Biomolecules and Biomedicine
ISSN: 2831-0896 (Print) | ISSN: 2831-090X (Online)
Journal Impact Factor® (2023): 3.1
CiteScore® (2023): 7.4
www.biomolbiomed.com | blog.biomolbiomed.com

RESEARCH ARTICLE **MOLECULAR BIOLOGY**

Molecular biomarkers involved in the progression of gallbladder inflammatory lesions to invasive cancer: A proteomic approach

Neetu Rawal[1], Gururao Hariprasad[2], Sabyasachi Bandyopadhyay[3], Nihar Ranjan Dash[4], Sunil Kumar[5], Prasenjit Das[6], Sharmistha Dey[2], Maroof Ahmad Khan[7], Amar Ranjan[1], Anita Chopra[1], Sundeep Saluja[8], Showket Hussain[9], G.K. Rath[10], Tanvir Kaur[11], and Pranay Tanwar[1*]

[1]Laboratory Oncology Unit, Dr. B.R.A. Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, India.

[2]Department of Biophysics, All India Institute of Medical Sciences, New Delhi, India.

[3]Proteomics Laboratory, Centralized Core Research Facility, All India Institute of Medical Sciences, New Delhi, India.

[4]Department of Gastrointestinal Surgery, All India Institute of Medical Sciences, New Delhi, India.

[5]Department of Surgical Oncology, All India Institute of Medical Sciences, New Delhi, India.
[6]Department of Pathology, All India Institute of Medical Sciences, New Delhi, India.
[7]Department of Biostatistics, All India Institute of Medical Sciences, New Delhi, India.
[8]Department of Gastrointestinal Surgery, Govind Ballabh Pant Hospital, New Delhi, India.
[9]Division of Molecular Oncology, National Institute of Cancer Prevention and Research, Indian Council of Medical Research, New Delhi, India.
[10]Dr. B.R.A. Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, India.
[11]Division of Non-Communicable Diseases, Indian Council of Medical Research, New Delhi, India.

***Corresponding author:** Pranay Tanwar; Email: pranaytanwar@aiims.edu

DOI: https://doi.org/10.17305/bb.2024.10704

Submitted: 05 May 2024/ **Accepted:** 17 June 2024/ **Published online:** 08 September 2024

Conflicts of interest: Authors declare no conflicts of interest.

Funding: This work was funded by the Indian Council of Medical Research (ICMR) (No. 5/13/1/TF/GBC/AIIMS/2015/NCD-III).

ABSTRACT

The progression of gallbladder inflammatory lesions to invasive cancer remains poorly understood, necessitating research on biomarkers involved in this transition. This study aims to identify and validate proteins associated with this progression, offering insights into potential diagnostic biomarkers for gallbladder cancer (GBC). Label-free liquid chromatography assisted tandem mass spectrometry (LC-MS/MS) proteomics was performed on samples from 10 cases each of GBC and inflammatory lesions, with technical duplicates. Validation was conducted through the enzyme-linked immunosorbent assay (ELISA) using 80 samples (40 GBC and 40 inflammatory lesions). Bioinformatics tools analyzed protein-protein interaction (PPI) networks and pathways. Statistical correlations with clinicopathological variables were assessed. Prognostic evaluation utilized Kaplan–Meier survival analysis and Cox regression analyses. mRNA expressions were studied using real time-polymerase chain reaction (RT-PCR). Out of 5,714 proteins analyzed, 621 were differentially expressed. Three upregulated (the S100 calcium-binding protein P [S100P], polymeric immunoglobulin receptor [PIGR], and complement C1q-binding protein [C1QBP]) and two downregulated (transgelin [TAGLN] and calponin 1 [CNN1]) proteins showed significant expression. Pathway analysis implicated involvement of proteoglycans in cancer and glycosaminoglycan metabolism. Significant correlations were observed between protein concentrations and clinicopathological variables. Prognostic factors such as tumor size, lymph node metastasis, and preoperative bilirubin levels were associated with overall survival. Protein-based assays demonstrated higher resolution compared to mRNA analysis, suggesting their utility in GBC risk stratification. S100P, PIGR, C1QBP, TAGLN, and CNN1 emerge as potential protein-based biomarkers involved in the progression from gallbladder inflammatory

lesions to invasive cancer. These findings hold promise for improved diagnostic and prognostic strategies in GBC management.

Keywords: Liquid chromatography assisted tandem mass spectrometry (LC-MS/MS); biomarker; differentially expressed proteins (DEPs); gallbladder cancer (GBC); enzyme-linked immunosorbent assay (ELISA); real time-polymerase chain reaction (RT-PCR); inflammatory lesion.

INTRODUCTION

Gallbladder cancer (GBC) is relatively rare but an aggressive malignancy that accounts for about 165,000 (1.7%) cancer deaths annually worldwide [1, 2]. The greatest incidence rates of GBC are found in Chile (27 per 100,000 people), with northern India following at 21.5 per 100,000 people. In India, GBC occurrence is significantly higher in the northern regions than in the southern states, with rates of 8.9 per 100,000 in Delhi compared to 0.8 per 100,000 in Chennai [3, 4]. The gallbladder is a small cystic organ that resides under the inferior surface of liver. Its main function is to store and concentrate the bile which is produced by the liver and transport the bile to the small intestine via the cystic duct [1]. The gallbladder is made up of a layer of tissue (a) mucosa (innermost layer of epithelial cells), (b) a muscular layer (smooth muscle cell layer) (c) perimuscular layer (connective tissue) (d) serosa (outer layer). The normal gallbladder epithelial layer shows metaplastic changes under the influence of Gallstones and other carcinogenic insults which further leads to dysplasia turning initially to carcinoma in situ (CIS), and finally invasive carcinoma. The multistage pathogenesis of gallbladder carcinoma arises due to prolonged exposure to gallstones (cholelithiasis) which creates an inflammatory environment (also, known as chronic cholecystitis) and increases the risk of GBC [5, 6]. The most common subtype of GBC is adenocarcinoma which accounts for 80 to 97% and arises from secretory cells. The papillary, mucinous, squamous, and adenosquamous are the other subtypes of GBC. The papillary subtype is considered as rarest GBC carcinoma with a better prognosis when compared to other subtypes [7]. Several risk factors include ethnicity, age, sex, chronic inflammation, gallstones, infections, exposure to heavy metals and environmental toxins, obesity, gallbladder polyps, genetic predispositions, and abnormalities in the pancreaticobiliary ductal junction. [8]. Till date, there is no diagnostic marker available for early detection of GBC. The combination of markers such as

CEA, CA125, CA242, and CA19-9 are used for the diagnosis of liver, gastric, colorectal, and pancreatic cancers. These markers when tested in GBC showed less specificity and sensitivity with inconsistent results and thus they cannot be used as a standalone marker for diagnosis of GBC [9, 10]. Identifying gallbladder carcinoma in its initial phases poses a challenge, despite improvements in ultrasound and computed tomography (CT) scanning, as merely half of gallbladder cancers are diagnosed prior to surgical intervention. The identification of GBC still relies on a clinical assessment, subsequently followed by minimally invasive imaging-guided fine needle aspiration cytology (FNAC). [11]. The radical surgical extirpation is the only effective treatment as majority of cases presents in advanced stage, since most of the time early detection is missed due to the fact that gallbladder lacks a sub mucosa layer to limit the spreading of cancer [12]. The gallbladder inflammatory lesion progresses into invasive cancer via metaplasia–dysplasia–carcinoma through recurrent epithelial damage. Thus, there is an impending need for research on biomarkers involved in the progression of gall bladder inflammatory lesions to invasive cancer. So, this study used comparative protein profiling between two clinical phenotypes of gall bladder (cancer and inflammatory lesion) that aimed to identify the differentially expressed protein panel which may be associated with the development of GBC in pre-existing inflammatory condition.

MATERIALS AND METHODS

Sample collection

The performa (questionnaire) was gathered for the clinical information from the patients. The tissues were collected from the resected gall bladder of patients during the surgery in 1XPBS vial and immediately stored at -80°C. The collected tissue was histopathologically checked for phenotype confirmation before further subsequent analysis. The percentage of tumor cells in each

tissue sample was meticulously assessed and recorded, and the samples containing more than 70% tumor cells in GBC cases were selected for further analysis.

Inclusion and exclusion criteria of patients

The staging of GBC was determined as per the American Joint Committee on Cancer (AJCC) cancer staging criteria. Patients were included in the study based on the following inclusion criteria: (a) Patients diagnosed with all types of the GBC (b) Patients diagnosed with inflammatory lesions of gall bladder. The exclusion criteria depend on (a) Patients who received chemotherapy/radiation therapy

Protein isolation and quantification

The homogenous samples from the two clinical phenotypes of the gallbladder which included stage 2, adenocarcinoma (n=10); and chronic cholecystitis lesions were taken for the discovery phase of proteomic experiments by the Liquid Chromatography assisted tandem Mass Spectrometry (LC-MS/MS) (Label-free). The study design is illustrated in (Figure 1). The protein was isolated by easy prep lysis buffer (Thermo Scientific, Catalog no. A45735). Around 5mg of gall bladder tissue was homogenized with 100µl of pre-warmed lysis buffer followed by the addition of 1µl of universal nuclease enzyme (Thermo Scientific, Catalog no. 88700). The homogenized tissue was centrifuged at 16000g for 10 minutes and the supernatant was collected in a vial. The protein concentration was determined through a UV-Vis spectrophotometer (Thermo Scientific, Nanodrop™) at A280nm wavelength with a baseline correction of 340nm.

Protein digestion and peptide cleanup

The 25ug of isolated protein was taken for protein digestion protocol (Thermo, Mini MS Sample kit, catalog A40006). 50µl of reduction and 50 µl of alkylation solution were added into the protein sample and incubated at 95℃ for 10 minutes. For protein digestion, 50µl of Trypsin/Lys C

protease mix was added to a reduced and alkylated sample and incubated overnight for 37°C. The Digestion reaction was stopped by the digestion stop solution. Further, the peptide clean-up protocol was followed as per the kit manual. Briefly, digested protein sample was transferred into the peptide desalting column and centrifuged. Flow through was discarded and the column was washed twice with a wash solution. Peptides were eluted and dried by lyophilization. The sample was resuspended in 0.1% formic acid for LC-MS/MS analysis.

Liquid Chromatography assisted tandem Mass Spectrometry (LC-MS/MS)

In the discovery phase, a total of 20 homogenous gall bladder tissues were used in each phenotype group. Ten samples of the GBC (stage 2, adenocarcinoma) and 10 samples from the inflammatory lesions (chronic cholecystitis) were taken up for label free LC-MS/MS experiment in technical duplicates. LC-MS/MS analysis was performed by Orbitrap Fusion tribrid mass spectrometer (Thermo Scientific, USA) connected with the ultra-pressure nano-flow liquid chromatography setup. 1 µg of peptide mixture was injected onto a reverse phase (RP) trap column, Acclaim PepMap 100 C18 (75 µm × 2 cm, 3 µm, 100 Å; Thermo Fisher Scientific) at a flow rate of 300 nl/min. 0.1% FA containing LC-MS grade water was used as loading buffer and 0.1% FA and 80% ACN containing LC-MS grade water was used as elution buffer. Peptides were washed with a loading buffer for 45 minutes to eliminate excess salt. Subsequently, the retained peptides were separated on an RP analytical column, specifically the Acclaim PepMap 100 C18 (75 µm × 15 cm, 2 µm, 100 Å; Thermo Fisher Scientific), prior to connection to the mass spectrometer. The elution gradient commenced with 5% elution buffer, gradually increasing at a linear rate of 8% over 5 minutes, followed by increments to 60% over 110 minutes, and finally to 95% over 2 minutes. Maintenance of the gradient at 95% elution buffer occurred for 5 minutes before re-equilibration with a 5% elution buffer for 20 minutes. Throughout the LCMS/MS analysis, the loading buffer

was utilized as a blank, while a tryptic digested Hela standard (200ng) served as the QC standard. A total of 1µg pooled peptide sample acted as the control for the LC-MS/MS run. The mass spectrometer operated under data dependent acquisition (DDA) mode. Full MS spectra were collected in positive ionization mode, with an ion-spray voltage of 2100 V and m/z ratio ranging from 350 to 2000 Da, using a 50-millisecond injection time. In the MS1 level, precursor/peptide isolation employed a quadrupole ion filter and orbitrap mass analyzer, set at a resolution of 60,000 (at 200 m/z) and AGC target of 1e6. The top 20 precursors were chosen for fragmentation (MS2 or MS/MS) via a linear ion trap, with a resolution setting of approximately 30,000 (at 200 m/z) and an AGC target of 1e5. DDA incorporated advanced "rolling collision energy" for subsequent MS/MS scans, with the normalized high energy collision-induced dissociation (HCD) fragmentation energy fixed at 30%.

Database search and analysis

The precision and accuracy of the raw data generated through the LC-MS/MS instrument were checked using X Caliber software. All 40 files obtained were normalized and analyzed using Proteome Discoverer 2.4 SP1 (Thermo Scientific) [13]. This software was used to analyze all raw files collectively, enabling the identification of differentially expressed proteins (DEPs) with more than 30% coverage across all 40 samples. To ensure accuracy and the identification of efficient biomarkers, DEPs that were consistently present across all samples were filtered and selected. Data is available via Proteome X change consortium through MassIVE data repository with the identifier PXD040704. Both the MS and MS/MS spectra were searched against the human Uniprot database appended to a list of common contaminants provided by thermo scientific using the Sequest HT algorithm. For each spectrum file, the spectrum files RC node calculates a constant mass shift (ppm) and the spectrum selector node selects the subset of the set of spectra that will be

searched in Sequest HT. The Sequest HT parameters were specified as trypsin enzyme, two allowed missed cleavages, 6 as minimum peptide length, 144 as maximum peptide length, 10 ppm as precursor mass tolerance, and 0.6 daltons as fragment mass tolerance. The static modification was set to carbamidomethylation (+57.021 Da) of cysteine. The dynamic modification applied was methionine oxidation (+15.995 Da) on the peptide terminus. Additionally, N-terminal modifications were implemented: Acetyl (+42.011 Da), Met-loss (-131.040 Da), and Met-loss+Acetyl (-89.030 Da). In the Percolator node, the false discovery rate (FDR) was determined using q-values from the Decoy database search. Peptide spectral match filtering was conducted with a stringent FDR threshold of 0.01 and a lenient threshold of 0.05, as determined by the percolator. Peaks and features were filtered using a minora feature detector node, with parameters set at a minimum trace length of 5 and a maximum ΔRT of isotope pattern multiplets of 0.2. Contaminant and decoy proteins were excluded from all datasets before downstream analysis, and MSF files were processed through the consensus workflow. In the consensus workflow, peptide spectrum matches were grouped with the site probability threshold of 75 and the peptide validator node was strictly set at the target FDR of 0.01 and relaxed at 0.05 for PSMs and peptide identification. The peptide and protein filter node was set to minimum peptide length of 6, peptide confidence at a least high, minimum number of peptide sequences at 1, count only rank 1 peptide as false, count peptides only for top scored protein as false for protein filter according to score thresholds. Further, proteins were scored, grouped and peptides in the protein was annotated depending on the position of the identified peptides and the proteins found in a sample. The number of falsely identified proteins were filtered out which was determined by protein FDR validator node strictly set at a target FDR of 0.01 and relaxed FDR of 0.05. The remaining proteins were annotated in respect of their biological process, cellular component, and molecular function. The

final proteins were marked as master protein along with their modification sites and peptide isoform group. The feature mapper node which performed a retention-time alignment and feature linking across data sets was set at the parameters of (a) mass tolerance: 10ppm, (b) maximum RT shift: 10 min, (c) minimum S/N threshold: 5. The precursor ion quantifier node controlled the peptide for their quantification based on the parameters set to (a) consider protein groups for peptide uniqueness as true, (b) precursor abundance based on intensity, (c) normalization mode of total peptide amount, (d) Pairwise ratio based protein ratio calculation, (e) maximum fold change of 100. The Log_2 fold change of ≥ 1 and ≤ 1 at $p < 0.05$ was set for the identification of differentially expressed proteins. A volcano plot with a p value <0.05 was generated graphically to represent the upregulated and downregulated proteins.

Enzyme-linked immunosorbent assay (ELISA*)*

The protein lysate from an independent set of gallbladder tissues (n=80; 40 cancerous and 40 inflammatory lesions) was prepared by the easy prep lysis buffer (Thermo Scientific). All the 5 differentially expressed proteins (S100P, C1QBP, PIGR, TAGLN, and CNN1) expressed at log_2 fold change ≥ 2 were assayed by following the manufacturer's protocol by a commercially available ELISA Kit (ELK, Catalog no. 1290, 1865, 1752, 3895, 3478). In summary, 96-well plates were subjected to an 80-minute incubation at 37°C with standard, sample, and blank duplicates. Following this, the liquid was removed from the microwells, and they were washed three times with 200 µL of wash buffer. Subsequently, diluted Biotin-Conjugate antibodies (100 µL) were added to all wells and incubated at 37°C for 50 minutes. After another washing step, diluted Streptavidin-HRP (100 µL) was added and incubated for 50 minutes at 37°C. Color development was initiated by adding 100 µL of TMB substrate solution to all wells and incubating at 37°C for 20 minutes. Finally, 50 µL of the stop reagent was added to all wells, and the

absorbance readings of both standards and samples were measured at 450nm, with readings tabulated according to their dilutions.

Pathway and protein-protein interaction (PPI) network analysis

The PPI network of the identified DEPs was generated through the STRING database (https://string-db.org/) [14] with the parameters set to 5% FDR stringent and species as homo-sapiens. Further, the significance of hub proteins in the interactome was calculated based on twelve topological methods (Closeness, Degree, Maximum neighborhood component (MNC), Maximum clique centrality (MCC), Edge percolated component (EPC), Bottleneck, EcCentricity, Density of maximum neighborhood component (DMNC), Betweeness, Radiality, Stress, and clustering coefficient) through cytohubba, a cytoscape plugin (https://cytoscape.org/) [15]. The pathway and Gene Ontology (GO) enrichment analyses of DEPs were studied through DAVID (https://david.ncifcrf.gov/tools.jsp) [16] software which includes three databases (KEGG, Reactome, and Wiki Pathway).

RNA isolation and cDNA synthesis

Total RNA was isolated from 80 gall bladder tissues, consisting of 40 cancer cases and 40 inflammatory controls, using the RNA later method with the Qiagen RNeasy Mini Kit according to the manufacturer's instructions. To ensure the RNA's suitability for downstream applications, purity and concentration were checked using a Nanodrop spectrophotometer and a Qubit fluorometer, respectively. Additionally, the integrity of the RNA samples was assessed on a 1.5% agarose gel and with the Bioanalyzer 2100 (Agilent Technologies Inc., USA). Only samples with an RNA Integrity Number (RIN) greater than 7.0 were used for cDNA synthesis followed by RT-PCR. Subsequently, reverse transcription was performed using the Improm II RT system from Promega to generate cDNA. Quantitative PCR was then conducted using SYBR green dye and

Oligo(dT) primer on the Agilent Mx3000P qPCR Platform from Agilent, Santa Clara, CA, USA. The specific primer sequences of S100P are Fwd: AGGTGCTGATGGAGAAGGAG, Rev: ACTCACTGAAGTCCACCTGG, C1QBP-Fwd:GGAGCTGGAACTGAATGGGA, Rev: GTTGGTGGGATGCTGTTGTT, PIGR-Fwd: Fwd: GAAAGGGCTCGGGACGATGG , Rev: TCTTCGTGGAGATGGCTGGGA, TAGLN- Fwd: GAGAGATGAGGATGGAGGCC, Rev: AGGATTGCTGCCAGAGAAGT, CNN1- Fwd: AGGTTAAGAACAAGCTGGCCC, Rev : CCGTCCATGAAGTTGTTGCC and *GAPDH* was utilized as the housekeeping gene, possessing the following sequence. of Fwd: TCGTGGAAGGACTCATGACC, Rev: ATGATGTTCTGGAGAGCCCC. The calculation of relative mRNA expression for all five genes involved utilizing the $2^{-\Delta Ct}$ equation.

Ethical statement

The institute ethics committee at All India Institute of Medical Sciences, New Delhi, granted approval for this study. (IECPG-608/25.11.2020,RT-25/23.12.2020), and protocols followed were in accordance with the ethical standards formulated in the Helsinki Declaration. Patients were informed in detail about the clinical study before screening at the GI surgery department, AIIMS, New Delhi. Voluntary written informed consent was obtained from each patient before recruitment.

Statistical analysis

Continuous variables were described using x±s and the median (range). Student's t-test was applied to ascertain the concentrations of S100P, PIGR, C1QBP, TAGLN, and CNN1 in both cancer cases and inflammatory lesions. The ROC curve analysis was utilized to determine the optimal cut-off value for the concentrations of S100P, PIGR, C1QBP, TAGLN, and CNN1 using the Youden index method. [17]. The chi-square test was used to check the association between

protein expression and clinic-pathological characteristics at optimal cut-off. Student's t-test was used to determine the S100P, PIGR, C1QBP, TAGLN, and CNN1 mRNA expression in cancer cases and inflammatory lesions. The Kaplan–Meier method with a log-rank test was used to predict the overall survival of GBC cases. The prognostic factors associated with survival were evaluated by univariate and multivariate Cox proportional hazard regression analysis. All the statistical analyses were performed with the Graphpad prism 8.0 and STATA version 11. $p<0.05$ was considered statistically significant.

RESULTS

Clinical profile of discovery phase

In the discovery phase, a total of 20 homogenous gall bladder tissues were used in each phenotype group. Ten samples of the GBC (stage 2, adenocarcinoma) and ten samples from the inflammatory lesions (chronic cholecystitis) were taken up for label-free LC-MS/MS experiment. Out of 10 cancer cases, 8 were females and 2 were males with a median age of 56 years. And none of the cases had a history of gallstones in cancer cases. In inflammatory lesion cases, 6 were females and 4 were males with a median age of 51 years. The three patients had a history of gallstones.

Differentially expressed protein identified through LC-MS/MS

The LC-MS/MS raw files (n=40) obtained from the samples were viewed and validated through X calibre software. The quality of the data is shown by the normalization curve and PLsDA-2D score plot in (Figure S1). In the raw output data, total 5714 proteins, 31675 peptides, 479279 PSMs, and 1569659 MS/MS spectrum were obtained through proteome discoverer software. A total of 3204 proteins were filtered out through the following initial parameters (a) Master proteins and contaminants are false, (b) Unique peptides greater than or equal to 2. The final set of DEPs was filtered with the additional parameter of proteins found to have at least confidence and peak

in every sample. Out of 621 proteins, total of 18 proteins were significantly differentially expressed at log_2 fold change of 1 at $p<0.05$ in which 3 were upregulated and 15 were downregulated as shown in (Table S1). Total of three significant upregulated proteins (S100P, PIGR, C1QBP) and two downregulated proteins (TAGLN and CNN1) were expressed at log_2 fold change ≥ 2 and the remaining 13 downregulated proteins were expressed at log_2 fold change ≤ 2 as represented in the volcano plot (Figure 2).

PPI and pathway analysis

The protein-protein interaction (PPI) network of 18 DEPs is represented in (Figure 3). TAGLN and CNN1 are the two most significant nodes identified in the PPI network with the highest score through cytohubba in the majority of the topological analysis methods as shown in (Table S2). A total of 15 pathways were found associated with DEPs as shown in (Figure 4a). Out of 15 pathways, 13 were from the Reactome database, 2 were from the wiki pathway database, and 1 from the KEGG database. The kyoto encyclopedia of genes and genomes (KEGG) database highlights the role of DEPs in the pathways of proteoglycans in cancer. The top pathways involved with the highest protein count in the reactome database were related to diseases associated with glycosaminoglycan metabolism, glycosylation and other related metabolism. The wiki database showed that DEPs discovered in our study are contributing a major role in burn wound healing pathway. The Gene Ontology for Cellular Component (GO_CC) enrichment that was carried out for the 18 DEPs showed that the majority of the proteins belong to the cytoplasm and extracellular exosome as represented in (Figure 4b). The extracellular space is the next most enriched cellular component with the highest protein count. The Gene Ontology for Biological process (GO_BP) and Molecular function (GO_MF) analysis showed that the enrichment proteins belongs to the

actomyosin structure organization and extracellular matrix structural constituent conferring compression resistance respectively.

Validation by ELISA

The DEPs which were higher and equal at log_2 fold change of 2 were further taken up for validation phase which included three up regulated proteins (S100P, PIGR, C1QBP) and two down regulated proteins (TAGLN and CNN1). A total of 80 independent technical duplicate set of samples (40 cancer cases and 40 inflammatory lesions) were used for validation. The protein concentration levels were tested by ELISA in the tissue lysate. The mean value of the S100P, PIGR, C1QBP, TAGLN, and CNN1 levels in GBC cases were 5.097 ± 0.3496 ng/mL, 1334 ± 67.55 pg/mL, 10.89 ± 0.9557 ng/mL, 9.377 ± 0.8957 ng/mL, 20.63 ± 2.082 ng/mL respectively, and in the inflammatory lesions were 3.546 ± 0.2538 ng/mL, 1089 ± 43.91 pg/mL, 7.010 ± 0.5599 ng/mL, 19.70 ± 1.294 ng/mL, 27.24 ± 1.353 ng/mL respectively. A significant difference was observed in the S100P, PIGR, C1QBP, TAGLN, and CNN1 protein concentration levels in cancer and inflammatory lesions ($p=0.0006$, $p=0.0032$, $p=0.0008$, $p<0.0001$, $p=0.0094$ respectively) as shown in (Figure 5).

The receiver operating characteristic (ROC) was established for signature proteins and the area under the curve AUC [95% confidence interval (CI)] obtained were statistically significant as shown in (Figure 6). The determined optimal cut-off value (sensitivity and specificity) of each signature protein through Youden index were S100P: 5.35 ng/µg (52.5% and 87.5%), PIGR: 1068 pg/µg (80%, 57.5%), C1QBP: 10.32 (52.5% and 92.5%), TAGLN: 12.66 ng/µg (70% and 92.5%) and CNN1: 19.79 ng/µg (70% and 90%) which is further used for clinicopathological correlation as shown in (Table 1).

Association of DEPs with overall survival in GBC cases

An examination of survival curves using the Kaplan-Meier method was conducted among 40 cases of GBC, comparing concentrations of S100P, PIGR, C1QBP, TAGLN, and CNN1 through log-rank testing. Overall survival (OS) was defined as the duration from treatment initiation to death, with patient follow-up concluding on February 28, 2023, and a median follow-up duration of 16 months for overall survival. The study employed an optimal concentration threshold for all five DEPs to predict overall survival based on high/low expression levels. It has been observed that a high expression of DEPs in cancer cases has relatively poor overall survival when compared to a low expression of DEPs as shown in (Figure 7). However, S100P, PIGR, C1QBP, TAGLN and CNN1 protein concentration showed no statistical significance (p=0.377, p=0.9206, p=0.7967, p=0.1393, p=0.5354, respectively) with overall survival.

Analysis of prognostic factors in GBC cases using univariate and multivariate approaches

Prognostic factors were analyzed through univariate and multivariate methods to predict overall survival, as detailed in Table 2. According to the univariate analysis, certain factors such as tumor size (T1/T2 vs. T3/T4), lymph node metastasis (N0 vs N1/N2), presence of distant metastasis, total bilirubin levels (mg/dL), and unconjugated bilirubin levels (mg/dL) exhibited significant associations with poor overall survival (p=0.005, p=0.0001, p=0.001, p=0.006, p=0.04, respectively). Conversely, the remaining prognostic factors did not show significant correlations with overall survival. Further, all the five signature proteins along with the significant univariable parameters were taken up for multivariate cox regression analysis and observed that the lymph node metastasis was significantly (p=0.004) associated with the poor overall survival in GBC cases.

Relative mRNA expression level of DEPs in GBC and inflammatory lesions

The S100P, PIGR, C1QBP, TAGLN, and CNN1 relative mRNA expression were checked among the matched 80 cases which included 40 GBC and 40 inflammatory lesions. A significant correlation was found between the S100P mRNA expression level in cancer and inflammatory lesions ($P = 0.0458$). It was noted that mRNA fold change levels in cancer cases exceeded those in inflammatory lesions across all five genes studied. However, no statistical significant correlation was found in PIGR, C1QBP, TAGLN, and CNN1 mRNA expression level in cancer and inflammatory lesions of the gallbladder ($P = 0.575$, $P = 0.594$, $P = 0.401$, $P = 0.320$, respectively), as shown in (Figure 8).

DISCUSSION

The aggressiveness of GBC and relative paucity of biological markers is the driving factor for this study. The study has focused on protein-based markers with an overall goal of community-based prevention and early detection in pre-existing inflammatory lesions. Our study appears to be the first comprehensive study which has analysed the signature proteins followed by their validation based on independent sets of protein lysate with their corresponding RNA based expression. Till date, there are only two previously published studies that establish the diagnostic marker for GBC by comparing potentially expressed protein in both inflammatory conditions and cancer. The first study (Huang et al., 2014) from china [18] identified that Annexin A4 was found upregulated, Hsp90B and Dync1h1 were found downregulated in the GBC (n=10) when compared it with the inflammatory lesions (n=10) through the 2DE-MALDI-TOF technique. The second is an Indian study (Sahasrabuddhe NA et al., 2014) [19] used LC-MS/MS (iTRAQ) technique and identified that prosaposin was upregulated and transgelin (TAGLN) was downregulated in GBC (n=10) in comparison to normal gallbladder tissue (n=10). Both the studies have the following limitations

(a) This study (Huang et al., 2014) was limited to discovery phase only with no further clinical validation (b) Clinical phenotypes of the sample used in the studies were not homogenous (c) Findings were not clinically correlated in the studies (d) Protein-Protein interaction analysis and functional role of identified proteins in the cancer were not established.

Regarding the presence of gallbladder inflammation and its link to cancer development, it's crucial to recognize that while gallstones commonly cause inflammation, they aren't the sole contributor. Other factors also play a role. Our focus is on chronic inflammation in various forms, including acalculous cholecystitis (inflammation without stones), inflammation due to chronic typhoid infection, and xanthogranulomatous cholecystitis. Incidental gallbladder cancer (IGBC) cases identified without suspicion of malignancy underscore this point, with an incidence ranging from 0.19 to 3.3% [20]. These diverse inflammatory conditions of the gallbladder, regardless of the presence of stones, are vital considerations in understanding GBC pathogenesis.

In our study, we used a label-free LC-MS/MS technique on 20 homogenous tissues (10 cancer cases and 10 inflammatory lesions) for the discovery phase. At a cutoff of log2 fold change ≥2, we identified three significantly upregulated proteins (S100P, PIGR, C1QBP) and two downregulated proteins (TAGLN and CNN1). The remaining 13 proteins were identified as downregulated at log2 fold change ≤ 2. The five differentially expressed proteins (S100P, PIGR, C1QBP, TAGLN, and CNN1) were validated by ELISA in the 80 independent set of gallbladder tissue samples (40 cancer cases and 40 inflammatory lesions). A significant difference was observed in the protein concentration levels of S100P, PIGR, C1QBP, TAGLN, and CNN1 between cancer and inflammatory lesions ($p=0.0006$, $p=0.0032$, $p=0.0008$, $p<0.0001$, $p=0.0094$, respectively). Exposure to carcinogens can potentially transform ordinary gallbladder epithelium into a state known as metaplasia, leading to the development of dysplasia and eventually carcinoma

in situ (CIS), which can progress to invasive carcinoma. It's worth noting that over 90% of individuals diagnosed with gallbladder carcinoma exhibit signs of dysplasia and CIS. [21]. In our study, the low expression of upregulating proteins in inflammatory lesions signifies the ongoing neoplastic process which is beyond morphological detection limits of the available modality.

To investigate the transcriptional profile of signature proteins i.e. S100P, PIGR, C1QBP, TAGLN, and CNN1 in carcinogenesis, we examined their corresponding mRNA expression in both of the clinical phenotypes (GBC and inflammatory lesion). The primary objective was to identify corresponding fold changes among mRNA with its level of protein expression. Even though a comparable difference is noted in both the clinical phenotype at the mRNA expression level of all the genes, the majority of the samples showed negligible fold change, which may or may not be detected by conventional real-time PCR. Therefore, our study results confirmed that a protein-based assay has relative higher resolution for detection of change of protein expression compared to mRNA expression and may help in risk stratification with better precision.

The Patients' clinicopathological characteristics of validation phase samples (n=80) were correlated with S100P, PIGR, C1QBP, TAGLN, and CNN1 concentration at their respective optimum cut-off values (5.353 ng/ µg, 1068 pg/ µg, 10.32 ng/µg, 12.66 ng/µg and 19.79 ng/ µg respectively) as shown in Table 1 and Table S4. The signal transduction mechanisms of signature proteins were established based on our clinical outcome and in reference to previously published studies on identified proteins.

S100P: This signaling molecule mediates multiple transduction pathways through Ca2+ ion activation, as depicted in (Figure S2). In our study, S100P expression was significantly correlated with the clinical phenotype of gallbladder conditions (cancer vs. inflammatory lesions; p=0.0001), differentiation (well/moderate vs. poor; p=0.0491), calcium levels (p=0.039), preoperative

bilirubin levels (conjugated; p=0.0189), and CEA marker (p=0.04). These clinical findings align with the outcomes of our study and are consistent with established pathways in GBC. Several studies have elucidated the role of S100P as a diagnostic marker in cancers. For example, Aishima et al., 2016 reported that S100P expression is associated with the progression from low-grade to high-grade biliary intraepithelial neoplasia and serves as a strong early detection marker for cholangiocarcinoma [21]. Another study by Mathai et al., 2021 found that S100P overexpression is strongly correlated with GBC advancement and poor survival [22]. Additionally, one study suggested that LASP-1 and S100P are two therapeutic targets that inhibit GBC aggressiveness and metastasis [23].

Parkila et al., 2008 [24] evaluated S100P protein and corresponding mRNA expression levels in normal and tumor tissues of various organs through immunohistochemistry (IHC) and Real-time PCR, respectively. They found that S100P protein expression was highly elevated in all tumor tissues, with the most prominent expression observed in gastric tumors. The authors [24] suggested that the high expression level of S100P in tumor tissues could serve as a potential target marker for diagnostic applications. Consistent with the above-published studies, our results also demonstrated a significant correlation between S100P mRNA expression levels in cancerous and inflammatory lesions of the gallbladder (p=0.0458). Thus, our study indicates that high expression of S100P protein and mRNA is found in GBC compared to inflammatory lesions of the gallbladder.

PIGR: The polymeric immunoglobulin receptor (PIGR) is a transmembrane protein involved in cancer signaling pathways, as illustrated in Figure S3. In the current study, PIGR expression was significantly correlated with the clinical phenotype (cancer vs. inflammatory lesions; p=0.0006) and AFP (p=0.045). Similar results are available in the Human Protein Atlas (HPA) database

(https://www.proteinatlas.org/), which indicates high PIGR expression in cancerous tissues of the gastrointestinal mucosa, kidney, gallbladder, and urinary bladder compared to non-cancerous tissues. Increased PIGR expression was also observed in the gastrointestinal tract and hepatocellular carcinoma [25]. Our study showed higher expression of PIGR in GBC tissues compared to inflammatory lesions of the gallbladder, suggesting its potential as an early detection marker in pre-existing inflammatory lesions.

A previously published study by Okhuma et al. (2020) [26] compared PIGR mRNA (data downloaded from the TCGA database) and protein (IHC) expression in pancreatic cancer. It was observed that expression was higher in the treated group versus the untreated group, further supporting that higher levels of PIGR mRNA and protein were independent prognostic factors. To the best of our knowledge, our study is the first to assess PIGR mRNA and protein levels in clinical samples. The results of our study are consistent with previously published findings.

C1QBP: Complement C1q binding protein, is involved in signaling pathways of cancer, as depicted in Figure S4. Our study revealed a significantly high expression of C1QBP in GBC compared to inflammatory lesions ($p<0.0001$). Previous studies (Chen et al., 2009; Niu et al., 2015; Wang et al., 2015; Gao et al., 2016) [27-30] have suggested that C1QBP acts as a diagnostic marker in cancer patients and is related to metastasis, progression, and poor overall survival. Our study on GBC aligns with these established findings, reporting a significant correlation between C1QBP and lymph node metastasis ($p=0.0010$), tumor differentiation (well/moderate vs. poor; $p=0.0491$), cholelithiasis ($p=0.0139$), and preoperative bilirubin levels (conjugated; $p=0.0408$ and unconjugated; $p=0.0406$).

In a study by Shen et al., 2014 [31], the C1QBP mRNA and protein expression were compared in cholangiocarcinoma cell lines (4 cell lines) and normal cell lines (1 cell line) using RT-PCR and

western blot, respectively. They found significantly higher C1QBP mRNA and protein expression in the cholangiocarcinoma cell lines compared to the normal cell lines. Our study is unique as there is no published data available to date that reports C1QBP protein and mRNA expression in tissue lysate in GBC cases.

TAGLN: Transgelin is an actin-binding protein that serves as a marker for smooth muscle differentiation [32]. The gradual loss of TAGLN function contributes to tumor progression and serves as a diagnostic marker in breast and colon cancer development [33, 34]. Our study also observed a similar trend, reporting lower expression of TAGLN in GBC compared to inflammatory lesions ($p<0.0001$). Additionally, we found that TAGLN depletion inversely correlates with preoperative bilirubin levels (unconjugated; $p= 0.003$), which serves as an indicative marker of tumorigenicity in cancer cells. The signal transduction pathway involved in cancer cells is illustrated in Figure S5.

A study by Tsui et al. in 2019 [35] compared the expression levels of TAGLN in bladder carcinoma cells to normal bladder tissues using RT-PCR and western blot analysis. They observed higher mRNA and protein expression levels of TAGLN in normal tissues compared to carcinoma cells. In our study, we found that the fold change level of mRNA in inflammatory lesions was relatively higher compared to the fold change level of mRNA expression in cancer cases. However, no significant correlation was found.

CNN1: The calponin protein plays a crucial role as a cytoskeletal protein and mediator of signal transduction, as illustrated in Figure S6. Several studies have identified CNN1 as a tumor suppressor, noting its decreased expression in various cancer types such as ovarian cancer, hepatocellular carcinoma, breast cancer, and colorectal cancer (CRC) [36-39]. Consistent with these findings, we also observed low expression of CNN1 in GBC compared to inflammatory

lesions. Furthermore, our significant correlation findings between clinical phenotypes (cancer vs. inflammatory lesions; $p<0.0001$), urea ($p=0.0077$), preoperative bilirubin level (unconjugated; $p=0.0106$), globulin ($p=0.0001$), CEA ($p=0.003$), and AFP ($p=0.05$) with CNN1 expression level suggest its involvement in cell invasion and progression through pre-existing inflammatory conditions.

A previously published study by Mamoor et al. [38] elucidated the role of CNN1 mRNA and protein expression in breast cancer tissues and adjacent normal tissues using a dataset from GEO2R. They reported decreased CNN1 expression in tumor tissues compared to adjacent normal tissues, correlating with poor overall survival in patients.

Our study is likely the first to report TAGLN and CNN1 protein expression and their corresponding mRNA expression in tissue lysates of GBC cases and inflammatory lesions.

As per Kaplan–Meier curve analysis, high S100P, PIGR, C1QBP, TAGLN and CNN1 protein concentration in GBC cases results in poor overall survival ($p=0.377$, $p=0.9206$, $p=0.7967$, $p=0.1393$, $p=0.5354$, respectively). Further, the univariate and multivariate analysis in GBC cases were studied to investigate the prognostic factors involved in overall survival. In the univariate analysis, the tumor size (T1/T2 vs. T3/T4), lymph node metastasis (N0 vs N1/N2), distant metastasis (with vs. without metastasis) showed a significant correlation with poor overall survival ($p=0.005$, $p=0.0001$, $p=0.001$ respectively). Few studies [40-42] also elucidated that tumor size is related to poor overall survival in GBC patients. Further, studies (Naveed et al., 2021, Shirai et al., 2012, Liu et al., 2020, Sachan et al., 2020, Xu et al., 2018) [43-47] reported that lymph node metastasis and distant metastasis are the most important prognostic factors. Our study results showed a significant correlation of preoperative lebel of bilirubin (total); mg/dL and bilirubin (unconjugated); mg/dL with poor overall survival ($p=0.006$, $p=0.04$ respectively). It was also

observed that the hazard ratio of tumor size increases (4.8 to 7.7) with the high bilirubin (total) level in the adjusted univariate analysis. Similar to our results, three previously published studies (Bandirmali et al., 2020, Farhat et al., 2008, Tran et al., 2017) [48-50] reported that a high level of preoperative bilirubin is an independent prognostic marker for poor overall survival in GBC. In the multivariate analysis, lymph node metastasis (N0 vs. N1/N2) revealed a significantly worse prognosis in cancer cases (p= 0.004**)**. In our multivariate analysis, we observed a significant correlation between the 5 DEPs and both tumor size (p-value: 0.09) and lymph node metastasis (p-value: 0.04). These findings suggest that the 5 DEPs, along with other parameters, impact patient survival by increasing the hazard ratio for tumor size from 4.89 to 34.22. When all five biomarkers were evaluated together for their impact on overall survival, TAGLN emerged as significant (p=0.049), indicating its potential as a prognostic marker. However, when the upregulated proteins (S100P, PIGR, C1QBP) and downregulated proteins (TAGLN, CNN1) were evaluated separately, neither group showed a significant impact on overall survival. Furthermore, our adjusted univariate analysis revealed that the hazard ratio for tumor size increases with the level of bilirubin from 4.89 to 5.7. Based on all the above results, this suggests a potential unexplored relationship between the expression level of the DEPs and bilirubin levels, which may contribute to the progression of gallbladder inflammatory lesions to invasive cancer.

The identified signature proteins need to be further validated in the serum/plasma for clinical diagnostic use in future in large cohort. This may establish the strong diagnostic parameters which may be missed due to tissue heterogeneity of samples and low sample size used in validation phase.

CONCLUSION

This study identified a panel of five protein-based diagnostic biomarkers (S100P, PIGR, C1QBP, TAGLN, and CNN1) potentially involved in the progression of gallbladder inflammatory lesions

to invasive cancer. The signal transduction mechanism established the role of these signature proteins in cancer metastasis and invasiveness. However, due to the small sample size, further analysis with a larger cohort is necessary to establish these biomarkers for diagnostic purposes. The study also suggests that protein-based assays may provide better resolution for GBC risk stratification compared to mRNA-based assays for future clinical use.

ACKNOWLEDGMENTS

The authors are thankful to the Proteomics lab at Central Core Research Facility, AIIMS for the Mass spectrometric analysis.

Data availability

Data is available via Proteome X change consortium through MassIVE data repository with the identifier PXD040704/ MSV000091448

REFERENCES

1. Abou-Alfa, G. K., Jarnagin, W., El Dika, I., D'Angelica, M., Lowery, M., Brown, K., … O'Reilly, E. M. (2019). Liver and Bile Duct Cancer. In *Abeloff's Clinical Oncology*. https://doi.org/10.1016/B978-0-323-47674-4.00077-3

2. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. https://doi.org/10.3322/caac.21492

3. Schmidt, M. A., Marcano-Bonilla, L., & Roberts, L. R. (2019). Gallbladder cancer: epidemiology and genetic risk associations. *Chinese clinical oncology*, *8*(4), 31-31.

4. (Dutta, Bush, Kalsi, Popli, & Kapoor, 2019)Dutta, U., Bush, N., Kalsi, D., Popli, P., & Kapoor, V. K. (2019). Epidemiology of gallbladder cancer in India. *Chinese Clinical Oncology*. https://doi.org/10.21037/cco.2019.08.03

5. Albores-Saavedra, J., Alcantra-Vazquez, A., Cruz-Ortiz, H., & Herrera-Goepfert, R. (1980). The precursor lesions of invasive gallbladder carcinoma. Hyperplasia, atypical hyperplasia and carcinoma in situ. *Cancer*. https://doi.org/10.1002/1097-0142(19800301)45:5<919::AID-CNCR2820450514>3.0.CO;2-4

6. Roa, I., Araya, J. C., Villaseca, M., De Aretxabala, X., Riedemann, P., Endoh, K., & Roa, J. (1996). Preneoplastic lesions and gallbladder cancer: An estimate of the period required for progression. *Gastroenterology*. https://doi.org/10.1053/gast.1996.v111.pm8698204

7. Nuzzo, G., Clemente, G., Cadeddu, F., Ardito, F., Ricci, R., & Vecchio, F. M. (2005). Papillary carcinoma of the gallbladder and anomalous pancreatico-biliary junction. Report of three cases and review of the literature. Hepato-gastroenterology, 52(64), 1034-1038.

8. Hundal, R., & Shaffer, E. A. (2014). Gallbladder cancer: epidemiology and outcome. Clinical epidemiology, 99-109. https://doi.org/10.2147/CLEP.S37357.

9. Hsing, A. W., Bai, Y., Andreotti, G., Rashid, A., Deng, J., Chen, J., … Gao, Y. T. (2007). Family history of gallstones and the risk of biliary tract cancer and gallstones: A population-based study in Shanghai, China. *International Journal of Cancer*. https://doi.org/10.1002/ijc.22756

10. Konstantinidis, I.T.; Deshpande, V.; Genevay, M.; Berger, D.; Castillo, C.F.-D.; Tanabe, K.K.; Zheng, H.; Lauwers, G.Y.; Ferrone, C.R.Trends in presentation and survival for gallbladder cancer during a period of more than 4 decades: A single-institution experience. Arch. Surg., 2009, 144, 441–447. https://doi.org/10.1001/archsurg.2009.46.

11. Rawal, N.; Awasthi, S; Dash, N.R.; Kumar, S.; Das, P.; Ranjan, A.; Chopra, A.; Khan, M.A.; Saluja, S.; Hussain, S.; Tanwar, P. Prognostic Relevance of PDL1 and CA19-9 Expression in Gallbladder Cancer vs. Inflammatory Lesions. Curr. Oncol. 2023, 30, 1571–1584. https://doi.org/10.3390/curroncol30020121

12. Tan, Y., Ma, S. Y., Wang, F. Q., Meng, H. P., Cuizhu, M., Liu, A., & Wu, H. R. (2011). Proteomic-based analysis for identification of potential serum biomarkers in gallbladder cancer. *Oncology Reports*. https://doi.org/10.3892/or.2011.1353

13. Orsburn, B. C. (2021). Proteome discoverer-a community enhanced data processing suite for protein informatics. *Proteomes*. https://doi.org/10.3390/proteomes9010015

14. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., … Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky1131

15. Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., & Lin, C. Y. (2014). cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*. https://doi.org/10.1186/1752-0509-8-S4-S11

16. Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis

of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. https://doi.org/10.1038/nprot.2008.211

17. Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal*. https://doi.org/10.1002/bimj.200410135

18. Huang, H. L., Yao, H. S., Wang, Y., Wang, W. J., Hu, Z. Q., & Jin, K. Z. (2014). Proteomic identification of tumor biomarkers associated with primary gallbladder cancer. *World Journal of Gastroenterology*. https://doi.org/10.3748/wjg.v20.i18.5511

19. Sahasrabuddhe, N. A., Barbhuiya, M. A., Bhunia, S., Subbannayya, T., Gowda, H., Advani, J., … Tiwari, P. K. (2014). Identification of prosaposin and transgelin as potential biomarkers for gallbladder cancer using quantitative proteomics. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2014.03.017

20. Waghmare, R. S., & Kamat, R. N. (2014). Incidental gall bladder carcinoma in patients undergoing cholecystectomy: A report of 7 cases. *J Assoc Physicians India*, *62*(9), 793-6.

21. Aishima, S., Fujita, N., Mano, Y., Kubo, Y., Tanaka, Y., Taketomi, A., … Oda, Y. (2011). Different roles of s100p overexpression in intrahepatic cholangiocarcinoma: Carcinogenesis of perihilar type and aggressive behavior of peripheral type. *American Journal of Surgical Pathology*. https://doi.org/10.1097/PAS.0b013e31820ffdf1

22. Mathai, A. M., Alexander, J., Huang, H. Y., Li, C. F., Jeng, Y. M., Fung, K. M., … Yeh, M. M. (2021). S100P as a marker for poor survival and advanced stage in gallbladder carcinoma. *Annals of Diagnostic Pathology*. https://doi.org/10.1016/j.anndiagpath.2021.151736

23. Wu, Z., Yu, X., Zhang, S., He, Y., & Guo, W. (2022). The role of PI3K/AKT signaling pathway in gallbladder carcinoma. American Journal of Translational Research, 14(7), 4426.

24. Parkkila, S., Pan, P. W., Ward, A., Gibadulinova, A., Oveckova, I., Pastorekova, S., … Isola, J. (2008). The calcium-binding protein S100P in normal and malignant human tissues. *BMC Clinical Pathology*. https://doi.org/10.1186/1472-6890-8-2

25. Zhang, L., Zhang, H., Yue, D., Wei, W., Chen, Y., Zhao, X., … Wang, C. (2019). The prognostic value of the preoperative albumin to alkaline phosphatase ratio in patients with non-small cell lung cancer after surgery. *Thoracic Cancer*, *10*(7), 1581–1589. https://doi.org/10.1111/1759-7714.13107

26. Ohkuma, R., Yada, E., Ishikawa, S., Komura, D., Kubota, Y., Hamada, K., … Wada, S. (2020). High expression levels of polymeric immunoglobulin receptor are correlated with chemoresistance and poor prognosis in pancreatic cancer. *Oncology Reports*. https://doi.org/10.3892/or.2020.7610

27. Chen, Y. B., Jiang, C. T., Zhang, G. Q., Wang, J. S., & Pang, D. (2009). Increased expression of hyaluronic acid binding protein 1 is correlated with poor prognosis in patients with breast cancer. *Journal of Surgical Oncology*. https://doi.org/10.1002/jso.21329

28. Niu, M., Sun, S., Zhang, G., Zhao, Y., Pang, D., & Chen, Y. (2015). Elevated expression of HABP1 is correlated with metastasis and poor survival in breast cancer patients. *American Journal of Cancer Research*, *5*(3), 1190.

29. Wang, J., Song, Y., Liu, T., Shi, Q., Zhong, Z., Wei, W., … Pang, D. (2015). Elevated expression of HABP1 is a novel prognostic indicator in triple-negative breast cancers. *Tumor Biology*. https://doi.org/10.1007/s13277-015-3131-x30.

30. Gao, H., Yao, Q., Lan, X., Li, S., Wu, J., Zeng, G., & Xue, Y. (2016). Elevated HABP1 protein expression correlates with progression and poor survival in patients with gastric cancer. *OncoTargets and Therapy*. https://doi.org/10.2147/OTT.S114756.

31. Shen, X., Han, B., Shen, Y., Yang, J., Ren, T., Sha, G., & Xiang, Y. (2014). Expression of C1QBP gene and its correlation with drug resistance in human resistance choriocarcinoma cell line. *Zhonghua fu Chan ke za zhi*, *49*(8), 616-620.

32. Assinder, S. J., Stanton, J. A. L., & Prasad, P. D. (2009). Transgelin: An actin-binding protein and tumour suppressor. *International Journal of Biochemistry and Cell Biology*. https://doi.org/10.1016/j.biocel.2008.02.011.

33. Shields, J. M., Rogers-Graham, K., & Der, C. J. (2002). Loss of transgelin in breast and colon tumors and in RIE-1 cells by Ras deregulation of gene expression through Raf-independent pathways. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M110086200

34. Thompson, O., Moghraby, J. S., Ayscough, K. R., & Winder, S. J. (2012). Depletion of the actin bundling protein SM22/transgelin increases actin dynamics and enhances the tumourigenic phenotypes of cells. *BMC Cell Biology*. https://doi.org/10.1186/1471-2121-13-1

35. Tsui, K. H., Lin, Y. H., Chang, K. S., Hou, C. P., Chen, P. J., Feng, T. H., & Juang, H. H. (2019). Transgelin, a p53 and PTEN-upregulated gene, inhibits the cell proliferation and invasion of human bladder carcinoma cells in vitro and in vivo. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms20194946.

36. Yamane, T., Asanoma, K., Kobayashi, H., Liu, G., Yagi, H., Ohgami, T., ... & Kato, K. (2015). Identification of the critical site of calponin 1 for suppression of ovarian cancer properties. *Anticancer research*, *35*(11), 5993-5999.

37. Lin, Z. Y., & Chuang, W. L. (2012). Genes responsible for the characteristics of primary cultured invasive phenotype hepatocellular carcinoma cells. *Biomedicine and Pharmacotherapy*. https://doi.org/10.1016/j.biopha.2012.04.001

38. Mamoor, S. (n.d.). *Differential expression of calponin 1 in cancers of the breast.* https://doi.org/10.31219/OSF.IO/U6VTW

39. Hammad, A., Elshaer, M., & Tang, X. (2021). Identification of potential biomarkers with colorectal cancer based on bioinformatics analysis and machine learning. *Mathematical Biosciences and Engineering*. https://doi.org/10.3934/mbe.2021443

40. Chen, M., Li, S., Topatana, W., Lv, X., Cao, J., Hu, J., … Cai, X. (2021). Development and Validation of a Nomogram for Predicting Survival in Gallbladder Cancer Patients With Recurrence After Surgery. *Frontiers in Oncology*. https://doi.org/10.3389/fonc.2020.537789

41. Higuchi, R., Yazawa, T., Uemura, S., Matsunaga, Y., Ota, T., Araida, T., … Yamamoto, M. (2020). Examination of prognostic factors affecting long-term survival of patients with stage 3/4 gallbladder cancer without distant metastasis. *Cancers*. https://doi.org/10.3390/cancers12082073.

42. Uzun, M. A., Tilki, M., Kayaoğlu, S. A., Okuyan, G. Ç., Kılıçoğlu, Z. G., & Gönültaş, A. (2022). Long-term results and prognostic factors after surgical treatment for gallbladder

cancer. *Turkish Journal of Surgery*, *38*(4), 334–344. https://doi.org/10.47717/TURKJSURG.2022.5861

43. Naveed, S., Qari, H., Thau, C. M., Burasakarn, P., Mir, A. W., & Panday, B. B. (2021). Lymph Node Ratio is an Important Prognostic Factor in Curatively Resected Gallbladder Carcinoma, Especially in Node-positive Patients: An Experience from Endemic Region in a Developing Country. *Euroasian Journal of Hepato-Gastroenterology*. https://doi.org/10.5005/jp-journals-10018-1336

44. Shirai, Y., Sakata, J., Wakai, T., Ohashi, T., Ajioka, Y., & Hatakeyama, K. (2012). Assessment of lymph node status in gallbladder cancer: Location, number, or ratio of positive nodes. *World Journal of Surgical Oncology*. https://doi.org/10.1186/1477-7819-10-87

45. Liu, Z., Zhu, G., Jiang, X., Zhao, Y., Zeng, H., Jing, J., & Ma, X. (2020). Survival Prediction in Gallbladder Cancer Using CT Based Machine Learning. *Frontiers in Oncology*. https://doi.org/10.3389/fonc.2020.604288

46. Sachan, A., Saluja, S. S., Nekarakanti, P. K., Nimisha, Mahajan, B., Nag, H. H., & Mishra, P. K. (2020). Raised CA19-9 and CEA have prognostic relevance in gallbladder carcinoma. *BMC Cancer*. https://doi.org/10.1186/s12885-020-07334-x

47. Xu, W. Y., Zhang, H. H., Xiong, J. P., Yang, X. B., Bai, Y., Lin, J. Z., … Sang, X. T. (2018). Prognostic significance of the fibrinogen-to-albumin ratio in gallbladder cancer patients. *World Journal of Gastroenterology*. https://doi.org/10.3748/wjg.v24.i29.3281

48. Bandırmalı, O. (2020). Prognostic Factors in Gallbladder Cancer. *Erciyes Medical Journal*. https://doi.org/10.14744/ETD.2020.58897

49. Farhat, M. H., Shamseddine, A. I., Tawil, A. N., Berjawi, G., Sidani, C., Shamseddeen, W., & Barada, K. A. (2008). Prognostic factors in patients with advanced cholangiocarcinoma: Role of surgery, chemotherapy and body mass index. *World Journal of Gastroenterology*. https://doi.org/10.3748/wjg.14.3224

50. Tran, T. B., Norton, J. A., Ethun, C. G., Pawlik, T. M., Buettner, S., Schmidt, C., … Poultsides, G. A. (2017). Gallbladder Cancer Presenting with Jaundice: Uniformly Fatal or Still Potentially Curable? *Journal of Gastrointestinal Surgery*. https://doi.org/10.1007/s11605-017-3440-z

TABLES AND FIGURES WITH LEGENDS

TABLE 1. Patients' clinico-pathological characteristics correlated with S100P, PIGR, C1QBP, TAGLN, CNN1 concentration at their respective optimum cut-off values (5.353 ng/µg, 1.068 ng/µg, 10.32 ng/µg, 12.66 ng/µg and 19.79 ng/µg, respectively) (*n* = 80)

Characteristics		**Frequency (*n*=80)**	**S1OOP**		***P***	**PIGR**		***P***	**C1QBP**		***P***	**TAGLN**		***P***	**CNN1**		***P***
			Low expression (<5.35 ng/µg)	High expression (≥5.35 ng/µg)		Low expression (<1068 pg/µg)	High expression (≥1068 pg/µg)		Low expression (<10.32 ng/µg)	High expression (≥10.32 ng/µg)		Low expression (<12.66 ng/µg)	High expression (≥12.66 ng/µg)		Low expression (<19.79 ng/µg)	High expression (≥19.79 ng/µg)	
Median age (years): 52																	
Phenotype	Cancer	40 (50%)	19	21	**0.0001***	8	32	**0.0006***	19	21	**<0.0001***	28	12	**<0.0001***	28	12	**<0.0001***
	Inflammatory lesions	40 (50%)	35	5		23	17		37	3		3	37		4	36	
Gallstone	Present	12 (15%)	9	3	0.54	7	5	0.13	12	0	**0.014***	3	9	0.31	5	7	0.9
	Absent	68 (85%)	45	23		24	44		44	24		28	41		27	41	
Lymph node metastasis	N0	29 (72.5%)	16	13	0.24	7	22	0.55	13	16	**0.001***	19	10	0.55	20	9	0.80
	N1	10 (25%)	3	7		1	9		42	8		8	2		7	3	
	N2	1 (2.5%)	0	1		0	1		1	0		1	0		1	0	
Differentiation	Well/ moderate	29(72.5%)	11	18	**0.049***	6	23	0.86	11	18	**0.049***	21	11	0.23	19	10	0.32
	Poor	11(27.5%)	8	3		2	9		8	3		7	1		9	2	
Tumor size	T1	9 (11.25%)	3	6	0.16	2	7	0.91	2	7	0.27	4	5	0.24	4	5	0.15

	T2	14 (17.5%)	10	4		3	11		9	5		11	3		12	2	
	T3	15 (18.75%)	5	10		3	12		7	8		11	4		10	5	
	T4	2 (2.5%)	1	1		0	2		1	1		2	0		2	0	
Metastasis	Yes	19 (47.50%)	8	11	0.52	3	16	0.53	11	8	0.21	13	4	0.44	14	5	0.63
	No	21 (52.5%)	11	10		5	16		8	13		15	8		14	7	
Preoperative laboratory values																	
Bilirubin (Total), mg/dL	<0.20	4(5%)	2	2	0.74	2	2	0.08	2	2	0.63	1	3	0.76	0	4	0.10
	0.20-1.20	64(80%)	44	20		21	43		46	18		26	38		29	35	
	>1.20	12 (15%)	8	4		8	4		8	4		4	8		3	9	
Bilirubin (conjugated), mg/dL	<0.10	5(6.25%)	1	4	**0.02***	1	4	0.32	1	4	**0.04***	2	3	0.66	2	3	0.96
	0.10-0.30	51(63.75%)	39	12		18	33		37	14		20	21		21	30	
	>0.30	24 (30%)	14	10		12	12		18	6		9	15		9	15	
Bilirubin (unconjugated), mg/dL	<0.20	10 (12.5%)	5	5	0.15	3	7	0.53	7	3	**0.04***	6	4	**0.003***	8	2	**0.01***
	0.20-0.90	65 (81.25%)	47	18		25	40		48	17		20	45		21	44	
	>0.90	5(6.25%)	2	3		3	2		1	4		5	0		3	2	
Blood test markers																	
CA125, U/mL	<35	2 (2.5%)	1	1	0.25	1	1	0.25	1	1	0.25	1	1	>0.9	2	0	0.25
	≥35	2 (2.5%)	0	2		2	0		0	2		1	1		1	1	
	NT	76 (95%)	53	23		28	48		55	21		29	47		29	47	
CA19.9, U/mL	<37	41(51.25%)	30	11	0.12	15	26	0.68	31	10	0.075	16	25	0.82	14	27	0.17

	≥37	19 (23.75 %)	10	9		8	11		10	9		8	11		10	9	
	NT	20 (25%)	14	6		8	12		15	5		7	13		8	12	
CEA, ng/mL	<5	38 (47.5%)	28	10	**0.04***	17	21	0.15	26	12	0.09	13	25	0.14	11	27	**0.003***
	≥5	10 (12.5%)	4	6		2	8		4	6		6	4		8	2	
	NT	32 (40%)	22	10		12	20		26	6		12	20		13	19	

*$P < 0.05$; NT: Not tested.

TABLE 2. Univariate and multivariate analysis of prognostic factors in GBC cases (*n* = 40)

Characteristics		*n*	Univariate cox regression		Multivariate cox regression	
			Hazard ratio [95% CI]	*P* value	Hazard ratio [95% CI]	*P* value
S100P	<5.35ng/µg (ref)	19	1.58 [0.56-4.46]	0.39	0.12 [0.01-2.66]	0.18
	≥5.35 ng/µg	21				
PIGR	<1068 pg/µg (ref)	8	0.94 [0.26-3.33]	0.92	0.22 [0.01-4.87]	0.34
	≥1068 pg/µg	32				
C1QBP	<10.32 ng/µg (ref)	19	1.14 [0.41-3.15]	0.8	4.87 [0.12-7.39]	0.94
	≥10.32 ng/µg	21				
TAGLN	<12.66 ng/µg (ref)	28	0.34 [0.08-1.51]	0.16	0.73 [0.03-16.89]	0.85
	≥12.66 ng/µg	12				
CNN1	<19.79 ng/µg (ref)	28	1.38 [0.43-4.44]	0.59	3.07 [0.44-21.15]	0.26
	≥19.79 ng/µg	12				
Age	<60 (ref)	32	1.71 [0.58-5.0]	0.33	1.24 [0.08-19.08]	0.88
	≥60	8				
Gender	Male (ref)	13	0.40 [0.14-1.16]	0.09	3.52 [0.21-59.41]	0.38
	Female	27				
Tumor size	T1/T2 (ref)	23	4.89 [1.62-14.77]	**0.005***	34.22 [0.60-1959.46]	0.09
	T3/T4	17				
Lymphnode metastasis	N0 (ref)	29	0.05 [0.01-0.24]	**0.0001***	0.01 [0.001-0.22]	**0.004***
	N1/N2	11				
Differentiation	Poor (ref)	8	1.31 [0.42-4.2]	0.64		
	Well/moderate	11				
Distant Metastasis	No (ref)	21	13.04 [2.77-61.37]	**0.001***	2.76 [0.07-107.49]	0.59
	Yes	19				
Gall stones	No (ref)	37	NE			
	Yes	3				
Dietary habit	Veg (ref)	24	0.78 [0.25-2.47]	0.67		

	Non veg	16				
Smoking	No (ref)	37	0.85 [0.11-6.45]	0.87		
	Yes	3				
Alcohol	No (ref)	37	1.19 [0.27-5.3]	0.82		
	Yes	3				
Urea (mg/dl)	<40.0 (ref)	36	1.33 [0.30-5.91]	0.79		
	≥40.0	4				
Creatine (mg/dl)	<1.0 (ref)	37	1.82 [0.41-8.08]	0.43		
	≥1.0	3				
Calcium (mg/dl)	<10.5 (ref)	37	1.02 [0.13-7.87]	0.98		
	≥10.5	3				
Bilirubin (Total); mg/dL	<1.20 (ref)	35	4.89 [1.58-15.11]	**0.006***	9.88 [0.21-462.90]	0.24
	≥1.20	5				
Bilirubin (conjugated; mg/dL	<0.30 (ref)	29	2.32 [0.73-7.35]	0.15		
	≥0.30	11				
Bilirubin (unconjugated); mg/dL	<0.90 (ref)	35	3.39 [1.04-11.09]	**0.04***	0.49 [0.06-4.24]	0.52
	≥0.90	5				
SGOT (AST); U/L	<40.00 (ref)	31	1.94 [0.60-6.28]	0.27		
	≥40.00	9				
SGPT (ALT); U/L	<45.00 (ref)	28	1.07 [0.34-3.40]	0.91		
	≥45.00	12				
Total Protein, g/dL	<8.00 (ref)	36	0.53 [0.07-4.03]	0.54		
	≥8.00	4				
Albumin;g/dL	<5.00 (ref)	39	NE	-		
	≥5.00	1				
Globulin (g/dL)	<3.5 (ref)	36	0.57 [0.07-4.36]	0.59		
	≥3.5	4				

*P <0.05; NE: Not estimable; CI: Confidence interval; ref: Reference value.

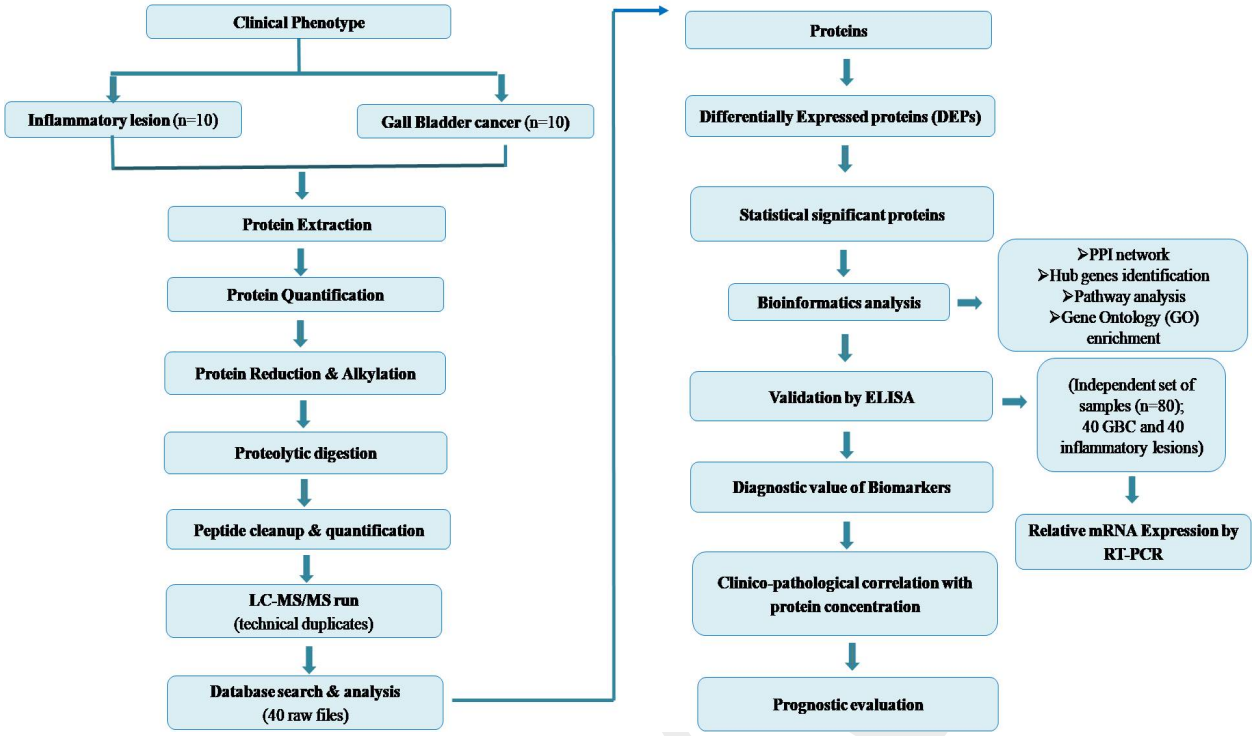


FIGURE 1. Study workflow.

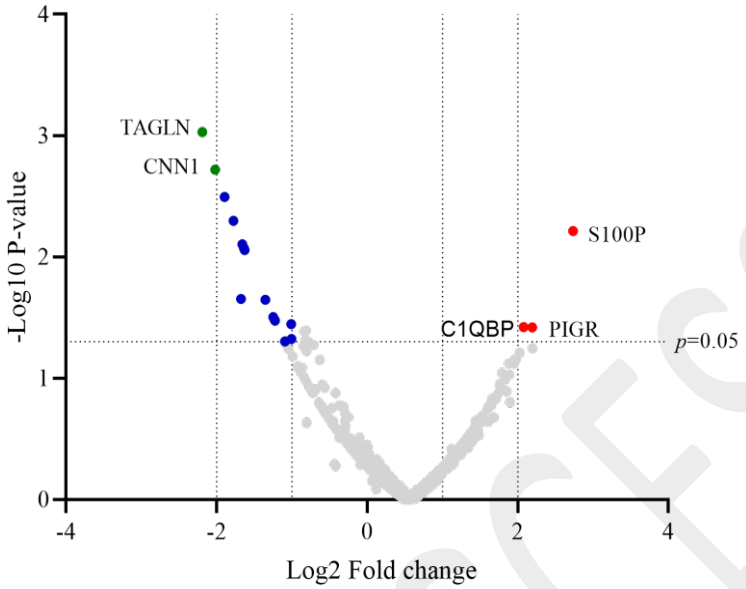


FIGURE 2. **Volcano plot of filtered proteins at Log_2 Fold change 2.** Green colored dots represent significant down-regulated proteins at log2 fold change 2 and P = 0.05. Red colored dots represent significant up-regulated proteins at log2 fold change 2 and P = 0.05. Blue colored dots represent significant down regulated proteins at log2 fold change <2 and P = 0.05. Grey color dots represent non-significant proteins.

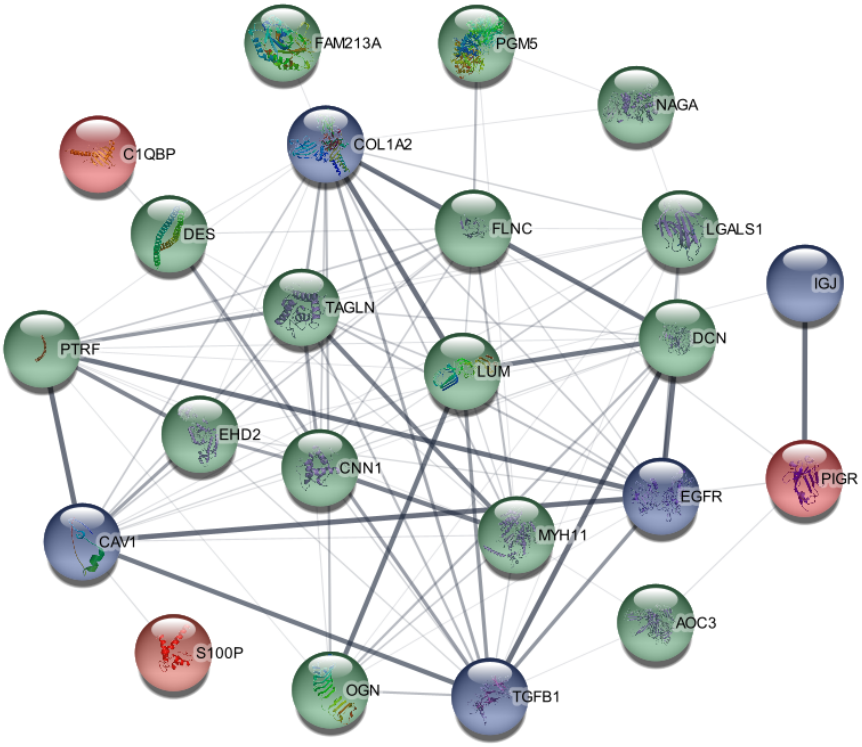
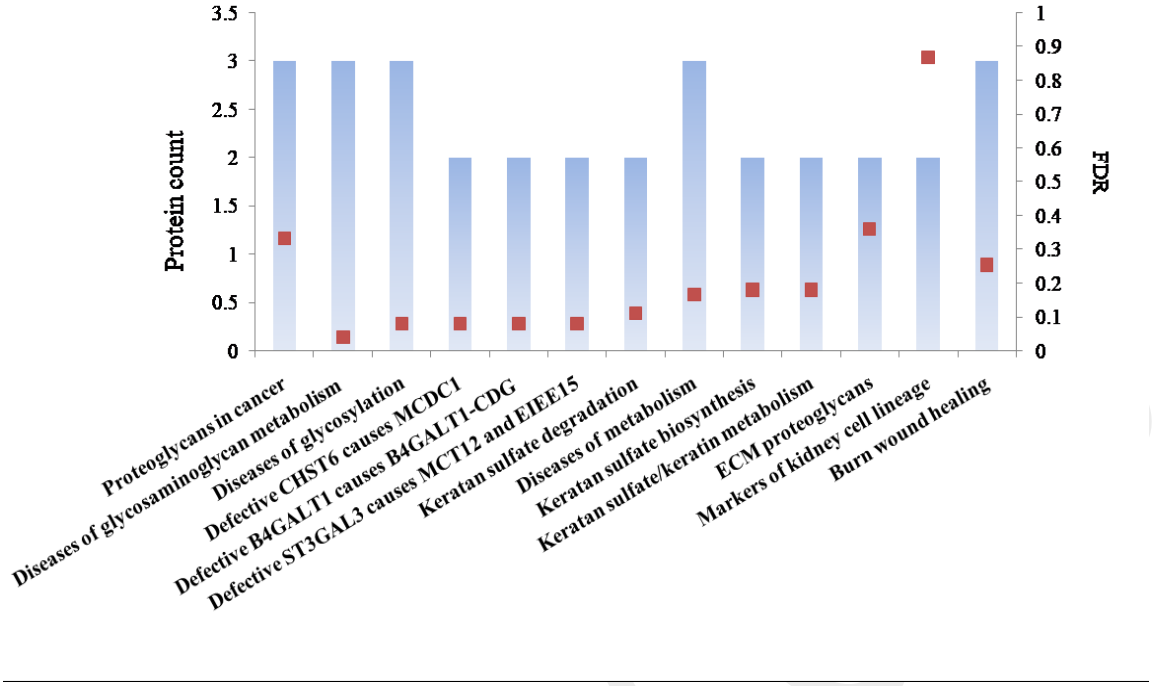
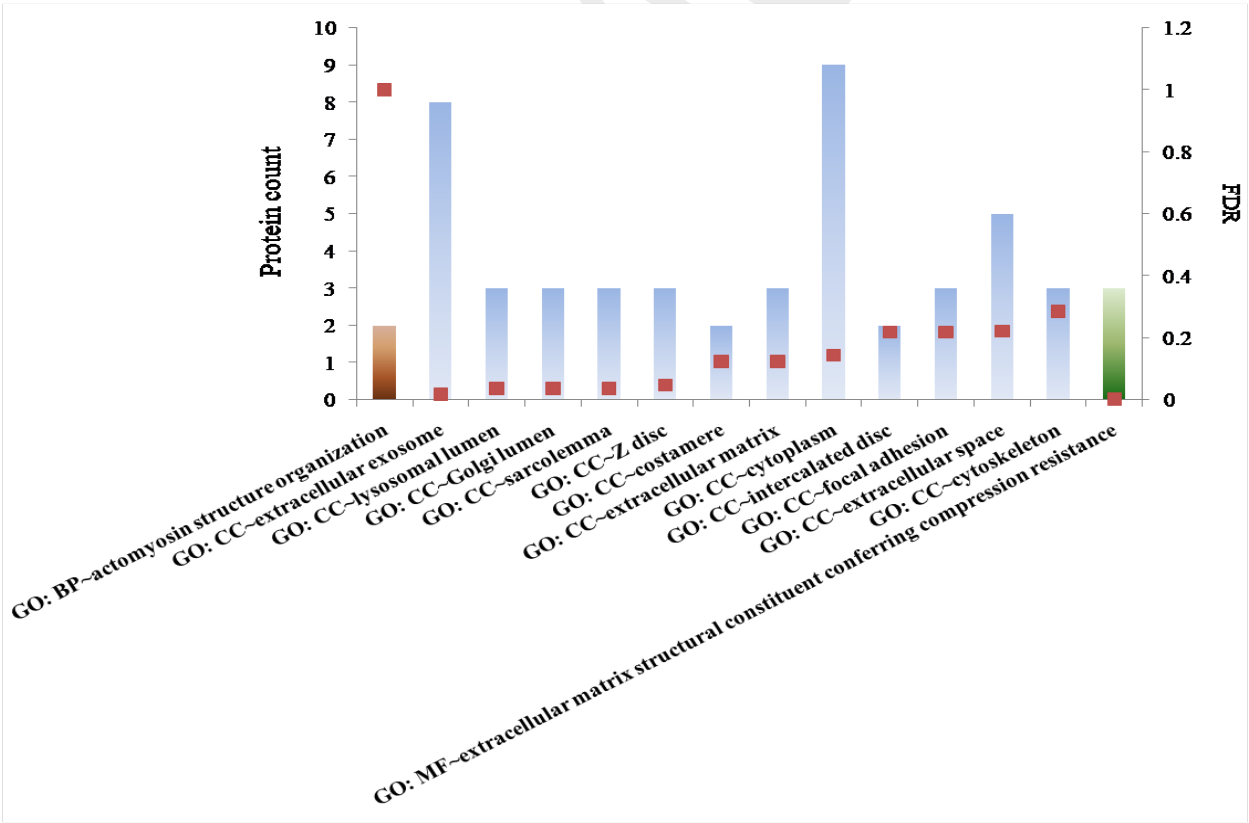


FIGURE 3. Protein-protein interaction (PPI) network of all the 18 differentially expressed proteins (DEPs). Red color nodes represented up regulated proteins ((S100P, PIGR, C1QBP), green color nodes represented down regulated proteins (TAGLN, CNN1, PTRF, DES, EHD2, PGM5, AOC3, NAGA, LUM, FLNC, DCN, MYH11, OGN, FAM213A, LGALS1) and blue color nodes were other common interacting proteins (CAV1, TGFB1, EGFR, IGJ). Thick edge in between the nodes indicates the high confidence strength of data support.

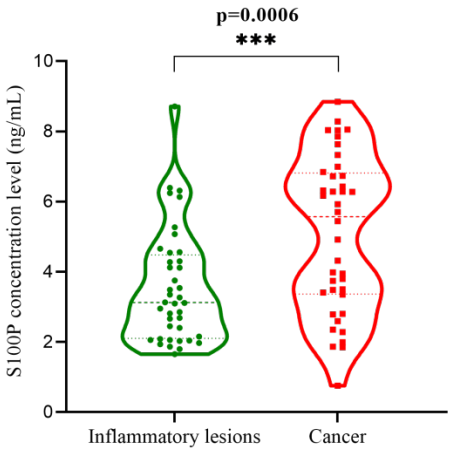

Protein count
FDR
3.5
3
2.5
2
1.5
1
0.5
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Proteoglycans in cancer
Diseases of glycosaminoglycan metabolism
Diseases of glycosylation
Defective CHST6 causes MCDC1
Defective B4GALT1 causes B4GALT1-CDG
Defective ST3GAL3 causes MCT12 and EIEE15
Keratan sulfate degradation
Diseases of metabolism
Keratan sulfate biosynthesis
Keratan sulfate/keratin metabolism
ECM proteoglycans
Markers of kidney cell lineage
Burn wound healing

(A)

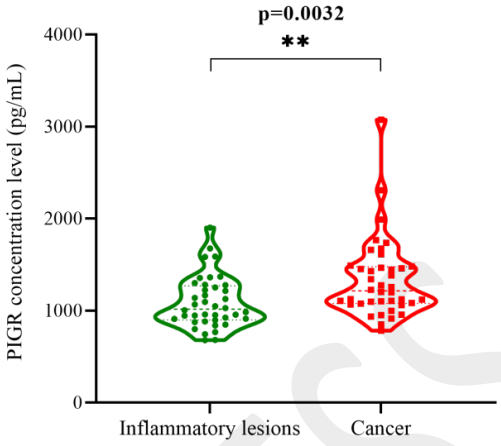

Protein count
FDR
10
9
8
7
6
5
4
3
2
1
0
1.2
1
0.8
0.6
0.4
0.2
0
GO: BP~actomyosin structure organization
GO: CC~extracellular exosome
GO: CC~lysosomal lumen
GO: CC~Golgi lumen
GO: CC~sarcolemma
GO: CC~Z disc
GO: CC~costamere
GO: CC~extracellular matrix
GO: CC~cytoplasm
GO: CC~intercalated disc
GO: CC~focal adhesion
GO: CC~extracellular space
GO: CC~cytoskeleton
GO: MF~extracellular matrix structural constituent conferring compression resistance

(B)

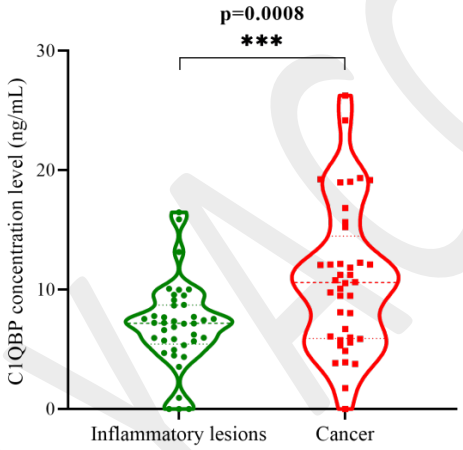
FIGURE 4. (A) Pathway analysis of DEPs through KEGG, Reactome and Wiki pathway databases; (B) Gene Ontologies of DEPs in which brown color bar represents the GO_BP, blue color bar represents GO_CC and green color bar represents GO_MF. The square boxes designate the FDR corrected *P* values.



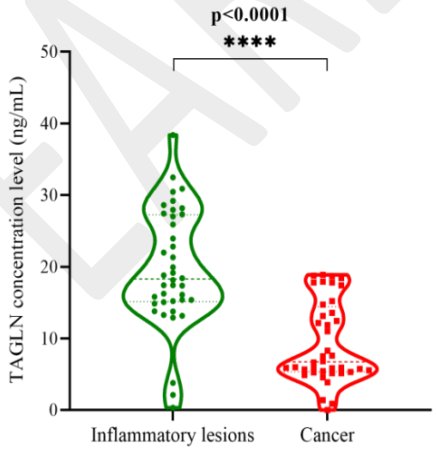
(a)



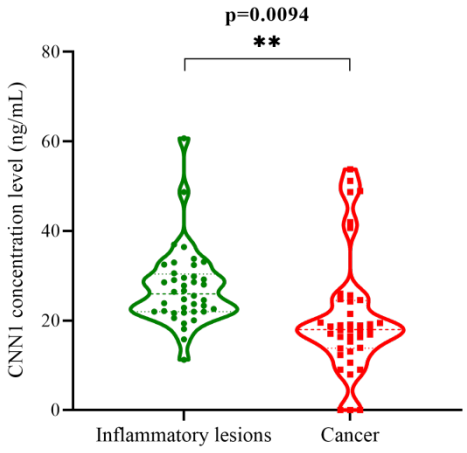
(B)



(C)



(D)



(E)

FIGURE 5. Violin box plots showing the concentration levels in inflammatory and cancer cases. (A) S100P concentration (ng/mL) ($p=0.0006$); (B) PIGR concentration (pg/mL) ($p=0.0032$); (C) C1QBP concentration (ng/mL) ($p=0.0008$); (D) TAGLN concentration (ng/mL) ($p<0.0001$); (E) CNN1 concentration (ng/mL) ($p=0.0094$). Each green color dot represents concentration level in individual inflammatory cases of gall bladder and red color dot represents concentration level in individual GBC cases.

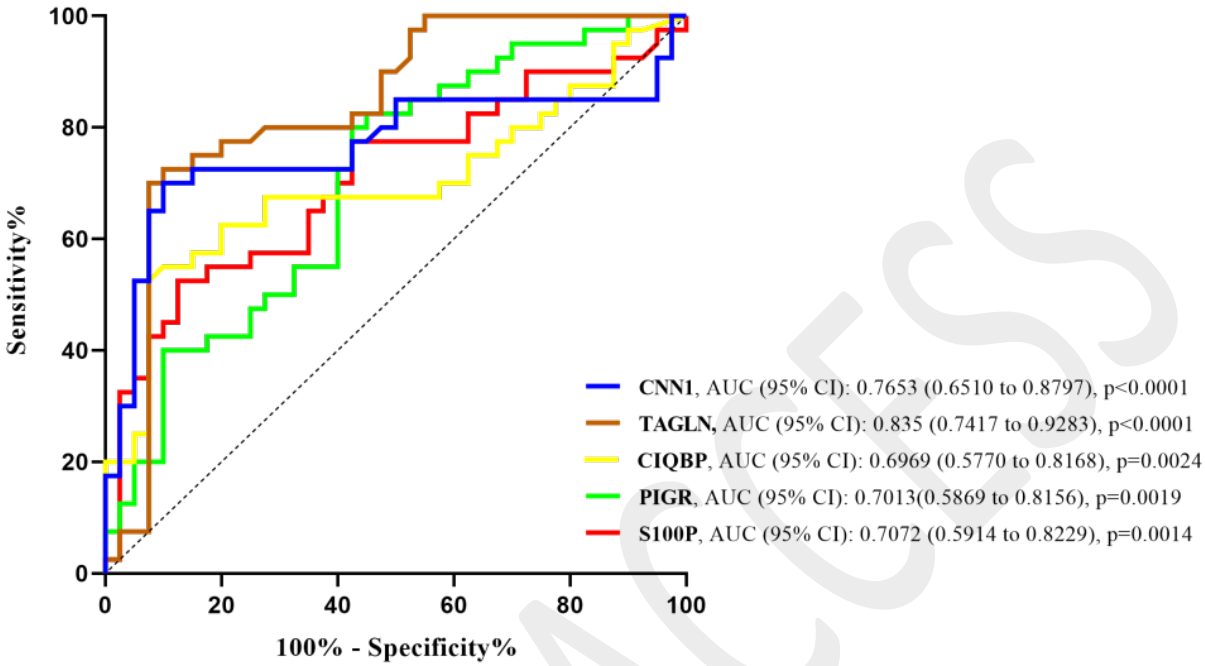


FIGURE 6. Receiver operating characteristic (ROC) curve of S100P (red color), PIGR (green color), C1QBP (yellow color), TAGLN (brown color) and CNN1 (blue color).

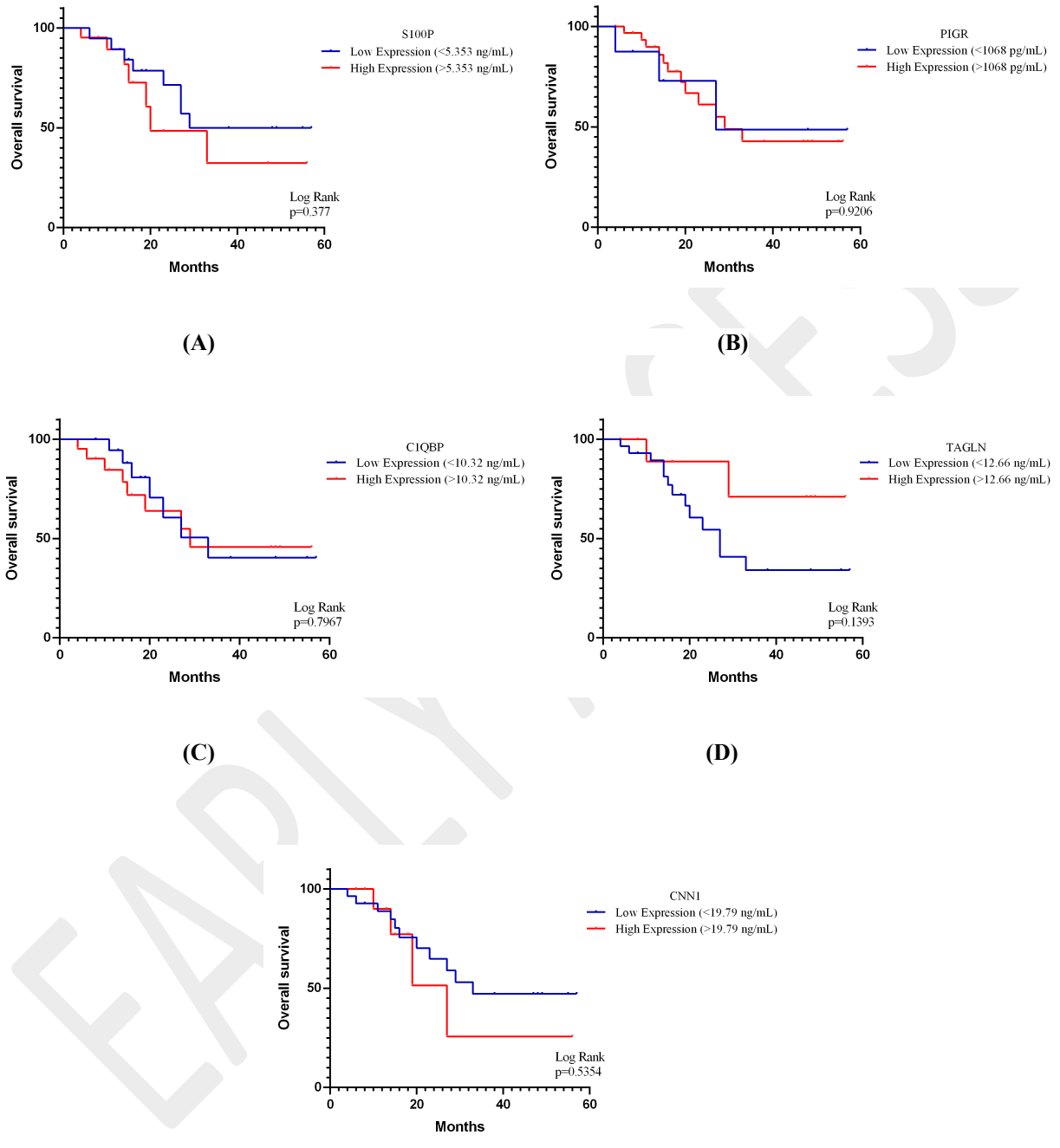
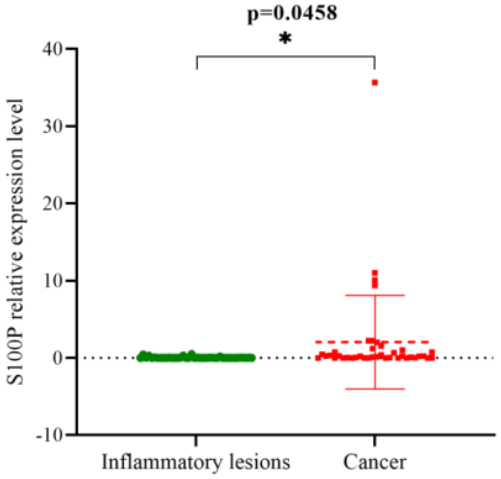
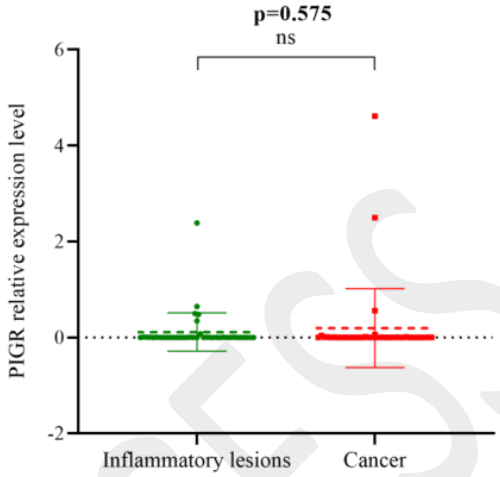


FIGURE 7. Kaplan–Meier survival curves. (A) S100P concentration (cut-off value of 5.353 ng/mL, *P* = 0.377); (B) PIGR concentration (cut-off value of 1068 pg/mL, *P* = 0.9206); (C) C1QBP concentration (cut-

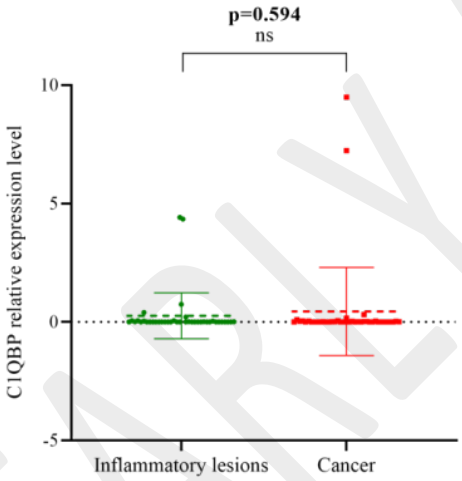
off value of 10.32 ng/mL, $P = 0.7967$); (D) TAGLN concentration (cut-off value of 12.66 ng/mL, $P = 0.1393$); (E) CNN1 concentration (cut-off value of 19.79, $P = 0.5354$).



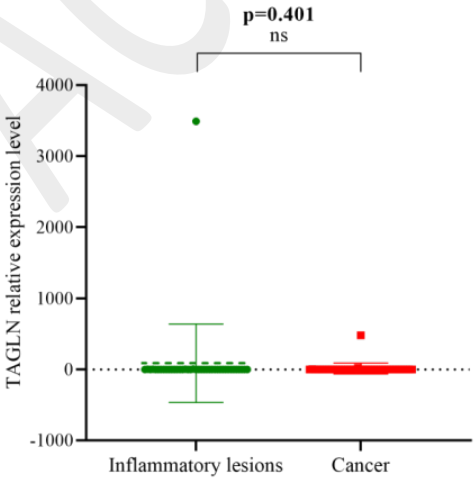
(A)



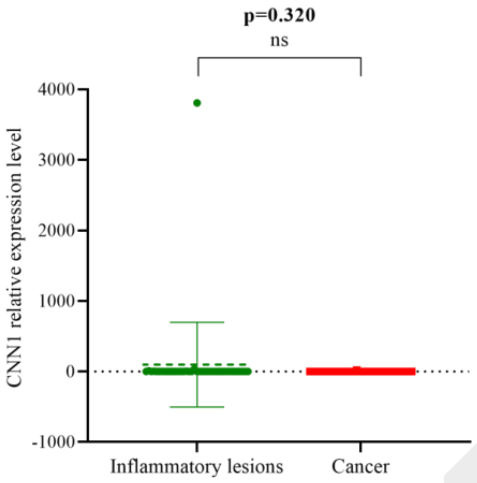
(B)



(C)



(D)



(E)

FIGURE 8. Box violin plot shows relative mRNA expression level in inflammatory and cancer cases. (A) S100P (p=0.0458); (B) PIGR (p=0.575); (C) C1QBP (p=0.594); (D) TAGLN (p=0.401); (E) CNN1 (p=0.320). Each green color dot represents relative mRNA expression level in individual inflammatory cases of gall bladder and red color dot represents relative mRNA expression level in individual GBC cases.

SUPPLEMENTAL DATA

TABLE S1. Differentially expressed proteins by LC-MS/MS at Log_2 fold change of 1 and *P* < 0.05

Accession	Gene symbol	Chromosome	Coverage [%]	MW [kDa]	calc. pI	PSMs	Unique peptides	Log_2FC at $P < 0.05$	Abundance ratio: (Cancer) / (Inflammatory lesion (control))
P25815	S100P	4	72	10.4	4.88	134	4	2.7	6.654
P01833	PIGR	1	61	83.2	5.74	347	32	2.1	4.569
A8K651	C1QBP	17	50	31.4	4.84	143	7	2	4.22
Q01995	TAGLN	11	67	22.6	8.84	785	14	2.1	0.219
P51911	CNN1	19	62	33.2	9.07	312	11	2	0.247
B4DPZ5	PTRF	17	26	40.5	5.25	148	8	1.8	0.269
Q53SB5	DES	2	68	53.5	5.27	990	24	1.7	0.292
Q9NZN4	EHD2	19	55	61.1	6.46	426	21	1.6	0.317
Q15124	PGM5	9	65	62.2	7.21	308	24	1.6	0.324
Q16853	AOC3	17	27	84.6	6.52	253	13	1.6	0.322
P17050	NAGA	22	21	46.5	5.19	67	6	1.6	0.313
P51884	LUM	12	41	38.4	6.61	863	8	1.3	0.392
Q14315	FLNC	7	59	290.8	5.97	1133	92	1.2	0.421
P07585	DCN	12	47	39.7	8.54	202	12	1.2	0.428
P35749	MYH11	16	56	227.2	5.5	2512	75	1.2	0.426
Q7Z532	OGN	9	44	33.9	5.48	247	11	1	0.497
Q9BRX8	FAM213A	10	21	25.7	8.84	59	4	1	0.498
A0A384MR27	LGALS1	22	90	14.7	5.5	677	9	1	0.469

TABLE S2. Identification of hub genes through cytohubba

Closeness method			Degree			MNC			MCC			EPC			Bottleneck		
Rank	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**
1	COL1A2	19	1	COL1A2	16	1	COL1A2	15	1	COL1A2	168003	1	COL1A2	9.546	1	COL1A2	5
2	MYH11	18	2	MYH11	14	2	MYH11	14	2	TAGLN	168000	2	TAGLN	9.515	2	MYH11	3
2	FLNC	18	2	FLNC	14	2	FLNC	14	3	DCN	167760	3	DCN	9.337	2	FLNC	3
2	TAGLN	18	2	TAGLN	14	2	TAGLN	14	4	LUM	156241	4	EGFR	9.237	4	LUM	2
5	TGFB1	17.5	5	EGFR	13	5	EGFR	13	5	CAV1	147601	5	CAV1	9.225	4	DES	2
6	EGFR	17.33333	5	DCN	13	5	DCN	13	6	PTRF	141120	6	MYH11	9.21	4	CAV1	2
6	DCN	17.33333	5	LUM	13	5	TGFB1	13	7	FLNC	131164	7	LUM	9.164	7	C1QBP	1
6	LUM	17.33333	5	CAV1	13	8	LUM	12	8	EGFR	116642	8	TGFB1	9.163	7	EGFR	1
6	CAV1	17.33333	5	TGFB1	13	8	CAV1	12	9	MYH11	101044	9	FLNC	9.056	7	DCN	1
10	PTRF	16.83333	10	PTRF	12	8	PTRF	12	10	LGALS1	55442	10	PTRF	8.899	7	EHD2	1

EcCentricity			DMNC			Betweeness			Radiality			Stress			Clustering Coefficient		
Rank	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**	**Rank**	**Gene name**	**Score**
1	COL1A2	0.5	1	LUM	0.848828	1	COL1A2	74.30815	1	COL1A2	3.954545	1	COL1A2	158	1	OGN	0.944444
1	MYH11	0.5	1	CAV1	0.848828	2	FLNC	44.1176	2	MYH11	3.863636	2	DES	136	2	EHD2	0.916667
1	CNN1	0.5	3	DCN	0.843025	3	CAV1	43.8899	2	FLNC	3.863636	3	FLNC	124	3	PTRF	0.848485
1	FLNC	0.5	4	PTRF	0.819558	4	DES	42.18182	2	TAGLN	3.863636	4	MYH11	110	4	DCN	0.846154
1	TAGLN	0.5	5	OGN	0.811459	5	MYH11	41.04856	5	TGFB1	3.818182	5	CAV1	104	5	CNN1	0.836364
1	TGFB1	0.5	6	TAGLN	0.799541	6	LUM	36.39206	6	EGFR	3.772727	6	LUM	96	6	TAGLN	0.78022
7	EGFR	0.333333	7	EHD2	0.787593	7	TGFB11	26.09452	6	DCN	3.772727	7	TGFB1	90	7	LGALS1	0.763636
7	DCN	0.333333	8	CNN1	0.780531	8	EGFR	16.78283	6	LUM	3.772727	8	TAGLN	68	8	EGFR	0.74359
7	EHD2	0.333333	9	EGFR	0.74084	9	LGALS1	13.02475	6	CAV1	3.772727	9	EGFR	54	8	LUM	0.74359
7	OGN	0.333333	10	COL1A2	0.72107	10	TAGLN	10.84149	10	CNN1	3.727273	10	CNN1	38	8	CAV1	0.74359

TABLE S3. Clinical staging characteristics of patients included in the validation phase of the study (*n* = 40)

S. No.	Clinical sample	Phenotype	Staging
1	Sample 1	Cancer	pT2N0
2	Sample 2	Cancer	pT2N2
3	Sample 3	Cancer	pT2aN0
4	Sample 4	Cancer	pT3pNo
5	Sample 5	Cancer	pT1bpN0
6	Sample 6	Cancer	pT2apN0
7	Sample 7	Cancer	pT2b pN0 cM0
8	Sample 8	Cancer	pT4N1M1
9	Sample 9	Cancer	pT2apN1cM0
10	Sample 10	Cancer	pT3pN1cM0
11	Sample 11	Cancer	pT3N0Mx
12	Sample 12	Cancer	pT2bN0
13	Sample 13	Cancer	pT2a
14	Sample 14	Cancer	pT3 No
15	Sample 15	Cancer	pT3pN0
16	Sample 16	Cancer	pT3pN0cM0
17	Sample 17	Cancer	pT3 pN0
18	Sample 18	Cancer	pT1b pNo
19	Sample 19	Cancer	pT3pN0
20	Sample 20	Cancer	pT3pN1
21	Sample 21	Cancer	pT2pN0Mx
22	Sample 22	Cancer	ypT3N0Mx
23	Sample 23	Cancer	pT2 pN0 cM0
24	Sample 24	Cancer	pT2apN1
25	Sample 25	Cancer	pT1bpN0
26	Sample 26	Cancer	pT3pN0
27	Sample 27	Cancer	pT1b pNo
28	Sample 28	Cancer	pT2b pNo
29	Sample 29	Cancer	pT3N1Mx
30	Sample 30	Cancer	pT3N1
31	Sample 31	Cancer	pTis pN0 cM0
32	Sample 32	Cancer	pT3No
33	Sample 33	Cancer	pT1bpN0Mx
34	Sample 34	Cancer	T1a No
35	Sample 35	Cancer	pT3pM1
36	Sample 36	Cancer	PT1b No
37	Sample 37	Cancer	pT1b pNo
38	Sample 38	Cancer	pT2 pN1 cM0
39	Sample 39	Cancer	pT4N1M0
40	Sample 40	Cancer	pT2aN0

TABLE S4: Patients' clinico-pathological characteristics correlated with S100P, PIGR, C1QBP, TAGLN, CNN1 concentration at their respective optimum cut-off values (5.353 ng/µg, 1.068 ng/µg, 10.32 ng/µg, 12.66 ng/µg and 19.79 ng/µg respectively) (n = 80)

Characteristics		Frequency (*n* = 80)	S1OOP		*P*	PIGR		*P*	C1QBP		*P*	TAGLN		*P*	CNN1		*P*
			Low expression (<5.35 ng/µg)	High expression (≥5.35 ng/µg)		Low expression (<1068 pg/µg)	High expression (≥1068 pg/µg)		Low expression (<10.32 ng/µg)	High expression (≥10.32 ng/µg)		Low expression (<12.66 ng/µg)	High expression (≥12.66 ng/µg)		Low expression (<19.79 ng/µg)	High expression (≥19.79 ng/µg)	
Median age, years: 52																	
Age	<60	56 (70%)	37	19	0.67	18	38	0.06	38	18	0.52	24	32	0.25	25	31	0.2
	≥60	24 (30%)	17	7		13	11		18	6		7	17		7	17	
Gender	Female	42 (52.5%)	25	17	0.10	14	28	0.3	27	15	0.24	18	24	0.43	21	21	0.05
	Male	38 (47.5%)	29	9		17	21		29	9		13	25		11	27	
Smoking	Yes	5 (6.25%)	3	2	0.71	2	3	0.95	4	1	0.61	3	2	0.313	3	2	0.35
	No	75 (93.7%)	51	24		29	46		52	23		28	47		29	46	
Chewing	Yes	3(3.75 %)	2	1	0.97 0.71	0	3	0.16 0.95	3	0	0.25 0.61	1	2	0.84 0.313	1	2	0.81 0.35

Smoking	Yes	5 (6.25%)	3	2		2	3		4	1		3	2		3	2	
Alcohol	Yes	6(7.5%)	4	2	0.96	2	4	0.77	4	2	0.85	3	3	0.56	3	3	0.60
	No	74(92.5%)	50	24		29	45		52	22		28	46		29	45	
Dietary habit	Vegetarian	44 (55%)	28	16	0.41	14	30	0.16	28	16	0.17	15	29	0.34	21	23	0.12
	Non-Vegetarian	36 (45%)	26	10		17	19		28	8		16	20		11	25	
Preoperative laboratory values																	
Urea (mg/dl) 10.00-40.00	<10.0	3 (3.7%)	2	1	0.99	1	2	0.83	3	0	0.51	1	2	0.83	1	2	**0.0077***
	10.00-40.00	71(88.75%)	48	23		27	44		49	22		27	44		25	46	
	>40.0	6 (7.5%)	4	2		3	3		4	2		3	3		6	0	
Creatine (mg/dl)	<0.5	12 (15%)	7	5	0.4	6	6	0.66	9	3	0.46	6	6	0.66	5	7	0.98
	0.50-1.00	56 (70%)	37	19		21	35		37	19		21	35		22	34	
	>1.00	12 (15%)	10	2		4	8		10	2		4	8		5	7	
Calcium (mg/dl)	<8.5	17 (21%)	12	5	**0.04***	7	10	0.37	12	5	0.99	6	11	0.33	5	12	0.18
	8.50-10.50	60(75%)	42	18		24	36		42	18		25	35		27	33	
	>10.50	3 (3.75%)	0	3		0	3		2	1		0	3		0	3	
SGOT (AST); U/L	<5.00	1 (1.25%)	0	1	0.23	0	1	0.48	0	1	0.07	0	1	0.67	1	0	0.38
	5.00-40.00	58 (72.5%)	38	20		21	37		38	20		22	36		24	34	

	>40.00	21 (26.25 %)	16	5		10	11		18	3		9	12		7	14	
SGPT (ALT); U/L	<5.00	0	0	0	0.41	0	0	0.69	0	0	0.99	0	0	0.51	0	0	0.29
	5.00-45.00	60 (75%)	39	21		24	36		42	18		22	38		22	38	
	>45.00	20 (25%)	15	5		7	13		14	6		9	11		10	10	
Total Protein, g/dL	<6.00	5 (6.25%)	2	3	0.37	1	4	0.23	2	3	0.32	3	2	0.09	2	3	0.63
	6.00-8.00	61 (76.2%)	41	19		22	39		44	17		26	35		26	35	
	>8.00	14 (17.5%)	10	4		8	6		10	4		2	12		4	10	
Albumin;g/dL	<3.50	14 (17.5%)	11	3	0.62	8	6	0.3	10	4	0.37	5	9	0.94	4	10	0.43
	3.50-5.00	63 (78.75 %)	41	22		22	41		45	18		25	38		26	37	
	>5.00	3(3.75 %)	2	1		1	2		1	2		1	2		2	1	
Globulin (g/dL)	<2.00	3 (3.75%)	2	1	0.19	1	2	0.25	1	2	0.07	2	1	0.07	1	2	**0.0001***
	2.00-3.50	59(73.75%)	37	22		20	39		39	20		26	33		16	43	
	>3.50	18 (22.5%)	15	3		10	8		16	2		3	15		15	3	
Blood test markers																	
AFP(ng/mL)	<8.78	3 (3.75%)	1	2	0.25	0	3	**0.045***	2	1	0.25	2	1	0.25	3	0	**0.05***
	≥8.78	1(1.25 %)	1	0		1	0		0	1		0	1		0	1	

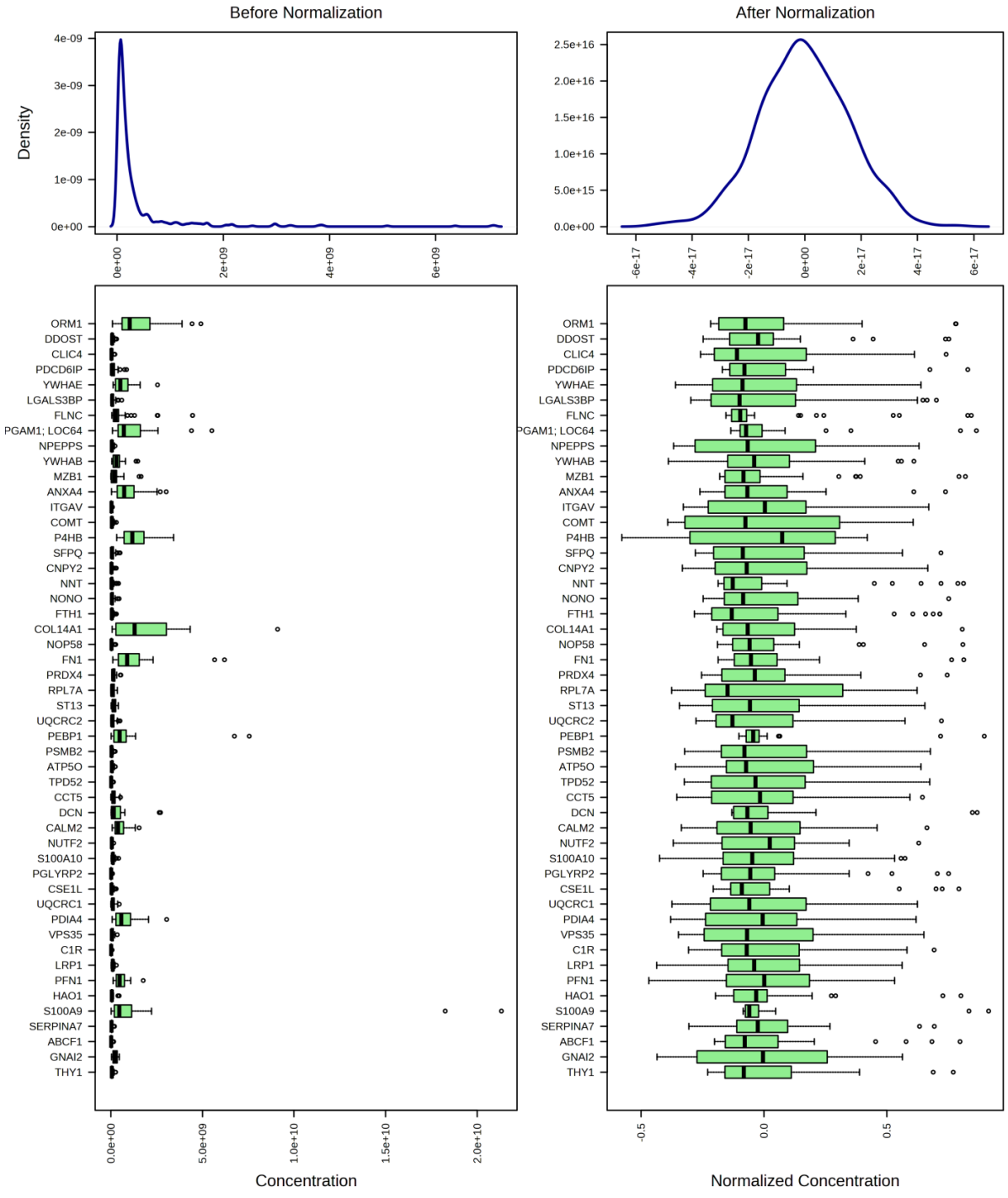
	NT	76(95%)	52	24		30	46		54	22		29	47		29	47	

*$P < 0.05$; NT: Not tested.

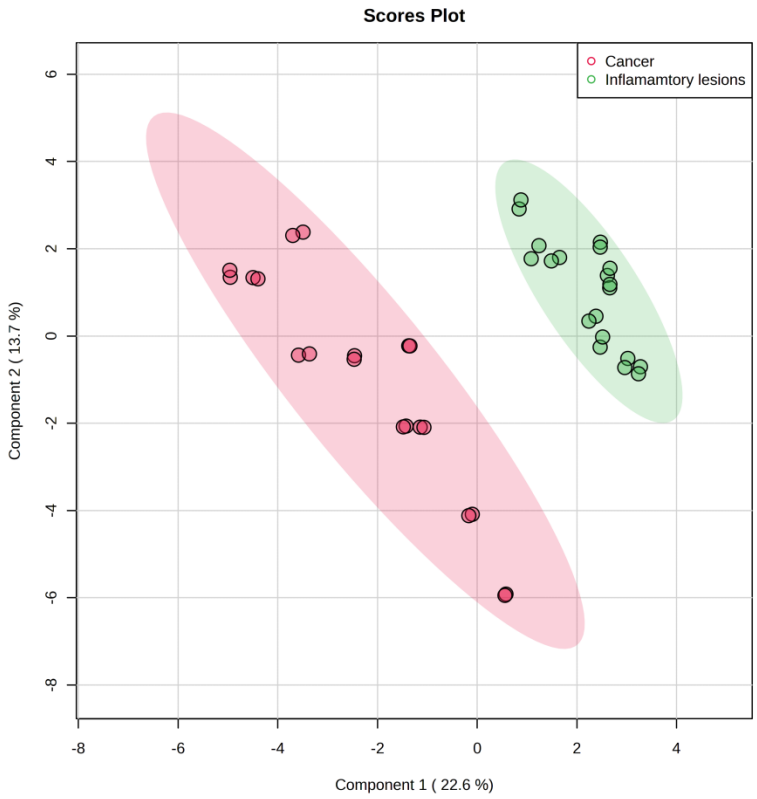
TABLE S5. Multivariate analysis of prognostic factors in GBC cases (*n* = 40

Characteristics		*n*	Multivariate cox regression (S100P, PIGR, C1QBP, TAGLN, CNN1)		Multivariate cox regression (Upregulating proteins-S100P, PIGR, C1QBP)		Multivariate cox regression (Downregulating proteins-TAGLN, CNN1)	
			Hazard ratio [95% CI]	*P* value	Hazard ratio [95% CI]	*P* value	Hazard ratio [95% CI]	*P* value
S100P	<5.35ng/µg (ref)	19	1.45[0.394-5.35]	0.57	1.7[0.535-5.42]	0.36	0.321[0.07-1.45]	0.14
	≥5.35 ng/µg	21						
PIGR	<1068 pg/µg (ref)	8	1.45[0.29-7.17]	0.644	0.768[0.198-2.97]	0.703	1.54[0.47-5.09]	0.47
	≥1068 pg/µg	32						
C1QBP	<10.32 ng/µg (ref)	19	2.07[0.525-8.21]	0.297	0.97[0.33-2.84]	0.967	-	-
	≥10.32 ng/µg	21						
TAGLN	<12.66 ng/µg (ref)	28	0.17[0.03-0.99]	**0.049***	-	-	-	-
	≥12.66 ng/µg	12						
CNN1	<19.79 ng/µg (ref)	28	1.18[0.32-4.23]	0.79	-	-	-	-
	≥19.79 ng/µg	12						

***P* < 0.05.**


Before Normalization
After Normalization
Density
4e-09
3e-09
2e-09
1e-09
0e+00
2.5e+16
2.0e+16
1.5e+16
1.0e+16
5.0e+15
0.0e+00
0e+00
2e+09
4e+09
6e+09
-6e-17
-4e-17
-2e-17
0e+00
2e-17
4e-17
6e-17
ORM1
DDOST
CLIC4
PDCD6IP
YWHAE
LGALS3BP
FLNC
PGAM1; LOC64
NPEPPS
YWHAB
MZB1
ANXA4
ITGAV
COMT
P4HB
SFPQ
CNPY2
NNT
NONO
FTH1
COL14A1
NOP58
FN1
PRDX4
RPL7A
ST13
UQCRC2
PEBP1
PSMB2
ATP5O
TPD52
CCT5
DCN
CALM2
NUTF2
S100A10
PGLYRP2
CSE1L
UQCRC1
PDIA4
VPS35
C1R
LRP1
PFN1
HAO1
S100A9
SERPINA7
ABCF1
GNAI2
THY1
0.0e+00
5.0e+09
1.0e+10
1.5e+10
2.0e+10
-0.5
0.0
0.5
Concentration
Normalized Concentration

(A)



(B)

FIGURE S1. (A) Normalization of differentially expressed proteins. Left panel represents before normalization curve and right panel represents after normalization curve. **(**B) 2D Score plot to differentiate gallbladder inflammatory lesions and cancer. Principal component analysis (PCA): Principal Component 1 (PC1) and Principal Component 2 (PC2) are plotted in x and y axes, respectively, and together have 36.3% of variation. Pink and green ovals represent the clustering regions of cancer and inflammatory lesion groups, respectively with 95% confidence interval.

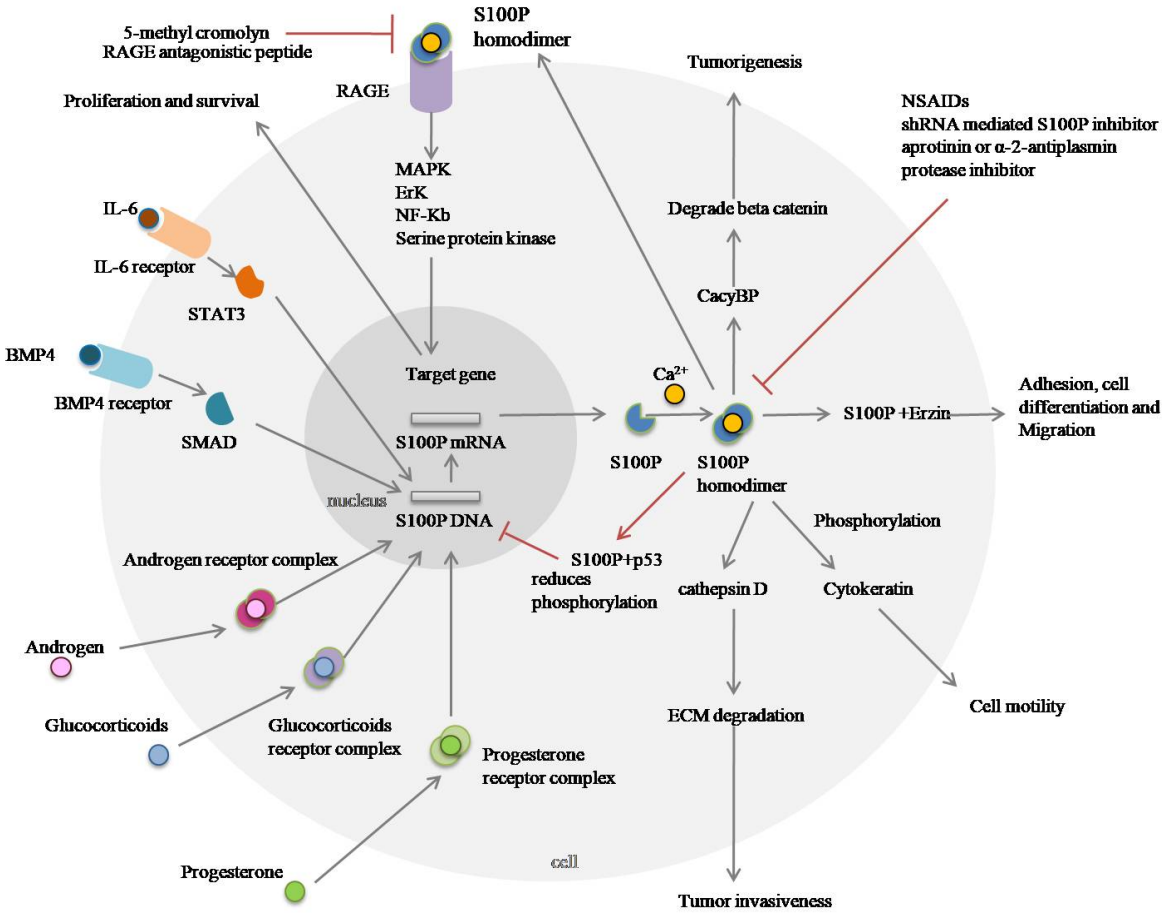


FIGURE S2. Some signal molecules such as glucocorticoids, androgen, progesterone, BMP4 (activates SMAD) and IL-6 (activates downstream signaling molecule STAT3) regulate S100P transcriptional activity in cancer cells. The S100P protein is expressed in an inactive state and triggered by calcium ions to form the active homodimers. Further, this S100P homodimer formed interacts with target proteins such as Erzin, which is the multidomain protein and has a role in adhesion, cell differentiation, and mirgration. S100P dimer activates the cytokeratins via phosphorylation and reduces the actin which ultimately affects the cell motility in cancer cells. The Receptors of advanced glycation end products (RAGE) is an immunoglobulin superfamily member receptor that is present on the cell surface. S100P acts as a ligand that binds to RAGE and activates the mitogen-activated protein kinase (MAPK), nuclear factor-KB (NF-KB), extracellular-regulated kinase (ERK) and serine protein kinase pathways. These activated pathways lead to cell proliferation and survival. Another signal transduction pathway is an interaction of CacyBP/SIP and S100P which leads to the degradation of beta-catenin and regulate tumorigenesis in the tumor cell. The S100P homodimer upregulate the cathepsin D expression level which further leads to extracellular matrix degradation and ultimately increases the tumor invasiveness. There is a feedback loop

in which a high level of S100P suppresses the endogenous S100P mRNA. The glucocorticoids-mediated S100P pathway plays an important role in cancer therapies. During tumor progression, some nonsteroidal anti- inflammatory drugs (NSAID) also impact S100P expression. In the pancreatic cells, 5-methyl cromolyn is the S100P inhibitor that binds to the RAGE receptor and inhibits tumor metastasis and growth. Another therapeutic pathway is the involvement of RAGE antagonistic peptide that blocks RAGE and S100P binding. The invasion of S100P positive cells is also suppressed by the protease inhibitors such as aprotinin or α-2-antiplasmin.

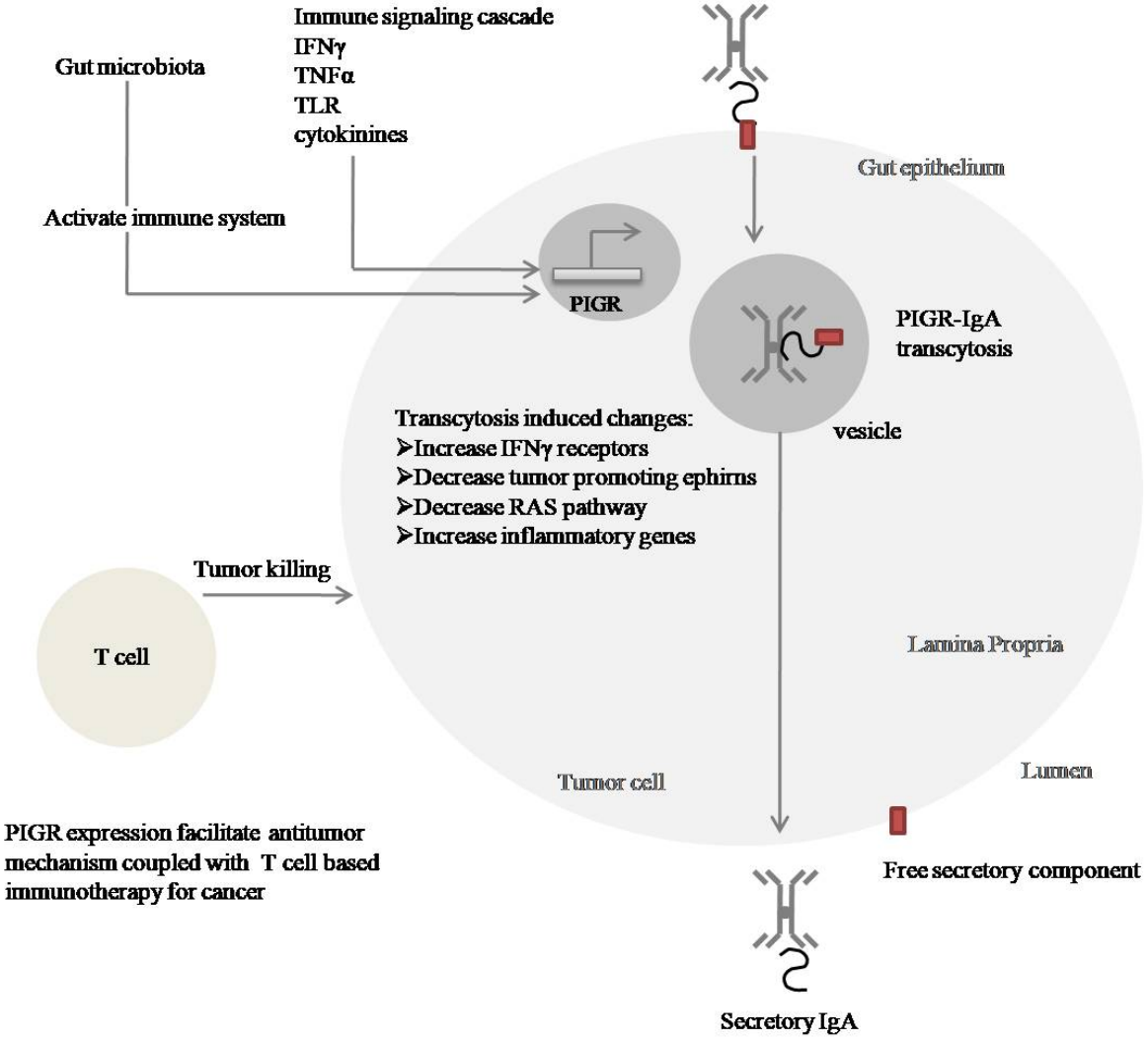


FIGURE S3. PIGR signalling pathway in cancer. Mechanistically, PIGR binds to dimerized IgA and IgM on the basolateral surface of the gut epithelium which is followed by endocytosis of IgA into vesicles. The PIGR transports IgA and becomes a part of the secreted IgA molecule which is further transported from the lamina propria across the epithelial barrier to the mucosal lumen. Therefore, PIGR act as a linkage between innate and adaptive immune responses. The PIGR expression is regulated by immune system mediators which include interferon γ (IFNγ) and tumor necrosis factor alpha (TNFα). Various immune signaling cascades (Toll like receptor (TLR) activation and inflammatory cytokinin signaling) also have a direct impact on the upregulation of PIGR gene. NF-κB activation subsequently upregulated the

transcytosis of PIGR: pIg. PIGR-IgA transcytosis induces transcriptional changes in the enhancement of inflammatory pathways in cancer cells which includes upregulation of interferon-gamma receptors, downregulation of tumor-promoting ephrins and antagonizes the RAS pathway which further sensitizes the tumor cells to cytolytic killing by T cells. Thus tumor dependent antibody facilitates the killing of cancer cells by antibody-dependent cellular cytotoxicity (ADCC) and phagocytosis (ADCP) mechanism.

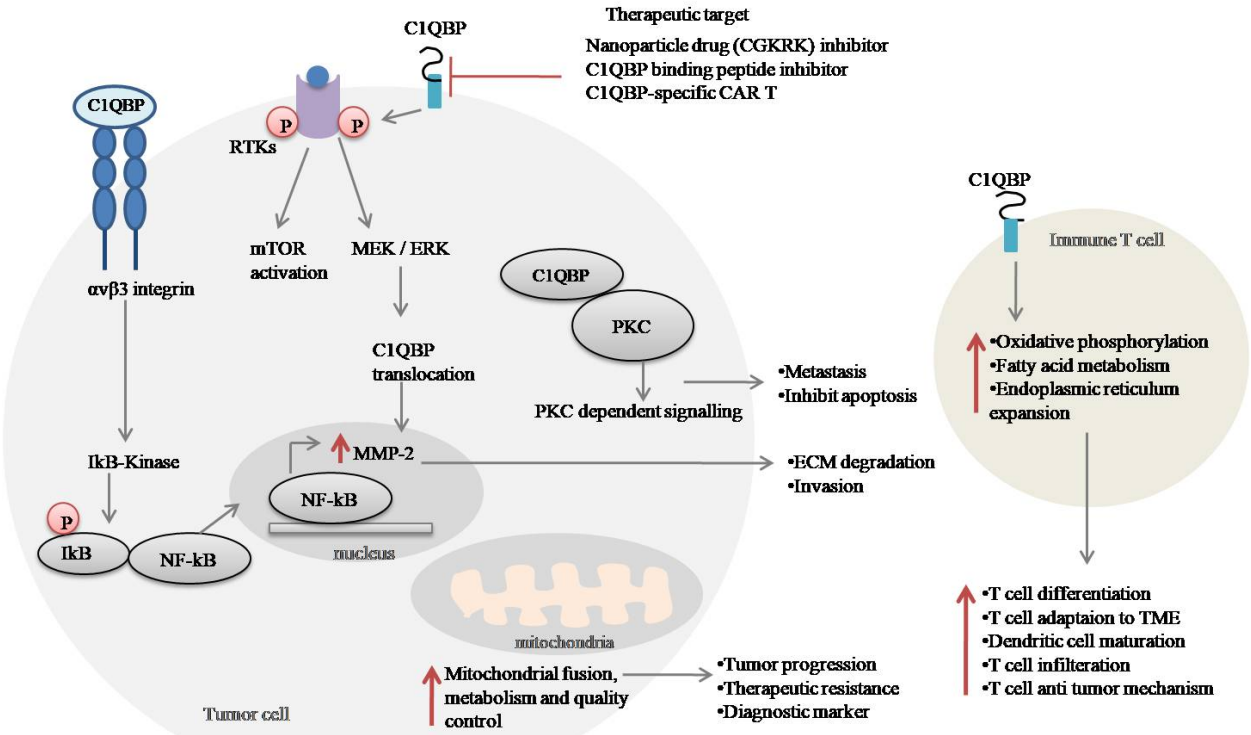


FIGURE S4. **C1QBP function as a ligand in cancer cell signalling pathway.** C1QBP binds to αvβ3 integrin and recruits IkB-Kinase that further phosphorylate IkB and translocate NF-kB to nucleus promoting the expression of metalloproteinase which leads to ECM degradation and invasion. C1QBP also phosphorylate receptor tyrosine kinase that activates mTOR and MEK/ERK pathway. ERK induces C1QBP translocation to nucleus. C1QBP binds to PKC and activates PKC dependent signalling which is cruical for metastasis and apoptosis inhibition. In mitochondria, C1QBP increases the mitochondrial fusion, metabolism and quality control which promote tumor progression, therapeutic resistance and diagnostic marker. In immune cells, C1QBP increases the oxidative phosphorylation, fatty acid metabolism and endoplasmic reticulum expansion that lead to T cell differentiation, T cell adaptation to tumor micro environment (TME), dendritic cell maturation, T cell infilteration and T cell antitumor mechanism. C1QBP specific CAR T cell approach is a therapeutic target for cancer treatment. Nanoparticle drug (CGKRK) and C1QBP binding peptide inhibitor also act as therapeutic target.

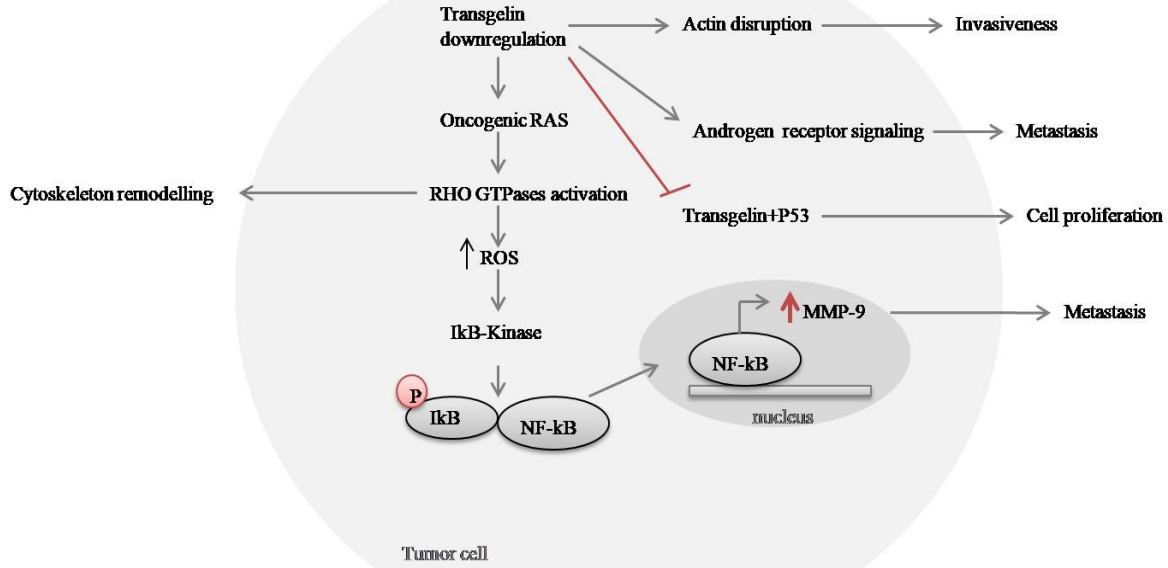


FIGURE S5. **TAGLN signaling pathway involved in cancer.** TAGLN act as a tumor suppressor gene and is found downregulated in tumor cells. Down regulation of TAGLN leads to actin disruption and causes cell invasiveness in tumor cells. In signal transduction pathway, decreased expression of TAGLN activates oncogenic RAS pathway that further activates RHO GTPases which is involved in cytoskeleton remodelling and increases the reactive oxygen species (ROS). Further, IkB-Kinase is recruited that phosphorylate IkB and translocate the NF-kB to nucleus. This leads to overexpression of metallomatrixprotease-9 (MMP9) and causes metastasis. The decreased expression of TAGLN activates androgen receptor signaling that causes metastasis. The low expression of TAGLN inhibits the binding of trangelin to p53 that causes cell proliferation.

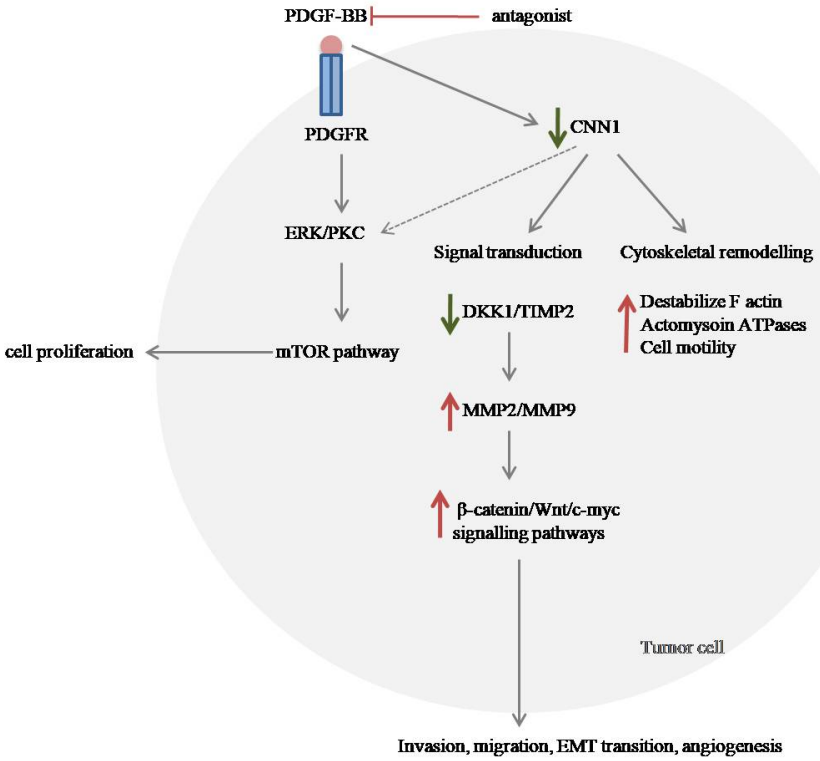


FIGURE S6. **CNN1 signaling pathway involved in cancer.** CNN1 recruits PKC/ERK due to its structural domain and further activates mTOR pathway leading to cell proliferation. Malignant tumor cell excrete various growth factors such as PDGR-BB that further suppresses the CNN1 expression. CNN1 plays an import role in cytoskeletal remodelling and signal transduction due to its structural domain. The decreased CNN1 expression, destabilize the F actin and further enhances the cell motility. In signal transduction, low expression of CNN1 decreases the dickkopf-1 (DKK1) and tissue inhibitor of metalloproteinases-2 (TIMP-2). This leads to over expression of metalloproteinases (MMP2/MMP9) and activates β-catenin/Wnt/c-myc signalling pathways that results into cell invasion, migration, epithelial-mesenchymal transition (EMT) and angiogenesis in tumor cells. PDGF antagonist such as neomycin may act as therapeutic target for cancer treatment.