-analysis. *Front. Oncol.* **10**, 565306 (2020).
27. Custódio, M., Biddle, A. & Tavassoli, M. Portrait of a CAF: The story of cancer-associated fibroblasts in head and neck cancer. *Oral Oncol.* **110**, 104972 (2020).
28. Ping, Q. *et al.* Cancer-associated fibroblasts: Overview, progress, challenges, and directions. *Cancer Gene Therapy.* **28**(9), 984–999 (2021).
29. Kang, S. H. *et al.* Cancer-associated fibroblast subgroups showing differential promoting effect on HNSCC progression. *Cancers (Basel).* **13**, 654 (2021).
30. Galbo, P. M., Zang, X. & Zheng, D. Molecular features of cancer-associated fibroblast subtypes and their implication on cancer pathogenesis, prognosis, and immunotherapy resistance. *Clin. Cancer Res.* **27**, 2636–2647 (2021).
31. Friedman, G. *et al.* Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4+ and PDPN+ CAFs to clinical outcome. *Nat. Cancer* **1**(7), 692–708 (2020).
32. Gascard, P. & Tlsty, T. D. Carcinoma-associated fibroblasts: Orchestrating the composition of malignancy. *Genes Dev.* **30**, 1002–1019 (2016).
33. Hsia, L. T. *et al.* Myofibroblasts are distinguished from activated skin fibroblasts by the expression of AOC3 and other associated markers. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2162–E2171 (2016).
34. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119**, 1420 (2009).
35. Liu, T., Zhou, L., Li, D., Andl, T. & Zhang, Y. Cancer-associated fibroblasts build and secure the tumor microenvironment. *Front. Cell Dev. Biol.* **7**, 60 (2019).
36. Puram, S. V. *et al.* Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).
37. Yang, Y. *et al.* Integrated single-cell and bulk RNA sequencing analyses reveal a prognostic signature of cancer-associated fibroblasts in head and neck squamous cell carcinoma. *Front. Genet.* **13**, 1028469 (2022).
38. Berdiel-Acer, M. *et al.* Differences between CAFs and their paired NCF from adjacent colonic mucosa reveal functional heterogeneity of CAFs, providing prognostic information. *Mol. Oncol.* **8**, 1290–1305 (2014).
39. Costa, A. *et al.* Fibroblast heterogeneity and immunosuppressive environment in human breast cancer. *Cancer Cell* **33**, 463–479. e10 (2018).
40. Takahashi, H. *et al.* Cancer-associated fibroblasts promote an immunosuppressive microenvironment through the induction and accumulation of protumoral macrophages. *Oncotarget* **8**, 8633–8647 (2017).
41. Subramanian, I., Verma, S., Kumar, S., Jere, A. & Anamika, K. Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights* **14**, 1177932219899051 (2020).
42. Duvvuri, U. & Myers, J. N. In brief. *Curr. Probl. Surg.* **46**, 114–117 (2009).
43. Dwivedi, N. *et al.* Establishment and characterization of novel autologous pair cell lines from two Indian non-habitual tongue carcinoma patients. *Oncol. Rep.* **48**, 1–12 (2022).
44. Dwivedi, N. *et al.* Establishment and characterization of three novel CAF cell lines from HNSCC patients. *bioRxiv* <https://doi.org/10.1101/2023.01.08.523131> (2023).
45. COL1A2 gene: MedlinePlus Genetics. <https://medlineplus.gov/genetics/gene/col1a2/>.
46. Misawa, K. *et al.* Hypermethylation of collagen $\alpha 2$ (I) gene (COL1A2) is an independent predictor of survival in head and neck cancer. *Cancer Biomark.* **10**, 135–144 (2011).
47. Lai, S. L. *et al.* Collagen induces a more proliferative, migratory and chemoresistant phenotype in head and neck cancer via DDR1. *Cancers (Basel)* **11**, 1766 (2019).
48. Bonazzi, V. F. *et al.* Cross-platform array screening identifies COL1A2, THBS1, TNFRSF10D and UCHL1 as genes frequently silenced by methylation in melanoma. *PLoS One* **6**, e26121 (2011).
49. Koga, Y. *et al.* Genome-wide screen of promoter methylation identifies novel markers in melanoma. *Genome Res.* **19**, 1462–1470 (2009).
50. Mori, K. *et al.* CpG hypermethylation of collagen type I alpha 2 contributes to proliferation and migration activity of human bladder cancer. *Int. J. Oncol.* **34**, 1593–1602 (2009).
51. Zou, X. *et al.* Up-regulation of type I collagen during tumorigenesis of colorectal cancer revealed by quantitative proteomic analysis. *J. Proteomics* **94**, 473–485 (2013).
52. Zhuo, C. *et al.* Elevated THBS2, COL1A2, and SPP1 expression levels as predictors of gastric cancer prognosis. *Cell. Physiol. Biochem.* **40**, 1316–1324 (2016).
53. Rong, L. *et al.* COL1A2 is a novel biomarker to improve clinical prediction in human gastric cancer: Integrating bioinformatics and meta-analysis. *Pathol. Oncol. Res.* **24**, 129–134 (2018).
54. Lin, J., Goldstein, L., Nesbit, A. & Chen, M. Y. Influence of hormone receptor status on spinal metastatic lesions in patients with breast cancer. *World Neurosurg.* **85**, 42–48 (2016).
55. Liang, Y., Diehn, M., Bollen, A. W., Israel, M. A. & Gupta, N. Type I collagen is overexpressed in medulloblastoma as a component of tumor microenvironment. *J. Neurooncol.* **86**, 133–141 (2008).
56. Li, I. M. H., Horwell, A. L., Chu, G., de Crombrugge, B. & Bou-Gharios, G. Characterization of mesenchymal-fibroblast cells using the Col1a2 promoter/enhancer. *Methods Mol. Biol.* **1627**, 139–161 (2017).
57. Muhl, L. *et al.* Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat. Commun.* **11**, 1–18 (2020).
58. Elyada, E. *et al.* Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019).
59. Lakins, M. A., Ghorani, E., Munir, H., Martins, C. P. & Shields, J. D. Cancer-associated fibroblasts induce antigen-specific deletion of CD8+ T Cells to protect tumour cells. *Nat. Commun.* **9**, 948 (2018).
60. Gong, Y. *et al.* TIMP-1 promotes accumulation of cancer associated fibroblasts and cancer progression. *PLoS One* **8**, e77366 (2013).
61. Song, T., Dou, C., Jia, Y., Tu, K. & Zheng, X. TIMP-1 activated carcinoma-associated fibroblasts inhibit tumor apoptosis by activating SDF1/CXCR4 signaling in hepatocellular carcinoma. *Oncotarget* **6**, 12061–12079 (2015).
62. Carmelita, S. *et al.* Analysis of TIMP-1 expression in leukoplakia and oral squamous cell carcinoma Análise da expressão de TIMP-1 em leucoplasia e carcinoma de boca de células escamosas. *J. Bras. Patol. Med. Lab.* **57**, 1–11 (2021).

63. Desai, N., Trieu, V., Damascelli, B. & Soon-Shiong, P. SPARC expression correlates with tumor response to albumin-bound paclitaxel in head and neck cancer patients. *Transl. Oncol.* **2**, 59 (2009).
64. Podhajcer, O. L. *et al.* The role of the matricellular protein SPARC in the dynamic interaction between the tumor and the host. *Cancer Metastasis Rev.* **27**, 523–537 (2008).
65. Chang, C. H. *et al.* Secreted protein acidic and rich in cysteine (SPARC) enhances cell proliferation, migration, and epithelial mesenchymal transition, and SPARC expression is associated with tumor grade in head and neck cancer. *Int. J. Mol. Sci.* **18**, 1556 (2017).
66. Galván, J. A. *et al.* Immunohistochemical analysis of the expression of cancer-associated fibroblast markers in esophageal cancer with and without neoadjuvant therapy. *Virchows Arch.* **476**, 725–734 (2020).
67. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
68. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
69. Home - GEO - NCBI. <https://www.ncbi.nlm.nih.gov/geo/>.
70. Broad GDAC Firehose. <https://gdac.broadinstitute.org/>.
71. Rio, D. C., Ares, M., Hannon, G. J. & Nilsen, T. W. Purification of RNA using TRIzol (TRI Reagent). *Cold Spring Harb. Protoc.* **5**, (2010).
72. Dwivedi, N., Dhar, S. K., Kuriakose, M. A., Suresh, A. & Das, M. Reference Genes for gene expression analysis in Head and Neck Squamous Cell Carcinoma: a Data Science Driven Approach. *Dent. Res. Oral Heal.* **5**, 21–037 (2022).
73. qPCR Efficiency Calculator, Thermo Fisher Scientific. <https://www.thermofisher.com/in/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/qpcr-efficiency-calculator.html>
74. Immunofluorescence protocol for culture cells. EuroMabNet. <https://www.euromabnet.com/protocols/immunofluorescence-for-culture-cells.php>

Acknowledgements

The authors thank Dr. Amritha Suresh, Dr. Moni A Kuriakose and Dr. Vijay Pillai for providing the patient samples from which the cell lines could be established.

Author contributions

M.D. and S.K.D. conceived and designed the study. N.D. performed the cell based experiments from the isolated fibroblast and epithelial cells. M.D. and N.D. wrote the manuscript. M.D., N.D., N.S. and S.K.D. analysed and interpreted the data from public databases and in-house RNAseq experiments. P.K.M. collected the HNSC patient sample and associated metadata used for IHC. M.D. and S.K.D. reviewed the manuscript. All authors read and approved the final manuscript.

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/s41598-023-40908-w>.

Correspondence and requests for materials should be addressed to S.K.D.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023