2-2023

CEACAM7 expression contributes to early events of pancreatic cancer

Anupam Dhasmana

Swati Dhasmana

Sudhir Kotnala

Partha Laskar

Sheema Khan

See next page for additional authors

Authors

Anupam Dhasmana, Swati Dhasmana, Sudhir Kotnala, Partha Laskar, Sheema Khan, Shafiul Haque, Meena Jaggi, Murali M. Yallapu, and Subhash C. Chauhan

Journal of Advanced Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Original Article

CEACAM7 expression contributes to early events of pancreatic cancer

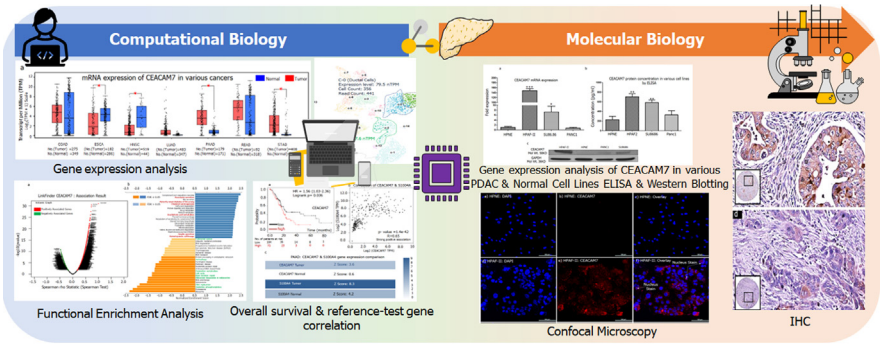
Anupam Dhasmana [a,b,c], Swati Dhasmana [a,b], Sudhir Kotnala [a,b], Partha Laskar [a,b], Sheema Khan [a,b], Shafiul Haque [d,e,f], Meena Jaggi [a,b], Murali M. Yallapu [a,b], Subhash C. Chauhan [a,b,*]

[a] Department of Immunology and Microbiology, School of Medicine, University of Texas Rio Grande Valley, McAllen, USA
[b] South Texas Center of Excellence in Cancer Research, School of Medicine, University of Texas Rio Grande Valley, McAllen, USA
[c] Himalayan School of Biosciences and Cancer Research Institute, Himalayan Institute of Medical Sciences, Swami Rama Himalayan University, Dehradun, India
[d] Research and Scientific Studies Unit, College of Nursing and Allied Health Sciences, Jazan University, Jazan, Saudi Arabia
[e] Centre of Medical and Bio-Allied Health Sciences Research, Ajman University, Ajman, United Arab Emirates
[f] Gilbert and Rose-Marie Chagoury School of Medicine, Lebanese American University, Beirut, Lebanon

HIGHLIGHTS

- The trends of pancreatic cancer (PanCa) incidence and mortality are on rising pattern, and it will a second leading cause of cancer related deaths by 2030.
- Approximately 85–90% PanCa are pancreatic ductal adenocarcinoma (PDAC), which is one of the most challenging and aggressive malignancy.
- PDAC exhibits with grim prognosis as mortality rate is very close to the incidence due lack of early detection methods and effective therapeutic regimen.
- Our team has identified a novel oncogenic protein, carcinoembryonic antigen-related cell adhesion molecule 7 (CEACAM7), that can be useful for early PDAC diagnosis and predictor of patient survival.
- We also observed an increase of CEACAM7 expression in PDAC cell line panel model. However, poorly differentiated, and normal cell lines did not show any expression.
- Commercially available human tissue analysis also strengthened our data by showing strong and positive IHC staining in early-stage tumors.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 June 2022
Revised 16 February 2023

ABSTRACT

Background: The trends of pancreatic cancer (PanCa) incidence and mortality are on rising pattern, and it will be a second leading cause of cancer related deaths by 2030. Pancreatic ductal adenocarcinoma

Peer review under responsibility of Cairo University.

* Corresponding author at: Department of Immunology and Microbiology, School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX, USA.
E-mail address: subhash.chauhan@utrgv.edu (S.C. Chauhan).

https://doi.org/10.1016/j.jare.2023.02.013
2090-1232/

Please cite this article as: A. Dhasmana, S. Dhasmana, S. Kotnala et al., CEACAM7 expression contributes to early events of pancreatic cancer, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2023.02.013

Accepted 19 February 2023
Available online xxxx

Keywords:
Pancreatic cancer
PDAC
CEACAM7
Early detection biomarker
Tumor grading & bioinformatics

(PDAC), major form of PanCa, exhibits a grim prognosis as mortality rate is very close to the incidence rate, due to lack of early detection methods and effective therapeutic regimen. Considering this alarming unmet clinic need, our team has identified a novel oncogenic protein, carcinoembryonic antigen-related cell adhesion molecule 7 (CEACAM7), that can be useful for spotting early events of PDAC.

Methodology: This study includes bioinformatics pre-screening using publicly available cancer databases followed by molecular biology techniques in PDAC progressive cell line panel and human tissues to evaluate CEACAM7 expression in early events of pancreatic cancer.

Results: PanCa gene and protein expression analysis demonstrated the significantly higher expression of CEACAM7 in PDAC, compared to other cancers and normal pancreas. Overall survival analysis demonstrated an association between the higher expression of CEACAM7 and poor patients' prognosis with high hazard ratio. Additionally, in a performance comparison analysis CEACAM7 outperformed S100A4 in relation to PDAC. We also observed an increase of CEACAM7 in PDAC cell line panel model. However, poorly differentiated, and normal cell lines did not show any expression. Human tissue analysis also strengthened our data by showing strong and positive IHC staining in early-stage tumors.

Conclusion: Our observations clearly cite that CEACAM7 can serve as a potential early diagnostic and/or prognostic marker of PDAC and may also potentiate the sensitivity of the existing biomarker panel of PDAC. However, further studies are warranted to determine its clinical significance.

Introduction

According to the American Cancer Society 2022, pancreatic cancer (PanCa) affects around 62,210 persons (32,970 men and 29,240 women) in the US [1]. PanCa is the 4th leading cause of cancer related deaths along with very poor prognosis as PanCa patients have only 11% of 5-year survival rate [2,3]. Generally, 85–90% of PanCa are adenocarcinomas (pancreatic adenocarcinomas), such as pancreatic ductal adenocarcinomas (PDAC) [4]. PDAC is characterized as a highly aggressive disease with dismal overall prognosis and mortality rate near to the incidence rate [5,6]. Surgery is one of the important therapeutic interventions for PDAC patients, however more than 80% of patients are diagnosed at advanced stages, which limits the intervention of surgical option. However, early-stage diagnosis can increase the survival chance up to 80% [4,5]. Unfortunately, no early and FDA-approved tumor specific biomarker for PDAC is available at present. Currently, multidetector computed tomography (CT) and the serum level of carbohydrate antigen 19-9 (CA19-9) are used for PDAC diagnosis, but both have their own limitations (7). CA19-9 has a broad range of cross immune-reactivity in multiple malignancies and nonmalignant indications, such as chronic pancreatitis and liver cirrhosis [6]. Thus, reliable specific molecular signals capable of predicting the primary detection and prognosis of PDAC are desired.

Herein, an understudied oncogenic protein, Carcinoembryonic antigen-related cell adhesion molecule 7 (UniProtKB – Q14002; also known as CEACAM7 or CGM2) was explored to enhance PDAC biomarker panel. CEACAM7 is a member of the carcinoembryonic antigen (CEA) family. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) belong to the immunoglobulin (Ig) family with differential expression over the surfaces of cells. The functional-cellular processes of CEACAMs are quite diverse like cell adhesion, tumor progression, proliferation, signal transductions and phagocytosis [7]. The CEA group consists of 29 genes or pseudogenes (18 are expressed, 07 are CEA subgroups and 11 related to pregnancy specific glycoprotein subgroups). CEAs are considered crucial cancer biomarkers for colorectal cancers (CRCs) and other malignancies. These subgroup proteins (i.e., CEA, NCA, BGP, CGM1, CGM2/CEACAM7, CGM6 and CGM7) are membrane associated and exhibit complex epithelial expression pattern in normal and cancerous tissues. Among all 7 subgroup proteins, CEA, NCA, CEACAM7, and CGM6 are anchored to the plasma membrane through glycosyl phosphatidylinositol (GPI) moiety without intracellular domain and a relatively small extracellular domain [8]. However, CEACAM7 is a deputized and neglected CEA family protein. Earlier study on CEACAM7 indicated that, in embryonic colon tissues, CEACAM7 migrates from the base of epithelial cells to the apical surface within a few days of birth. Afterwards, CEACAM7 locates on well differentiated epithelial cells of the colon and rectum [9].

According to BioProject (PRJEB4337) entitled, 'HPA RNA-seq normal tissue, CEACAM7 does not express in most of the normal tissues except low level in colon. Remarkably, CEACAM7 expression is found to be down regulated in colorectal carcinoma in comparison with normal colon. A recent study reported CEACAM7 as a viable target for CAR-T as it has significantly high level only in PDAC tumor epithelial cells with no expression in normal pancreas [9,10]. Therefore, the present study was aimed to investigate the diagnostic/prognostic potential of CEACAM7 in early events of PDAC, thus, potentiating the existing PDAC biomarker panel.

Material and methods

Gene expression analysis

BioProject (PRJEB4337) entitled, 'HPA RNA-seq normal tissues', shows the normal levels of CEACAM7 in various tissues of human. In this BioProject database, RNA-seq analysis was performed in 95 normal human tissue samples. The pan cancer gene expression analysis was done by GEPIA database [11]. The parameters used for the analysis were ANOVA differential method, 1 Log_2FC cutoff, 0.01 is characterized as median tumor and normal tissues, q-value Cutoff, log_2 (TPM + 1) for log-scale and 33 differential cancer databases. For a better and significant analysis of the data visualization, some important cancers (COAD/Colon cancer, ESCA/Esophageal cancer, HNSC/Head & Neck Cancer, LUAD/Lung Cancer, PAAD/Pancreatic Cancer, READ/Renal cancer, STAD/Stomach Cancer) were selected and compared for the gene expression using box style method with p-value cutoff 0.01, Log_2FC cutoff 1 and 0.4 jitter size. For PanCa protein expression analysis of CEACAM7, CPTAC database was used to retrieve the data from UALCAN [12]. From this database, 10 important cancers were selected to detect the protein expression against the query protein.

Overall survival and reference-test gene correlation analysis

For survival analysis, Kaplan Meier (KM) Plotter was employed to identify the Overall Survival (OS) in 261 PDAC cases, 743 cutoff value, 0-31517 expression range of probe and hazard ratio of more than 1 was selected [13].

The pairwise gene correlation analysis between two genes (TPM value of S100A4 as a reference gene and CEACAM7 as a test gene) in PAAD tissues were computed, using the GEPIA database [14,15] to get positive or negative association (correlation coefficients guidelines: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module9-Correlation-Regression/PH717-Module9-Correlation-Regression4.html). The log and non-log scale axis were employed to calculate and visualize, respectively. The comparison of expression matrix of reference and test genes in tumor and normal tissues was plotted [11].

Assessment of CEACAM7 associated co-expressed genes

LinkFinder (for gene association) and LinkInterpreter (for enrichment analysis) sub-modules of LinkedOmics [16] were used for exploring the genes that exhibited disparity in association with CEACAM7 in Pancreatic Cancer. From the UNC Institute's module, RNA-seq datasets, with 178 sample sizes were chosen and scrutinized using Spearman's correlation coefficient and characterized in volcano plots. The positive or negative correlated genes were screened by the p-value ($p < 0.01$) [17].

The co-expressed genes were utilized for the gene enrichment analysis by using the Gene Set Enrichment Analysis (GSEA) database, under the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways plugin. The parameters for enrichment analysis were lowest no. of IDs in the 25 categories, highest no. of IDs in the 2000 categories, top 25 significance levels, 10,000 number of permutations.

Spot prediction of CEACAM7 in single cell type cluster of pancreatic cells

Single cell RNA expression of CEACAM7 in pancreas cell clusters was visualized by using Uniform Manifold Approximation and Projection (UMAP) plot, from the Protein Atlas database [18]. Using this method, the predictive positioning of CEACAM7 was detected in specific cell clusters of pancreatic cells from the range of 0–280 nTPM value and the number of included cells were expressed in the form of read count and cell count.

PDAC cell line cultures

The commercially available PDAC cell lines were procured from ATCC. For this study we have selected only new or early passage number cell lines with varying differentiation degrees of PDAC, like normal pancreatic cell line (HPNE), well differentiated cell line (HPAF-II), moderately differentiated cell line (SU86.86), and poorly differentiated cell line (PANC1) [19]. The DMEM, RPMI or DMEM/F12 media supplemented with 10% fetal bovine serum (FBS) and 1% (w/v) penicillin–streptomycin was used for the cell cultivation (Gibco, Thermo Fisher Scientific, Grand Island, NY, USA) under humidified atmosphere with 5% CO_2 at 37 °C [20].

Gene expression analysis of CEACAM7 in various PDAC cell lines

TRIzol (Invitrogen, USA) was utilized for the extraction of RNA from various PDAC cell lines. Reverse transcription was done using High-Capacity cDNA reverse transcriptase kit (ThermoFisher) according to manufacturer's protocol. CEACAM7 and β-Actin specific primers were used to amplify the collected cDNA. qPCR was performed using SSO Fast Eva Green supermix (Bio-Rad). The differential expression levels of CEACAM7 mRNA were measured by Bio-Rad CFX96 using sequence-specific primers using a previously published protocol [21]. The forward and reverse primers of CEACAM7 and β-actin were purchased from IDT and sequences were A) CEACAM7 forward primer: 5′-GTT ACC CAC AAT GAC GCA GGA-3′; CEACAM7 reverse primer: 5′-TCC ACC GGA TTG AAG TTG TTG-3′. B) beta-actin forward primer: 5′-GTG CTA TCC CTG TAC GCC TC-3′; β-actin reverse primer: 5′-GAG GGC ATA CCC CTC GTA GA-3′.

ELISA and immunoblotting

To quantify the protein concentration of CECAM7 in various PDAC cell lines, commercially available R&D System, Human CEACAM7 DuoSet ELISA kit (Cat. No. DY4478-05, Lot: P187793) was used as per the manufacturer's protocol. CEACAM7 antibody (Monoclonal mouse # MAB44782, Lot No. CKUS011804, 1:500; R&D) was used as a test primary antibody and GAPDH was used as endogenous control (anti rabbit Cat:14C10 Cell Signaling) for the western blot as per previously published protocol [22]. Secondary antibodies were procured from Promega (anti rabbit W4018 and anti-mouse W4028).

Confocal microscopy

Confocal microscopy was used to determine the expression and subcellular localization of CEACAM7 in various PDAC cell lines. Cells were seeded in four chamber slides and after the cells attained 70% confluency, they were prepared for confocal microscopy [23]. Briefly, media was aspirated, and cells were washed twice with phosphate-buffered saline-Glycine (PBS-G), followed by fixing of cells with 2–4% of paraformaldehyde for 15 mins. Cells were incubated with Triton X-100 (0.2%) to permeabilize cells, followed by two PBS washes. Proceeded with 1 h blocking with 10% donkey serum in 1X PBS. Primary antibody (Monoclonal mouse CEACAM7 BAC2:sc-59946 dissolved in 5% of donkey serum; SIGMA-D9663-10 ml in PBS) was added to all test PDAC cell lines and incubated overnight. Primary antibody (1:200) incubated cell lines were washed three times with 1x PBST for 5 min each followed by the addition of donkey anti-mouse Cy3 secondary antibody (1:200, Jackson Immunoresearch Lab, Code: 715-1660151) and incubated for 1 h in the dark. Cells were washed and mounted using DAPI containing-Vecta shield mounting medium (Vector Laboratories; Burlingame, CA H-1200). Immunofluorescence was examined and photographed under A1R HD25 Confocal Microscope (Nikon Instruments Inc., Melville, NY, USA). Photomicrographs were captured in the red channel.

Immunohistochemical analysis of CEACAM7 expression and composite scoring in stained pancreatic normal, pancreatitis and cancerous tissues.

Commercially available human pancreatitis and pancreatic carcinoma Tissues Micro Array (TMAs) with normal tissue (PA691 and TMAFC1006_14) were obtained from US Biolab Corporation, Inc (Rockville, MD, USA). CEACAM7 expression was assessed by immunohistochemistry (IHC) staining using a commercially available kit (Biocare Medical) as reported earlier [20]. Briefly, the tissue sections were deparaffinized and rehydrated using different grades of alcohol, followed by endogenous peroxidase quenching. Blocking and incubation with monoclonal mouse CEACAM7 antibody (R&D # MAB44782, Lot No. CKUS011804, conc. 5 μg/ml) was completed overnight. The sections were then incubated with secondary MACH4 Universal HRP mouse polymer (#M4U534H) followed by staining with 3-diaminobenzidine (DAB) chromogen (#DB801L) and counterstaining with hematoxylin.

Following histochemical processing, slides were digitally scanned by 3DHISTECH Panoramic MIDI and analyzed for CEACAM7 staining. Blinded to the patient's record, scanned images were minutely scrutinized, and a consensus scoring was deduced as described earlier [20]. The intensity of immunostaining of the CEACAM7 was graded on a scale from 0 to 4 (0 for no staining, 1 for weak immunostaining; 2 for moderate immunostaining; 3 for

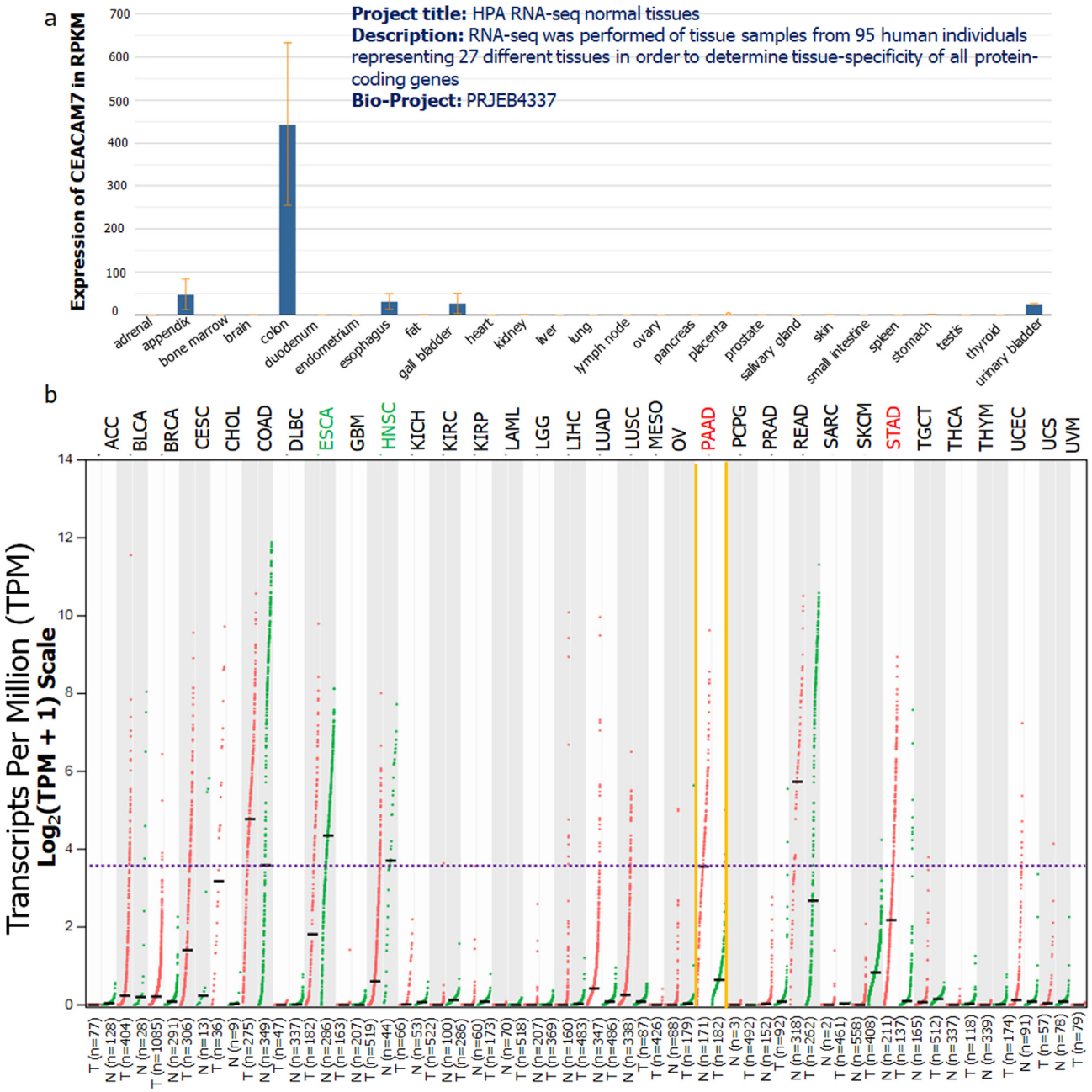


Fig. 1. Expression profiling of CEACAM7 in normal and cancerous tissues: (a) CEACAM7 in normal tissues: CEACAM7 is cell adhesion molecule 7 (Gene ID: 1087). According to BioProject (PRJEB4337) entitled "HPA RNA-seq normal tissues", the expression level of CEACAM7 at normal condition in various tissue of human was showcased. This data was retrieved from the NCBI BioProject database. In this BioProject database, RNA-seq analysis was performed on normal tissue samples from 95 human individuals. This expression profiling clearly depicted that CEACAM7 does not express in most of the tissue in normal condition expect colon tissue. (**b) Pan cancer gene expression of CEACAM7:** The gene expression profile across all tumor samples and paired normal tissues is indicated in the dot plot. Each dot represents the expression of samples. The expression data was scaled by $\log_2$(TPM + 1). Red: Probable high tumor expression of CEACAM7; Green: Normal cell expression of CEACAM7. Source: GEPIA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

strong immunostaining, and 4 for very strong immunostaining). The percentage of the cells positive for CEACAM7 immunostaining within the normal, adjacent normal and tumor cores, were scored as follows: 0–25% as 1, 26–50% as 2, 51–75% as 3 and 76–100% as 4.

The composite score (CS) values ranging from 0 to 16, were calculated by multiplying the value for staining intensity (0–4) and the percentage of immune-stained cells (0–4) for each individual sample. Finally, the mean composite score (MCS) was determined by calculating the average of the composite scores of respective samples in each category. The cores were not consented, and stroma was excluded from the study. In order to evaluate the accuracy of the test, an ROC curve was plotted to detect the area under curve (AUC) of CEACAM7 using PA2072b and TMAFC1006_14 TMAs. The trial version of Medcalc software was used to compute the complete test.

Statistical analysis

To assess the difference between relevant groups in various experiments, the analyses were conducted using unpaired, one-tailed Student's *t*-tests. All error bars used in the attached figures indicate the standard error of mean (SEM). The number of stars

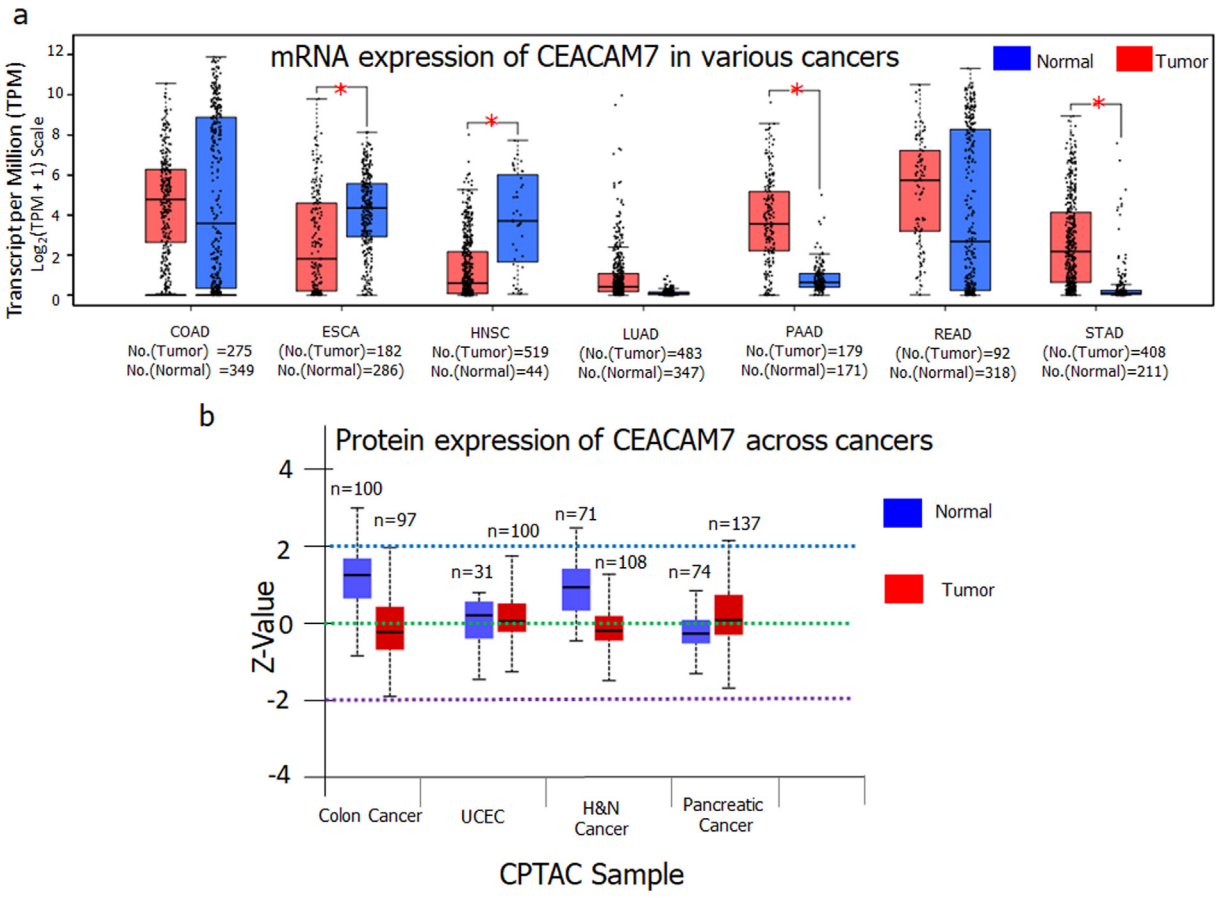


Fig. 2. (a) Differential gene expression box plot analysis of CEACAM7: GEPIA box plot analysis to observe significantly higher mRNA expression of CEACAM7 in PAAD followed by STAD. The red colored box is tumor tissue dataset and blue colored box is normal tissue dataset. Red * denoted significant data. The method for differential analysis is a one tailed unpaired student *t*-test, using disease state (tumor or normal) as a variable for calculating differential expression. The level of significance is p-value cutoff 0.01. **(b) Protein expression level of CEACAM7:** CPTAC analysis to observe mass-spectrometry-based proteomic characterization of CEACAM7 in various cancers. The statistics of each cancer indication are based on the Z-values, which denote standard deviations (SD) from the median value of all samples. The statistical significance p-value of normal vs Pancreatic adenocarcinoma (PAAD): $3.47e^{-04}$, normal vs colon cancer: $6.34e^{-18}$, normal vs Uterine corpus endometrial carcinoma (UCEC): $6.99e^{-02}$, normal vs Head and neck carcinoma (H&N Cancer): $4.03e^{-14}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

used on each graph signifies the level of significance – i.e., one-star (*) indicates P-values below 0.05, two stars (**) values below 0.01 and three stars (***) for values below 0.001.

Ethical statement

No ethical approval was necessary for this study, as this study didn't involve any animal or human participants.

Result

Differential expression levels of CEACAM7 in different types of normal and cancer tissues

As per BioProject (PRJEB4337) entitled, HPA RNA-seq normal tissues, the expression level of CEACAM7 in normal condition in various human tissues was illustrated. This expression profiling clearly depicted that CEACAM7 significantly expressed in colon. Apart from colon, CEACAM7 was not observed to express in most of the normal tissues (Fig. 1**a**).

To appraise the CEACAM7 disparity in human cancers, the GEPIA database was employed. Using this database, we obtained CEACAM7 gene expression data of RNA sequencing from TCGA and GTEx datasets through tumor samples along with paired normal tissues. 33 tumor types expression analyses for CEACAM7 were available in this database. Among all tumor types, CEACAM7 expression was higher in pancreatic adenocarcinoma (PAAD) and stomach adenocarcinoma (STAD), both of which are denoted with a red color (Fig. 1**b**). The median of line of PAAD was higher (around 3.8 TPM) than the STAD (around 2.0 TPM), this feature again reflects the strong association of CEACAM7 with PAAD. However, the expression of CEACAM7 was lower in esophageal carcinoma (ESCA) and head and neck squamous cell carcinoma (HNSC), denoted with a green color (Fig. 1**b**). This data revealed that CEACAM7 is over expressed in PAAD.

To further validate the differential expression of CEACAM7 in important cancers (COAD, ESCA, HNSC, LUAD, PAAD, READ, STAD), box plot analysis was performed, where we have noticed that PAAD was the only cancer condition showing remarkably higher CEACAM7 expression than normal pancreatic tissues followed by STAD. ESCA and HNSC exhibited significantly lower expressions in cancer conditions, however non-significant expression was observed in COAD and READ cancer than normal conditions. Albeit LUAD exhibited a bit higher CEACAM7 expression but was non-significant (Fig. 2**a**).

Differential expression analysis was continued utilizing UALCAN for protein expression assessment with data from CPTAC dataset. Mass-spectrometry-based proteomic profiling demonstrated

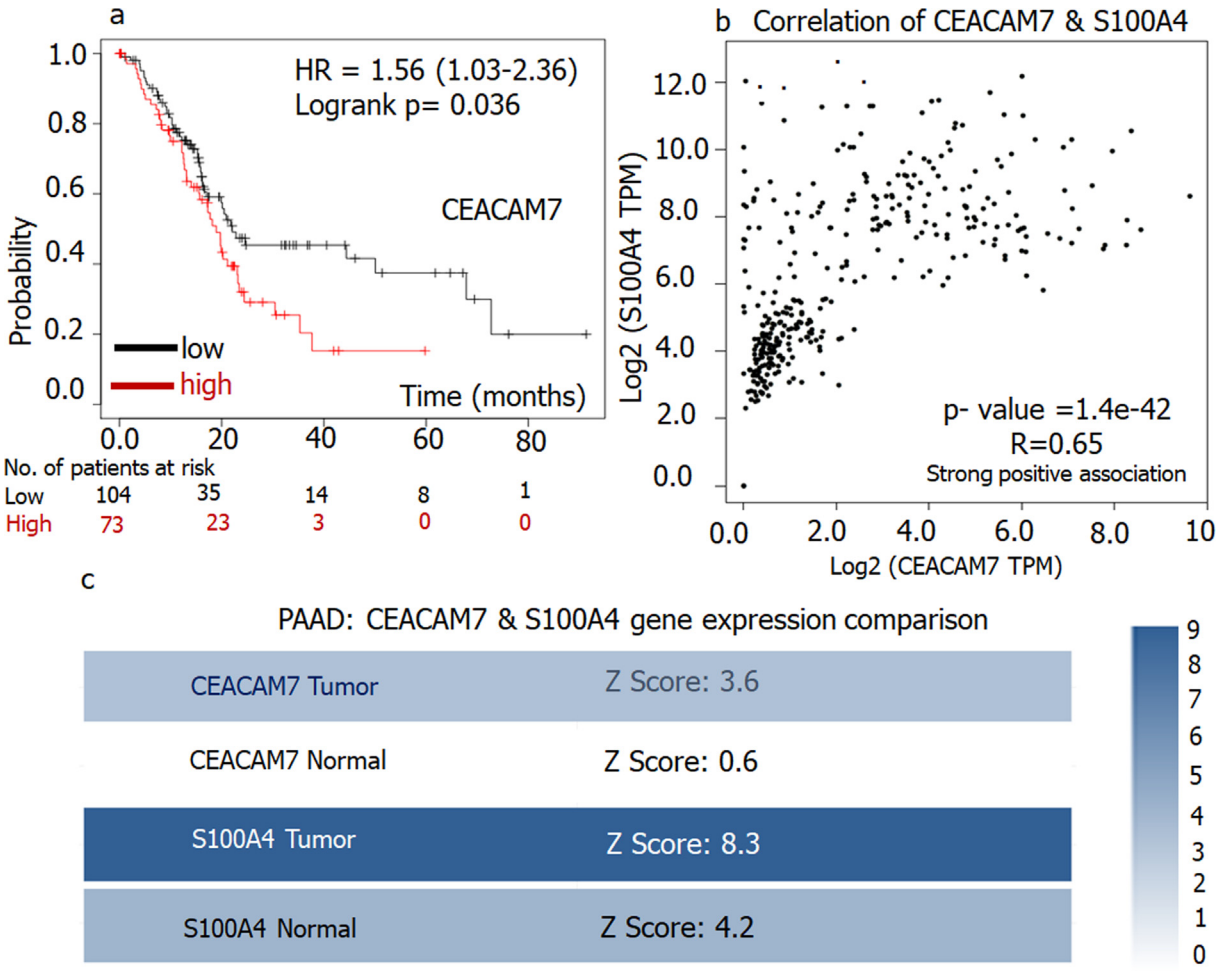


Fig. 3. Patients' survival and gene correlation plots of CEACAM7: (a) This survival plot was generated by Kaplan-Meier Plotter (Hazard Ratio (HR): 1.56, Logrank p = 0.036, total no. of patients: 261); the cutoff value that was used in the analysis is 743 (ranging from 154 to 1426). This analysis was done based on mRNA expression. **(b)** Correlation graph of CEACAM7 and S100A4, which demonstrates the strong positive association of both of genes in pancreatic cancer. **(c)** Comparison of z-score gene-based expression of CEACAM7 and S100A4in pancreatic cancer.

that high CEACAM7 protein expression was especially observed in pancreatic cancer condition, while other cancers like colon cancer and head and neck cancer show lower expression, whereas UCEC was showing almost same level of expression in cancerous and normal condition **(**Fig. 2**b).**

Correlation of CEACAM7 expression with overall survival and reference gene

Kaplan-Meier Plotter (Hazard Ratio (HR): 1.56, Logrank p = 0.036, total no. of patients: 261) was employed to ascertain the relationship between CEACAM7 expression and patients' survival rates in pancreatic cancer. Higher expression of CEACAM7 correlates with reduced overall survival in pancreatic cancer patients **(**Fig. 3**a).**

A correlation analysis was performed to assess the significance of test gene (CEACAM7) in pancreatic cancer against the reference gene (S100A4). The final R and p-value were 0.65 (strong positive association) and $1.4e^{-42}$, respectively **(**Fig. 3**b).** The gene expression comparison results show the expression profile of CEACAM7 and S100A4 with a z-score of 3.6 vs 8.3 in pancreatic tumor and 0.6 vs 4.2 in normal pancreatic cells, respectively (Fig. 3**c).**

CEACAM7 associated co-expressed genes and its functional enrichment

LinkFinder & LinkInterpreter, were used to explore the possible potential mechanisms of CEACAM7 and its associated genes. TCGA_PAAD search, and target dataset of UNC institute were selected for the RNAseq analysis using HiSeq RNA platform and Firehose_RSEM_log2 pipeline. The patient sample size was 178. A total of 19,774 genes were enriched in this study, out of which 7,572 genes showed significant negative correlations (green dots) with CEACAM7 and the remaining 12,201 genes showed significant positive correlation (red dots) with CEACAM7 **(**Fig. 4**a).**

For functional enrichment of the co-expressing genes with CEACAM7, enrichment studies using GSEA method and KEGG pathways category in the LinkInterpreter platform of LinkedOmics were performed. As demonstrated in Fig. 4**b,** the CEACAM7 co-expression genes have positive significant involvement in pancreatic physiology like pancreatic secretion, insulin secretion, bile secretion and maturity onset diabetes of the young. Apart from that, other physiological functions include chemical carcinogenesis, hematopoietic cell lineage, etc. Whereas colorectal cancer, autophagy, pyrimidine metabolism, cell cycle, base excision repair, ribosome biogenesis in eukaryotes, DNA replication, and oxidative phosphorylation pathways had negative significant correlation with CEACAM7.

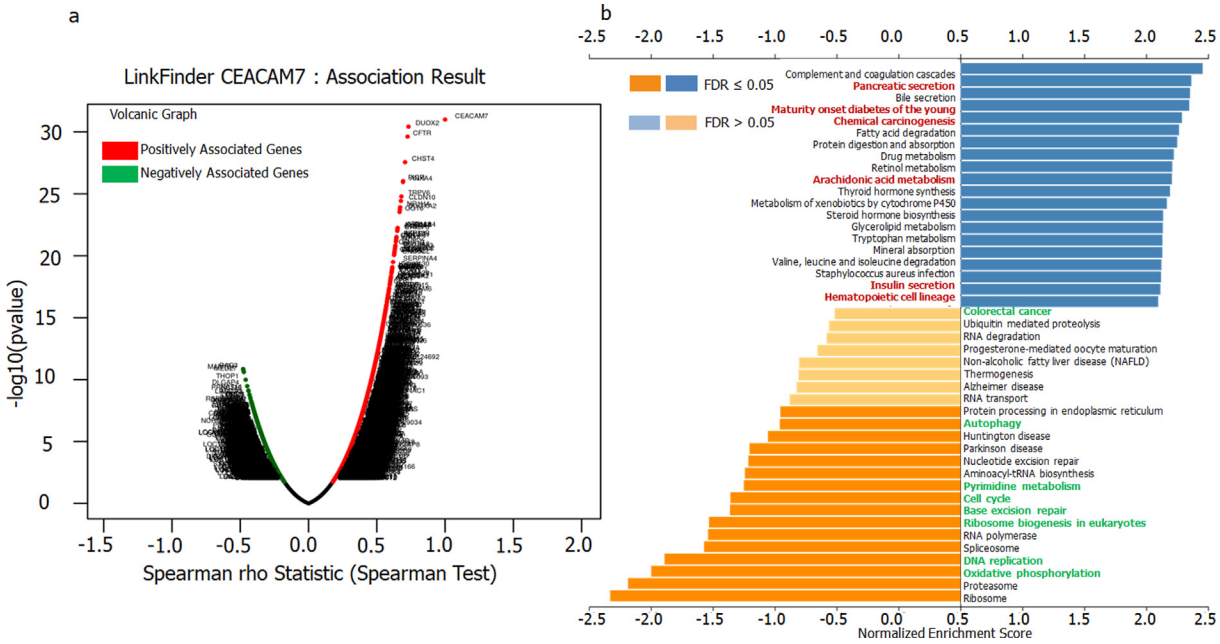


Fig. 4. Volcano graph and pathways enrichment analyses of CEACAM7 and associated genes: (a) Co-expression genes of CEACAM7 in pancreatic cancers (LinkedOmics: LinkFinder). The significantly correlated genes with CEACAM7 were analyzed by Spearman test. Green and red dots represented the negative and positive correlations with CEACAM7, respectively. (**b)** Pathway enrichment analysis of CEACAM7 and associated genes by using LinkInterpreter. Blue and orange bar graphs represent pathway enrichment of positively and negatively associated genes with CEACAM7, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

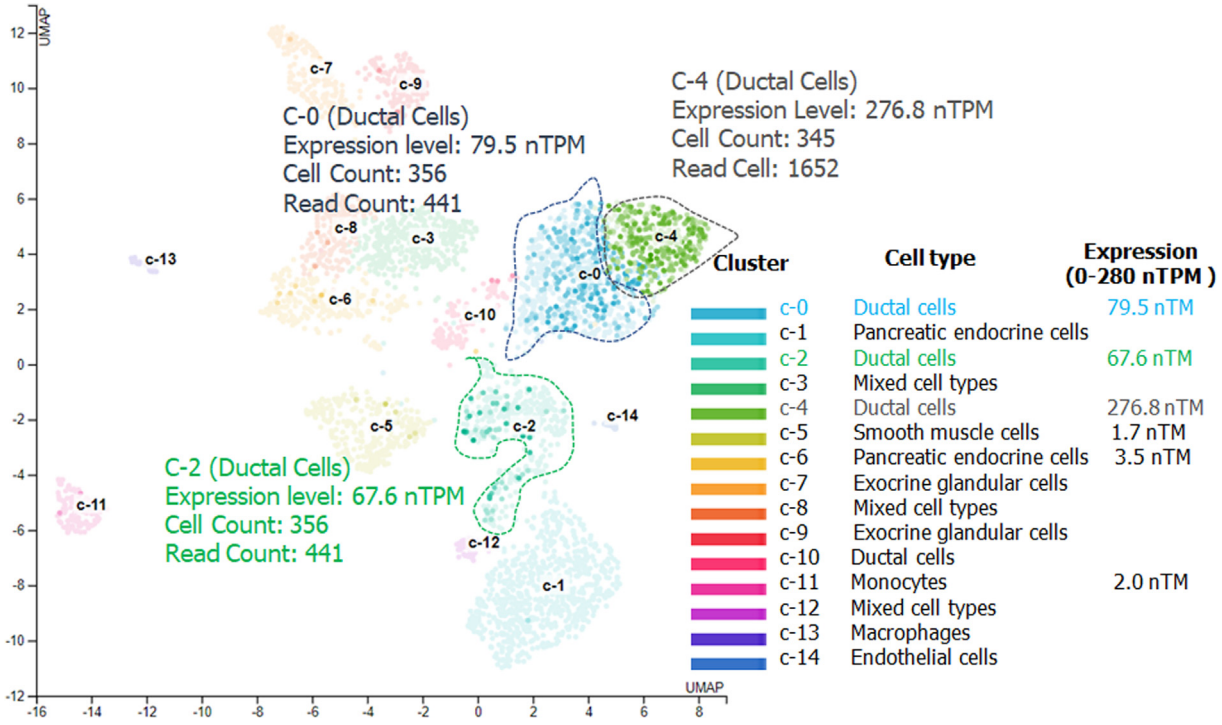


Fig. 5. CEACAM7 positioning in pancreas: RNA expression of CEACAM7 in pancreas tissues. The single cell type clusters identified in this tissue were visualized by a UMAP plot.

CEACAM7 expression localization in pancreas by using single cell clustering

The predictive localization of CEACAM7 in the pancreas was demonstrated in this section. Maximum expression was observed in the ductal cells. These single cell type clusters were indicating the higher expression of CEACAM7 in ductal cell clusters like C0 (79.5 nTPM), C2 (67.6 nTPM), and C4 (276.8nTPM). Remaining clusters like C6 (3.5 nTPM) belonged to pancreatic endocrine cells, C11 to monocytes (2.0 nTPM), and C5 to smooth muscle cells (1.7 nTPM). This graph was plotted based on a UMAP plot as shown in Fig. 5.

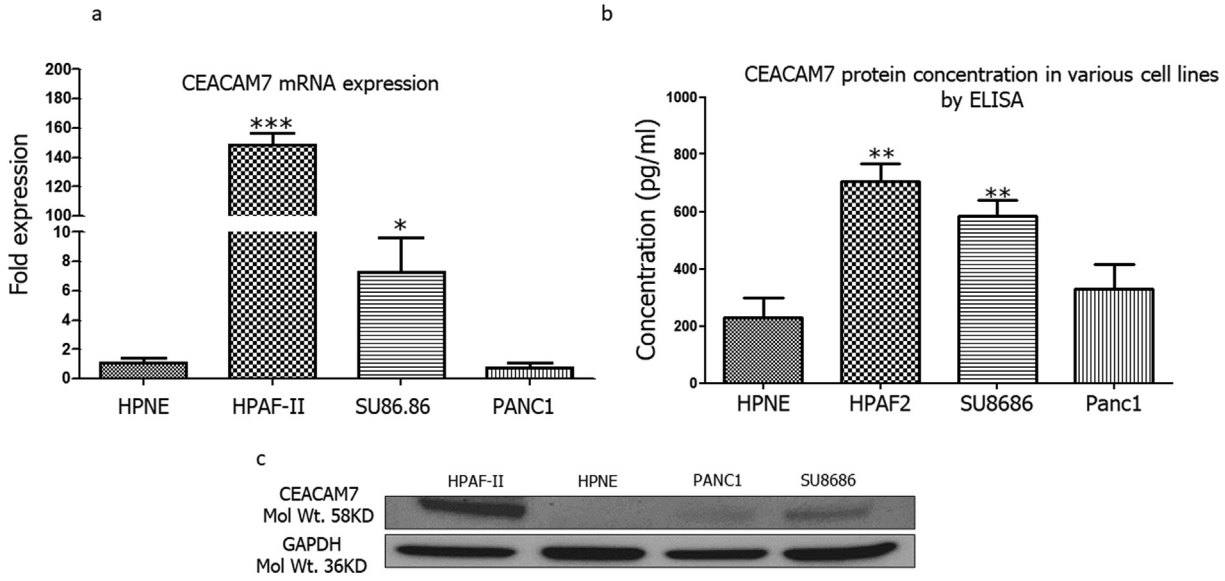


Fig. 6. Transcriptomic and translational expression of CEACAM7: (a) mRNA expression analysis: Gene expression of CEACAM7 in progressive PDAC cell lines with respect to HPNE, a normal pancreatic cell line. n = 3; p-value *<= 0.05 and ***<=0.001. (**b) Protein expression:** Quantification of CEACAM7 protein expression in progressive PDAC cell lines was demonstrated using ELISA. n = 3; p-value *<= 0.05, ** <=0.01 and *** <=0.001. (**c**) Representative image of western blot recognition pattern of CEACAM7 in lysates PDAC cell lines (n = 3).

Transcriptional and translational analysis of PDAC progressive cell lines

qPCR analysis was performed to screen the mRNA transcript expression of CEACAM7 in various PDAC cell lines like HPAF-II, SU86.86, and PANC1 with respect to HPNE. This fold change experiment provides the important trends of CEACAM7 expression. With respect to HPNE, HPAF-II showed the highest mRNA expression of CEACAM7, followed by SU86.86 and PANC1 (Fig. 6**a).** To determine the translational potential and precise protein concentration of CEACAM7 in various PDAC cell lines and HPNE, an ELISA test was performed with a 1:10 sample dilution. According to this sensitive test, HPAF-II showed the highest CEACAM7 protein concentration (704.435 pg/ml), followed by SU86.86 (583.806 pg/ml), PANC1 (329.8719 pg/ml), and HPNE (231.237 pg/ml). Furthermore, western blot analysis validated the expression pattern of CEACAM7 in various PDAC cell lines. The findings of ELISA and western blot correlated and demonstrated significant expression of CEACAM7 in HPAF-II, followed by SU86.86, but no significant expression was observed in HPNE and PANC1 (Fig. 6**b & c).**

Confocal microscopy (Fig. 7**)** was employed to evaluate the CEACAM7 expression pattern in PDAC and normal (HPNE) cell lines. This follows a similar trend of mRNA transcript expression of CEACAM7. Normal pancreatic cell line HPNE did not show a detectable level of CEACAM7 expression (Fig. 7**a, b, c**), while it was highest in HPAF-II (Fig. 7**d, e, f**), followed by SU.86.86 (Fig. 7**g, h, i**), and very faint in PANC1 (Fig. 7**j, k, l**). The CEACAM7 expression in HPAF-II was observed majorly in cytoplasm followed by the nucleus (Fig. 7**f**), whereas in SU 86.86 this expression was confined to cytoplasm with lesser intensity and coverage (Fig. 7**i).**

Aberrant expression of CEACAM7 in pancreatitis and PDAC samples

Negative CEACAM7 expression or negligible staining was observed in pancreatitis cores (MCS of acute pancreatitis = 0.75 and MCS of chronic pancreatitis = 0.65, the representative IHC images of pancreatitis tissues cores are shown in supplementary Fig. 1). However, 84.4% of PDAC sample cores were found to be reactive with CEACAM7 antibody (38 positive stained cores; 7 negative stained core and 3 normal cores with no stain). Among all PDAC sample cores, 8 cores belong to adjacent normal cores and exhibited faint to moderate staining. CEACAM7 expression was the highest in well differentiated tumors (MCS = 10.5), followed by moderately differentiated tumors (MCS = 8.1) while poorly differentiated tumors demonstrated relatively low expression (Fig. 8**a–d**). Additionally, the quantification of immunostaining in the form of MCS is shown in Fig. 8**e**. IHC analysis shows that some of the normal cores, which were considered as adjacent normal tissue cores (MCS = 7.875) also showed immunostaining because of the cancer influenced molecular changes. ROC analysis was performed to calculate the sensitivity of CEACAM7, which was computed 0.724 of the area under curve with a p-value of 0.033 (Supplementary Fig. 2), which means this model has about 72% of capability to distinguish between positive and negative class.

Discussion

PanCa remains a major health concern; it will become the second leading cause of cancer deaths in the United States by 2030 [24,25]. It is generally diagnosed at a late stage, with extensive metastases [26,27]. The ability of PanCa diagnosis is concerning for healthcare professionals as the conventional biomarkers such as carbohydrate antigen 19-9 (CA19-9), CA125, MUC1 and carcinoembryonic antigen (CEA), have relatively less sensitivity/specificity. Clinically, these biomarkers are inadequate as early detection markers of PDAC, at present [6,14,28-30].

The data presented in this amalgamated (*in silico* supported with *in vitro* experimental) study suggests that CEACAM7 can be an early diagnostic biomarker candidate because of its special presence in early events of PDAC. According to Stefan et al. 2000, CEACAM7 is differentially expressed in normal colon tissues and inversely downregulated in hyperplastic colorectal polyps and early adenomas [9]. Koji Yoshida et al 2003, conducted DNA microarray analysis of 3,456 human genes from pancreatic ductal carcinoma tissues [31]. In this study, for the first time, CEACAM7 was identified as an activated gene. A recent study by Raj et al, 2021 has reported CECAM7 as a molecular target for CAR-T for PDAC [10]. In this current study, bioinformatics analyses revealed

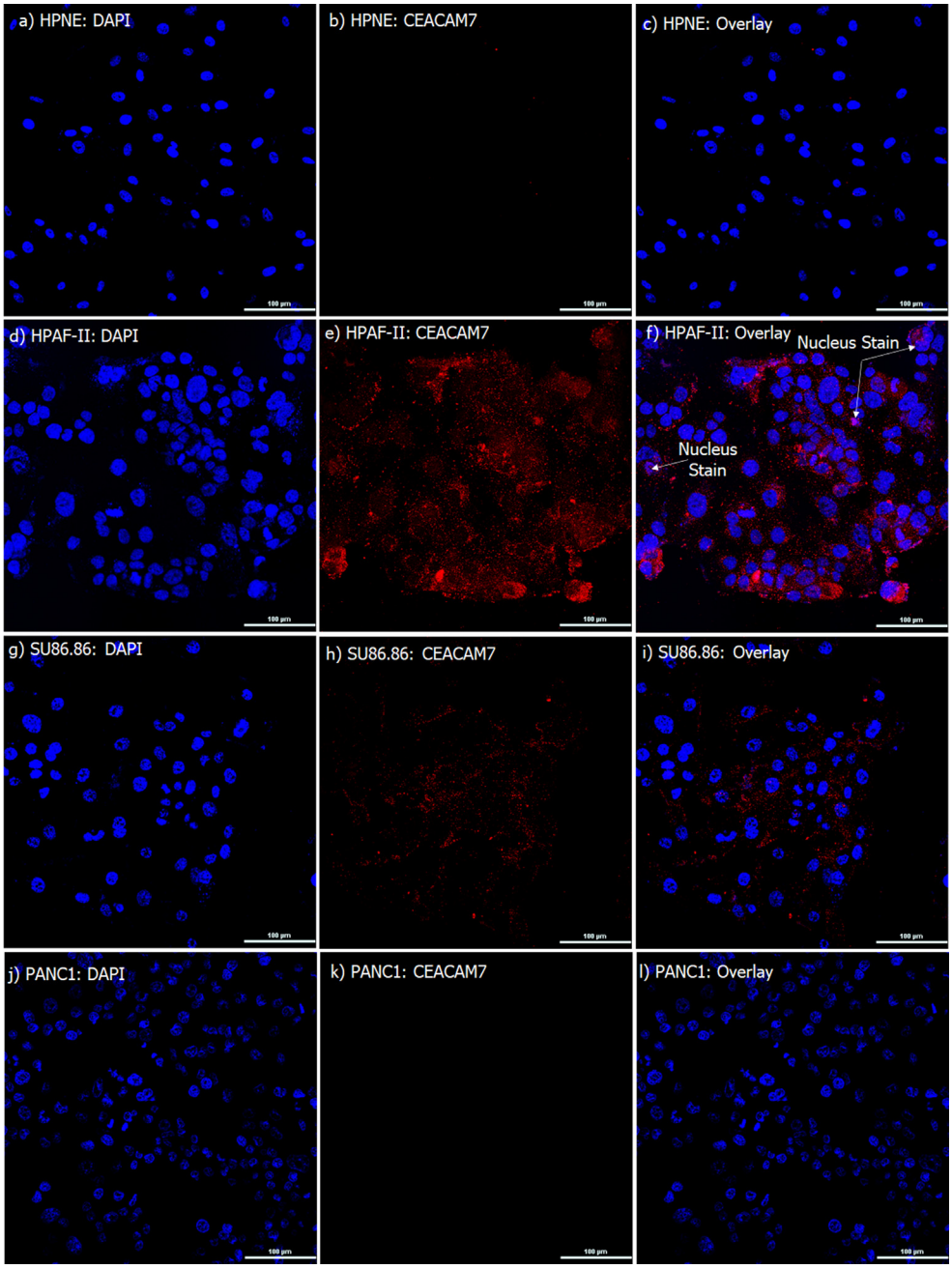


Fig. 7. Confocal microscopy of CEACAM7 in normal pancreas and PDAC cell line panels: Protein expression analysis of CEACAM7 in progressive PDAC cell lines with respect to HPNE, a normal pancreatic cell line. Cells were processed for CEACAM7 immunostaining and confocal microscopy. Original Magnifications 40×.

that PAAD possesses the highest expression of CEACAM7 in the cancerous stage as compared to normal conditions. Furthermore, high CEACAM7 expression level and low survival probability were discovered to be directly proportional. To further cross validate the feasibility and self-assurance of CEACAM7 in PanCa, we have correlated the expression profiling of CEACAM7 with an important PDAC biomarker, S100A4. In this correlation, the R value of CEACAM7 and S100A4 had strong positive correlation with a significant p-value. GEPIA gene expression comparison analysis indicated that CEACAM7 has almost 6 times higher significant expression level in cancerous condition as compared to normal condition, while S100A4 possessed <2 times higher expression level in cancerous condition than normal condition.

The results of functional enrichment analysis revealed that CEACAM7, and its co-expressed genes were significantly involved in normal molecular functions followed by the pathological indications of pancreas. Positively correlated genes were involved in various pancreatic functions like pancreatic secretion, insulin secretion, fatty acid degradation, protein digestion and absorption, glycerolipid metabolism and hematopoietic cell lineage. Some CEACAM7 co-expressed genes were also involved in pancreatic pathological conditions like maturity onset diabetes in young persons. Although PDAC rarely occurs in the young population, a recent study asserted that PanCa is now more frequently disseminating in the young African population [29]. The data indicates that approximately 80% of PanCa patients suffer with either new-onset type 2 diabetes or impaired glucose tolerance at the time of diagnosis. Latest reports recommend that type 2 diabetes is a major risk and prognostic factor for PanCa [32]. Hence, it can be inferred that in future, CEACAM7 and other associated genes can be consid-

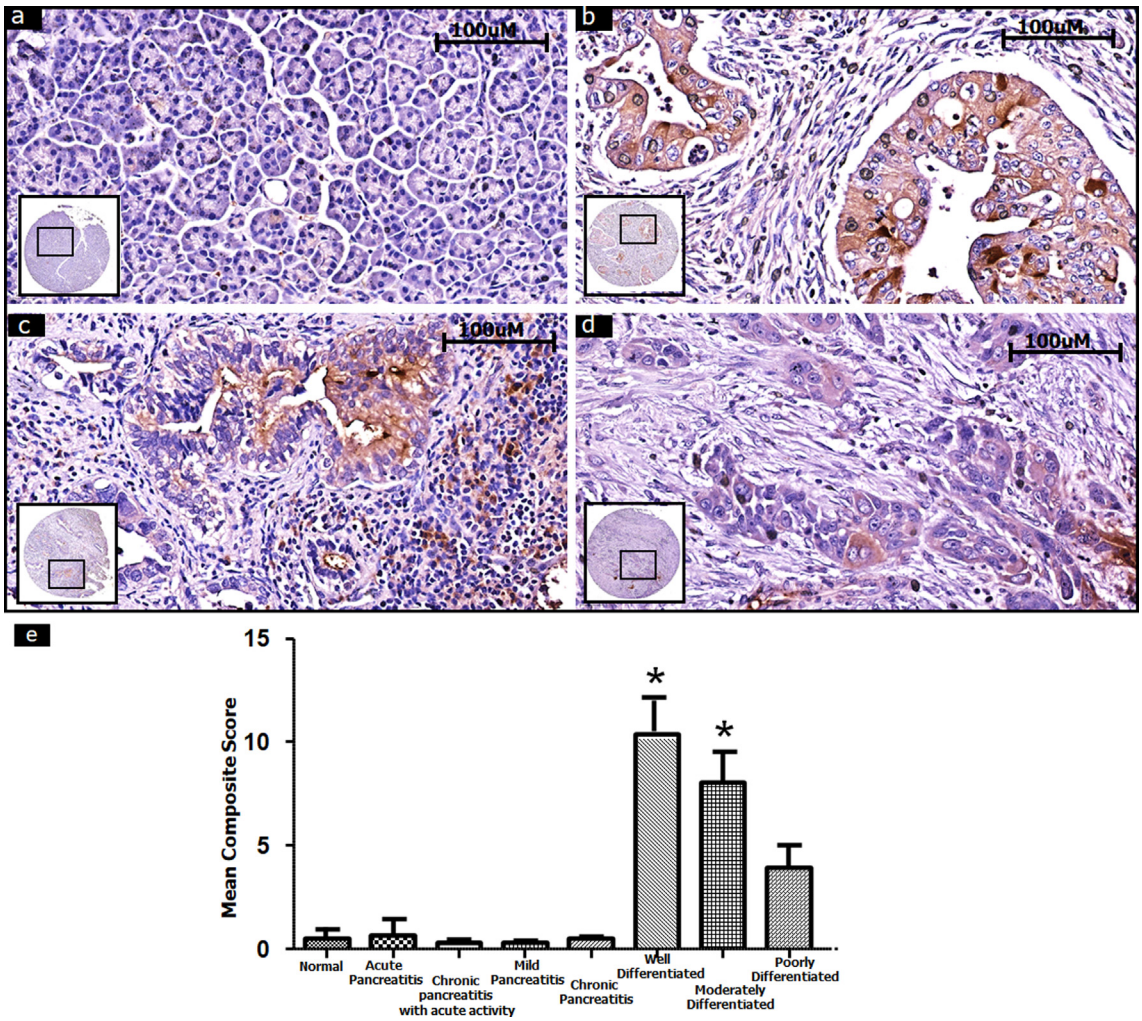


Fig. 8. IHC staining of commercially available Tumor Micro Arrays (TMAs) and its Mean Composite Scoring (MSC): Representative images of the tissues stained for CEACAM7 expression, depicted in color brown. Figure a, b, c, & d were 40× zoomed images of inset images (5–7.5× magnification full core of images) by using Case Viewer 2.4 software (3DHistech Ltd., Budapest, Hungary). In this IHC image (**a**) normal tissue with no staining; (**b**) well differentiated (high coverage and intense staining is present in duct); (**c**) moderately (high coverage but less intense staining is present in duct); and (**d**) poorly (low coverage and low intense staining is present in deformed duct) are the differentiated cores of PDAC samples, respectively. (**e**) **Mean Composite Scoring of CEACAM7:** Bar diagram of Mean Composite Score (MSC) of CEACAM7 expressions in various stages of pancreatitis, PDAC, and normal tissue cores.

ered as biomarkers to identify the diabetes associated PanCa. Few more enriched pathways are like arachidonic acid metabolism, [33], bile acids [34] and chemical carcinogenesis associated pathways are also enriched with CEACAM7 associated positive correlated genes [35]. Negatively correlated genes to CEACAM7 were also involved in colorectal cancer (CRC), as previous studies reported the downregulation of CEACAM7 in CRC [9]. These negatively correlated genes are also involved in some other important molecular processes like pyrimidine metabolism, autophagy, cell cycle, base excision repair, ribosome biogenesis in eukaryotes, DNA replication and oxidative phosphorylation, etc. The predictive positioning of CEACAM7 demonstrated its high expressions in ductal cells. In the molecular biology segment of study, we have opted for four different kinds of cell lines which simulate the broad range of pancreatic cancer tumor grades. HPNE was considered as a reference for normal pancreatic cell lines. HPAF-II (well differentiated cell line), SU86.86 (Moderately differentiated cell line) and PANC1 (Poorly differentiated cell line) were considered as differentiated grades of PDAC [19]. Using this experimental set, mRNA and protein expression profiling was performed. Real time PCR data demonstrated that the CEACAM7 expression in HPAF-II and SU86.86 was significantly higher compared to HPNE and PANC1. The same trend was observed at protein level in ELISA and western blotting analyses. In western blot, CEACAM7 was observed at 58KDa (it might be a dimer since the molecular weight mentioned in Uniprot (Q14002) is 29KDa). To confirm this, we looked for the structural information of CEACAM7, and found that CEACAM7 forms a dimer, which is tenfold tighter than that measured for CEACAM5 [7]. Additionally, a previously published article states that CEACAM5 was observed at 153KDa; however, as per Uniprot (P06731), the actual molecular weight is 76.79KDa. Therefore, we can anticipate that the dimer of CEACAM7 was observed in our western blot analysis.

To cross validate the results of western and ELISA, confocal microscopy was performed to visualize and check the localization of CEACAM7′s expression in the same set of experiments. Here we found that the results of confocal microscopy were precisely replicating the pattern of CEACAM7 expression as observed in mRNA and ELISA protein expressions. In HPAF-II, CEACAM7 was highly expressed in the cytoplasm (Red TRITC filter) and getting expression over nucleus (magenta color). While SU86.86 cells showed CEACAM7′s expression only in the cytoplasm, and very faint or no expression was observed in PANC1 and HPNE. All these results clearly depicted the expression profiling of CEACAM7 that gradu-

ally reduced from well differentiated grade of tumor to poorly differentiated grade of tumors. Although very less, CEACAM7 expression was observed in normal HPNE cell line at mRNA and no expression was observed in confocal microscopy. This result indicates that CEACAM7 can be used as an important biomolecular signature of early events of PDAC.

To cross validate our molecular cell biology data, IHC analysis was done on commercially available PDAC TMAs. Interestingly, well differentiated samples showed the highest expression followed by moderately differentiated, and poorly differentiated samples demonstrated very faint but higher than the normal, acute, and chronic pancreatitis samples. Considering these important indicators, we minutely scrutinized the histopathology of PDAC core. It was found that the expression intensity pattern was very high in well differentiated core especially at the apical, lateral, and basal part of the duct. Additionally, staining can be noted in some regions of the nucleus but mostly in the cytoplasm. However, the intensity pattern gradually fainted and diffused in moderate and poorly differentiated cores, the staining was observed more in cytoplasm. Consistency in our results from multiplatform assessments indicate significantly high levels of CEACAM7 expression in pancreatic cancer especially in early events. However, further studies in a bigger cohort are required. CEACAM7′s expression in early events of PanCa could be further exploited for molecular signature of PDAC. Moreover, it can work as a quality predictor for determining low versus high grade dysplasia.

Conclusion

This study suggests that CEACAM7 could be a promising biomolecule, contributing to the early events of PDAC and may potentiate the existing biomarker panel. This could be used in finding the vulnerable population, especially those who have a family history of PDAC. This development can also provide the perfect time frame to identify the risk population of PDAC early on to start the management protocols. The present study is of high clinical significance, as its layout offers a solid foundation for future investigation on CEACAM7 to determine its potential as a novel early detection biomarker of PDAC.

CRediT authorship contribution statement

Anupam Dhasmana: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing. **Swati Dhasmana:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing. **Sudhir Kotnala:** Software, Validation, Formal analysis, Investigation, Data curation. **Partha Laskar:** Methodology, Writing – review & editing, Visualization, Data curation. **Sheema Khan:** Methodology, Software, Validation, Formal analysis. **Shafiul Haque:** Methodology, Software, Validation, Formal analysis. **Meena Jaggi:** Supervision, Project administration, Writing – review & editing, Visualization, Funding acquisition. **Murali M. Yallapu:** Supervision, Project administration, Writing – review & editing, Visualization, Funding acquisition. **Subhash C. Chauhan:** Conceptualization, Methodology, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors gratefully acknowledge the start-up research grants from the Dept. of Immunology and Microbiology, School of Medicine, University of Texas Rio Grande Valley awarded to Meena Jaggi, Murali M. Yallapu, and Subhash C Chauhan. This work was supported by the NIH/NCI's funding R01s: CA210192; CA206069; CA204552; SC1GM139727 and Herb Kosten Foundation, Memphis. This study utilizes core facilities of CPRIT "Integrated Cancer Research Core (ICRC) (RP210180)" & "RP230419" and UT-System. Authors are also grateful to Ms. Stella Rios for the proofreading of this research article.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2023.02.013.

References

[1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Stat 2022, 2022,;72(1):7–33.
[2] https://www.cancer.org/cancer/pancreatic-cancer/detection-diagnosis-staging/survival-rates.html.
[3] Lippi G, Mattiuzzi C. The global burden of pancreatic cancer. Arch Med Sci 2020;16(4):820–4.
[4] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2020, CA: a cancer journal for clinicians 70(1) (2020) 7–30.
[5] Z. Yu, S. Zhao, Allicin depresses biological activities of pancreatic cancer cells by regulating miRNA-339-5p/ZNF-689 axis, 2021.
[6] Y. Cao, R. Zhao, K. Guo, S. Ren, Y. Zhang, Z. Lu, L. Tian, T. Li, X. Chen, Z. Wang, Potential metabolite biomarkers for early detection of stage-i pancreatic ductal adenocarcinoma 11 (2022).
[7] Bonsor DA, Beckett D, Sundberg EJ. Structure of the N-terminal dimerization domain of CEACAM7. Acta Crystallogr Sect F, Struct Biol Commun 2015;71(Pt 9):1169–75.
[8] Hammarström S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. Semin Cancer Biol 1999;9(2):67–81.
[9] Schölzel S, Zimmermann W, Schwarzkopf G, Grunert F, Rogaczewski B, Thompson J. Carcinoembryonic antigen family members CEACAM6 and CEACAM7 are differentially expressed in normal tissues and oppositely deregulated in hyperplastic colorectal polyps and early adenomas. Am J Pathol 2000;156(2):595–605.
[10] Raj D, Nikolaidi M, Garces I, Lorizio D, Castro NM, Caiafa SG, et al. CEACAM7 is an effective target for CAR T-cell therapy of pancreatic ductal adenocarcinoma. Clin Cancer Res: Off J Am Assoc Cancer Res 2021;27(5):1538–52.
[11] Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucl Acids Res 2017;45(W1):W98–W102.
[12] Chen F, Chandrashekar DS, Varambally S, Creighton CJ. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. Nat Commun 2019;10(1):5679.
[13] Nagy Á, Munkácsy G, Győrffy B. Pancancer survival analysis of cancer hallmark genes. Sci Rep 2021;11(1):6047.
[14] Liu L, Xu HX, Wang WQ, Wu CT, Xiang JF, Liu C, et al. Serum CA125 is a novel predictive marker for pancreatic cancer metastasis and correlates with the metastasis-associated burden. Oncotarget 2016;7(5):5943–56.
[15] Mehta S, Bhimani N, Gill AJ, Samra JS, Sahni S, Mittal A. Serum biomarker panel for diagnosis and prognosis of pancreatic ductal adenocarcinomas. Front Oncol 2021;11:708963.
[16] Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucl Acids Res 2017;46(D1).
[17] Wu B, Tang X, Ke H, Zhou Q, Zhou Z, Tang S, et al. Gene regulation network of prognostic biomarker YAP1 in human cancers: an integrated bioinformatics study. Pathol. Oncol. Res.: POR 2021;27:1609768.
[18] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science (New York, N.Y.) 2015;347(6220):1260419.
[19] Deer EL, González-Hernández J, Coursen JD, Shea JE, Ngatia J, Scaife CL, et al. Phenotype and genotype of pancreatic cancer cell lines. Pancreas 2010;39(4):425–35.
[20] Khan S, Zafar N, Khan SS, Setua S, Behrman SW, Stiles ZE, et al. Clinical significance of MUC13 in pancreatic ductal adenocarcinoma. HPB (Oxford) 2018;20(6):563–72.
[21] Massey AE, Doxtater KA, Yallapu MM, Chauhan SC. Biophysical changes caused by altered MUC13 expression in pancreatic cancer cells. Micron (Oxford England) 1993;130(2020):102822.
[22] Kumari S, Sikander M, Malik S, Tripathi MK, Hafeez BB, Yallapu MM, et al. Steviol represses glucose metabolism and translation initiation in pancreatic cancer cells. Biomedicines 2021;9(12).

[23] Gupta BK, Maher DM, Ebeling MC, Sundram V, Koch MD, Lynch DW, et al. Increased expression and aberrant localization of mucin 13 in metastatic colon cancer. J Histochem Cytochem 2012;60(11):822–31.
[24] Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the united states. Cancer Res 2014;74(11):2913.
[25] Noel M, Fiscella K. Disparities in pancreatic cancer treatment and outcomes. Health equity 2019;3(1):532–40.
[26] Kleeff J, Korc M, Apte M, La Vecchia C, Johnson CD, Biankin AV, et al. Pancreatic cancer. Nat Rev Dis Primers 2016;2:16022.
[27] Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. Lancet (London, England) 2016;388(10039):73–85.
[28] Tarasiuk A, Mackiewicz T, Małecka-Panas E, Fichna J. Biomarkers for early detection of pancreatic cancer - miRNAs as a potential diagnostic and therapeutic tool? Cancer Biol Ther 2021;22(5–6):347–56.
[29] Sellam F, Harir N, Khaled MB, Mrabent NM, Salah R, Benchouk A, et al. Delayed diagnosis of pancreatic cancer reported as more common in a population of North African young adults. J Gastrointest Oncol 2015;6(5):505–10.
[30] Striefler JK, Riess H, Lohneis P, Bischoff S, Kurreck A, Modest DP, et al. Mucin-1 protein is a prognostic marker for pancreatic ductal adenocarcinoma: results from the CONKO-001 study. Front Oncol 2021;11:670396.
[31] Yoshida K, Ueno S, Iwao T, Yamasaki S, Tsuchida A, Ohmine K, et al. Screening of genes specifically activated in the pancreatic juice ductal cells from the patients with pancreatic ductal carcinoma. Cancer Sci 2003;94(3):263–70.
[32] De Souza A, Irfan K, Masud F, Saif MW. Diabetes type 2 and pancreatic cancer: a history unfolding. JOP: J Pancreas 2016;17(2):144–8.
[33] Ding X-Z, Hennig R, Adrian TE. Lipoxygenase and cyclooxygenase metabolism: new insights in treatment and chemoprevention of pancreatic cancer. Mol Cancer 2003;2(1):10.
[34] Gál E, Veréb Z, Kemény L, Rakk D, Szekeres A, Becskeházi E, et al. Bile accelerates carcinogenic processes in pancreatic ductal adenocarcinoma cells through the overexpression of MUC4. Sci Rep 2020;10(1):22088.
[35] Bates RR. Chemical carcinogenesis and the pancreas. J Surg Oncol 1975;7(2):143–9.