



OPEN Comprehensive analysis of inhibin- β A as a potential biomarker for gastrointestinal tract cancers through bioinformatics approaches

Rohit Kumar Verma¹, Prashant Kumar Srivastava²✉ & Ashutosh Singh¹✉

Inhibin, β , which is also known as INHBA, encodes a protein that belongs to the Transforming Growth factor- β (TGF- β) superfamily, which plays a pivotal role in cancer. Gastrointestinal tract (GI tract) cancer refers to the cancers that develop in the colon, liver, esophagus, stomach, rectum, pancreas, and bile ducts of the digestive system. The role of INHBA in all GI tract cancers remains understudied. By utilizing GEPIA2, which uses transcriptomic data from TCGA, we examined the expression of INHBA across different GI tract cancers. The results revealed consistent upregulation of INHBA in all TCGA GI tract cancers, except for liver hepatocellular carcinoma, where it showed downregulation compared to normal tissues, along with GTEx normal samples. Significant differences in INHBA expression were noted in adenocarcinomas of the colon, pancreas, rectum, and stomach, while no such differences were observed in cholangiocarcinoma and liver cancer. Moreover, a comprehensive bioinformatics analysis has been done to demonstrate that the differences in expression levels are significantly related to pathological tumor stages and prognosis in different GI tract cancers. Mucinous adenocarcinoma, esophageal squamous cell carcinoma, and stomach adenocarcinoma show a higher frequency of INHBA alteration and are primarily linked to mutations and amplifications. DNA methylation, immune infiltration, functional enrichment analysis, the genes associated with INHBA, and survival analysis in all TCGA GI tract cancers have been extensively analyzed. In colon and stomach cancers, increased INHBA expression significantly correlates with poorer overall survival (OS). However, in colon and pancreatic adenocarcinoma, higher expression is significantly associated with worse disease-free survival (DFS). Additionally, INHBA expression exhibited a positive correlation with cancer-associated fibroblasts across all gastrointestinal (GI) tract cancers. The KEGG pathway analysis revealed that INHBA and its interacting proteins are involved in several pathways, including TGF-beta signaling, Signalling pathways regulating pluripotency of stem cells, colorectal cancer, pancreatic cancer, AGE-RAGE signaling, and so on as major pathways. These findings demonstrate that INHBA could serve as a potential biomarker therapeutic target for GI tract cancer.

Keywords INHBA Inhibin- β A, Transforming growth factor- β , Gastrointestinal tract cancers, Prognostic biomarker, DNA methylation

Abbreviations

INHBA	Inhibin subunit beta A
GI	Gastrointestinal
CPTAC	Clinical Proteomic Tumor Analysis Consortium
BP	Biological process
CC	Cellular component
DFS	Disease-free survival
GEPIA	Gene expression profiling interactive analysis
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
GTEx	Genotype-tissue expression

¹Department of Life Sciences, School of Natural Sciences (SONS), Shiv Nadar Institution of Eminence, Delhi NCR, India. ²National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, UK. ✉email: prashant.srivastava@imperial.ac.uk; ashutosh.singh@snu.edu.in

TIMER	Tumor immune estimation resource
AA	Amino acid
ACC	Adrenocortical carcinoma
BLCA	Bladder Urothelial Carcinoma
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
GBM	Glioblastoma multiforme
HNSC	Head and Neck squamous cell carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute Myeloid Leukemia
LGG	Brain Lower Grade Glioma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
OV	Ovarian serous cystadenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
TGCT	Testicular Germ Cell Tumors
THCA	Thyroid carcinoma
THYM	Thymoma
UCEC	Uterine Corpus Endometrial Carcinoma
UCS	Uterine Carcinosarcoma
CHOL	Cholangiocarcinoma
COAD	Colon adenocarcinoma
ESCA	Esophageal carcinoma
LIHC	Liver hepatocellular carcinoma
PAAD	Pancreatic adenocarcinoma
READ	Rectum adenocarcinoma
STAD	Stomach adenocarcinoma

Gastrointestinal cancer refers to cancer of the digestive tract, which includes cancer of the associated organ such as the esophagus, stomach, anal, pancreas, liver, gallbladder, colon, and rectum. These tumors often develop slowly and are difficult to recognize early. Treatment may include surgery, radiation therapy, chemotherapy, and targeted medications. There were approximately 19.9 million novel instances of cancer (including nonmelanoma skin cancers (NMSCs)) and over 9.7 million fatalities (including NMSCs) attributed to the disease reported in Globacancer 2022¹. By 2050, these numbers will increase if advancements in cancer prevention, diagnosis, and treatment don't progress; there could be around 35 million instances and around 17 million cancer-related fatalities¹. A major cause of death in the world is gastrointestinal (GI) cancer. Among all types of cancer, GI-tract cancer is the third most common cause of death worldwide, after lung cancer and breast cancer¹. Despite the heterogeneity, many GI-tract cancers share common molecular pathways. For example, aberrations in the Wnt/ β -catenin pathway, mutations in the TP53 gene, and alterations in the PI3K/AKT/mTOR pathway are observed across multiple GI tract cancers^{2,3}. Chronic inflammation is a known risk factor for many GI tract cancers⁴. Biomarkers associated with inflammatory processes or components of the tumor microenvironment, such as cytokines and immune cells, might be shared among different cancer types. Therefore, it is crucial to identify a common biomarker for GI-tract cancers that aims to leverage shared molecular features and pathways to improve early detection, streamline screening, and reduce costs.

Inhibin β A (INHBA), which is a member of the transforming growth factor- β (TGF- β) superfamily, is encoded by the activin/inhibin gene, which is located at the 7p14.1 chromosome⁵. It controls immunological responses, tissue homeostasis, cell proliferation, differentiation, and embryonic development. Transforming proliferation factor beta (TGF- β), which is known as a multifunctional protein, and regulates the immunological response, cell proliferation, and cell differentiation. It is essential for development, tissue repair, and the progression of disease⁶. It is frequently linked to being associated with metastasis, immune evasion, and tumor growth in malignancies such as gastric^{7,8}, cervical⁹, ovarian¹⁰, breast^{11,12}, pancreatic¹³, lung¹⁴, Head and Neck squamous cell carcinoma¹⁵, colorectal cancer¹⁶. The roles of INHBA and TGF- β in cancer are complicated and interconnected. INHBA can operate as a ligand for TGF- β receptors, activating the TGF- β signaling pathway when it binds to these receptors. The TGF-signalling pathway can promote cancer progression by different mechanisms¹⁷, like inhibiting the immune responses, apoptosis, or induce an epithelial-mesenchymal transition (EMT). In a study of gastric and esophageal adenocarcinoma, high activin protein expression was detected in tumors, and patients with elevated activin levels had significantly longer overall survival compared to those with lower activin expression. This indicates that the activin subunit inhibin beta A might serve as a prognostic biomarker¹⁸. In another study of gastro-esophageal (GE) cancers, several genes were identified as differentially expressed compared to normal tissues. The top overexpressed genes across all stages of GE cancers were CST1, INHBA, ACAN, HSP90AB1, and HSPD1 when examining plasma-proteome-associated genes. Machine learning techniques were used in this study to evaluate the predictive accuracy of these identified biomarkers with an area-under curve (AUC) > 0.9¹⁹.

The molecular mechanism through which INHBA contributes to the GI tract cancer progression needs to be elucidated, but the enhanced expression of INHBA has been associated with tumor progression, cell migration, and poor prognosis in various cancers, including colorectal cancer, gastric cancer, and esophageal cancer. The exact role of INHBA is not known. However, there are studies that reported and investigated the relationship of INHBA with various cancers. A study reported that INHBA and circularTHBS1 both shared the same miRNA response elements, and CircTHBS1 serves as a sponge for miR-204-5p. Dual luciferase reporter assay revealed that MiR-204-5p could bind to the 3'UTR region of INHBA; based on the results, they hypothesized that CircTHBS1 might regulate the INHBA expression and promote gastric cancer progression²⁰. In colon cancer also, INHBA upregulation is positively correlated with colon cancer cells malignant progression by the influence of BHIHE40 Transcription factor. SiRNA-based silencing of INHBA in HCT116 cell lines suggested the remarkable suppression of cell proliferation, and wound healing assay revealed the suppressed migration of colon cancer cells. Results suggested INHBA modulates cell proliferation as well as migration; however, the mechanism behind this is still unclear²¹. A comprehensive study also reported INHBA as a promising predictor of immunotherapy response as they evaluated the expression of INHBA and the correlation between INHBA and GC prognosis and revealed that INHBA expression was positively correlated with the type markers of CD8 + T cells, neutrophils, macrophages, and dendritic cell²². A recent study in colorectal cancer suggested an interplay of HNF1A-AS1/ miR-214/ INHBA axis regulation in TGF β / SMAD signaling based on their preliminary results, where they showed the expression of HNF1A-AS1 was negatively correlated with hsa-miR-214 and hsa-miR-214 targets INHBA, and expression INHBA and HNF1A-AS1 were found to be increased in colon adenocarcinoma (COAD), and rectum adenocarcinoma (READ)²³.

The interaction of INHBA with other receptor subunits/different molecular signatures may contribute to cancer, depending on their relative gene expression. The current study aimed to undertake a comprehensive bioinformatics analysis of the publicly available “The Cancer Genome Atlas (TCGA)” datasets and GEO datasets that provide expression datasets for many tumor types. Our study seeks to examine the role and mechanism of INHBA in all malignancies of the GI tract, including liver, cholangiocarcinoma, stomach, colon, rectum, esophageal, and pancreatic cancers. Different bioinformatics approaches are utilized to uncover potential highlights of INHBA in GI tract cancers. This study would be significant for a number of reasons, as it might aid in identifying patients on the verge of high risk for developing certain tumors and developing novel medications to target INHBA for cancer therapy to improve effectiveness.

Additionally, we investigated the putative molecular processes of INHBA in GI tract cancers using gene expression, patient survival analysis, gene mutation status, immunological infiltration, methylation, protein interactions, and pathway analysis. These results may improve our knowledge of how INHBA interacts biologically with various GI tract cancer types. However, the precise involvement of INHBA in diverse GI tract malignancies is still unexplored.

Results

INHBA expression in GI-tract cancers and their pathological stages

In this study, we investigated the possible significance of INHBA genes in cancer development and initiation in gastrointestinal tract cancers. Figure 1 represents the total number of TCGA cancer and TCGA normal and GTEx samples in (Fig. 1(A)) TCGA 7 GI tract cancers and (Fig. 1(B)) TCGA 24 other cancers from the GEPIA2 online server. We found significantly higher INHBA expression in cancers like Bladder Urothelial Carcinoma (BLCA), Breast invasive carcinoma (BRCA), and Head and Neck squamous cell carcinoma (HNSC). However, notable higher expression of INHBA was found in Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC), Glioblastoma multiforme (GBM), Kidney renal clear cell carcinoma (KIRC), Acute Myeloid Leukemia (LAML), Brain Lower Grade Glioma (LGG), Lung squamous cell carcinoma (LUSC), Pheochromocytoma and Paraganglioma (PCPG), Pancreatic adenocarcinoma (PRAD), Sarcoma (SARC), and Thyroid carcinoma (THCA) when compared with the normal including GTEx normal. INHBA was found to be low expressed in cancers such as Adrenocortical carcinoma (ACC), Kidney Chromophobe (KICH), Kidney renal papillary cell carcinoma (KIRP), Ovarian serous cystadenocarcinoma (OV), Skin Cutaneous Melanoma (SKCM), Lung adenocarcinoma (LUAD), Testicular Germ Cell Tumors (TGCT), Thymoma (THYM), Uterine Corpus Endometrial Carcinoma (UCEC) and Uterine Carcinosarcoma (UCS) in Fig. 2(A). As for GI-tract cancers, INHBA was found to be upregulated in cancers with significant differences in Colon adenocarcinoma (COAD), Esophageal carcinoma (ESCA), Pancreatic adenocarcinoma (PAAD), Rectum adenocarcinoma (READ), and Stomach adenocarcinoma (STAD) but there is no significance was found in Cholangiocarcinoma (CHOL). However, it was found to be downregulated in Liver hepatocellular carcinoma (LIHC) when compared with normal and GTEx normal tissues, which suggests the role of INHBA in GI tract-related cancer Fig. 2(B). This also indicates that INHBA might play an important role in GI tract cancers. Each TCGA tumor (red) is accompanied by its matching normal data (blue). Here, the Y axis defines the number of transcripts per million ($\log_2(\text{TPM} + 1)$), and the X axis defines the various cancers with a significance level of (p -value < 0.05). We analyzed INHBA expression levels in various cancer types using data from TCGA, utilizing TIMER2, UALCAN, and TNMplot. Consistent results were observed with TIMER2 and UALCAN, showing upregulation of INHBA in all GI-tract cancers and downregulation in liver hepatocellular carcinoma (LIHC). In TNMplot, significant differences were illustrated on the x-axis, highlighted in red. Notably, data for cholangiocarcinoma was not available for this analysis, as shown in **Supplementary Figure S1 (A-C)**. To further investigate the proteomics data of INHBA in cancer patients, we analyzed the CPTAC dataset using UALCAN. Our analysis revealed that INHBA protein expression levels were higher in breast cancer, lung cancer, HNSC, and PAAD, and lower in OV and UCEC compared to normal tissues, as shown in **Supplementary Figure S1 (D)**. Among GI-tract cancers, proteomics data were available only for PAAD, where a significant difference was observed between cancer and normal samples (p -value < 0.001).

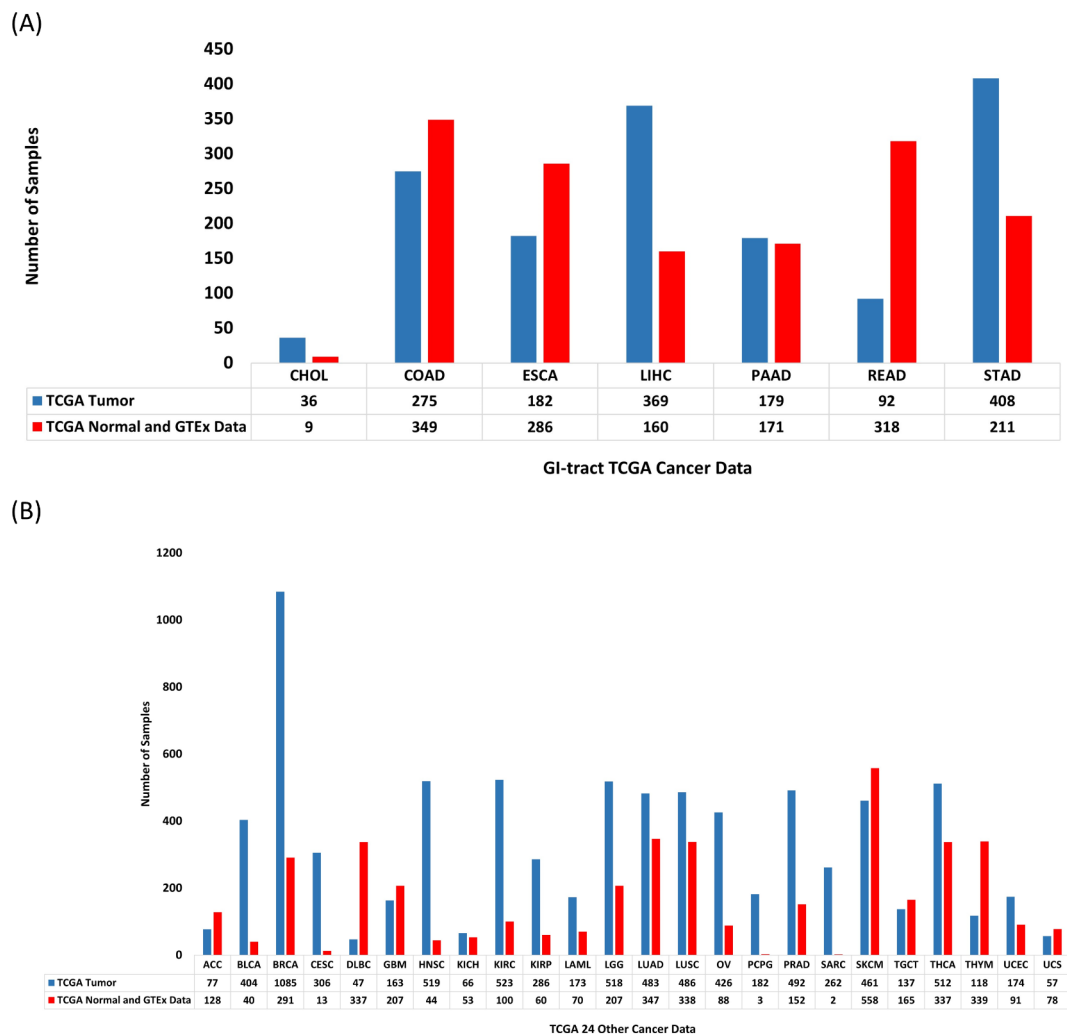


Fig. 1. The Number of samples in (A) TCGA 7 GI tract cancers and (B) TCGA 24 other cancers.

The GTEx data was used to measure the expression of INHBA in normal tissue, and we found that the expression of INHBA is higher in blood vessels and lungs. It was also observed that the expression of INHBA is relatively low in blood, muscle, pancreas, and spleen (Fig. 3(A)). Moreover, we examined significant differences (p -value < 0.001) in GI-tract TCGA cancers (colon, esophageal, pancreatic, liver, and stomach) and GTEx normal samples Fig. 3(B) and for cholangiocarcinoma and rectum adenocarcinoma, we didn't find any GTEx samples. Therefore, we compared the TCGA tumor with TCGA normal samples for cholangiocarcinoma and rectum adenocarcinoma. When the expression level of INHBA was analyzed using the microarray datasets for Intrahepatic Cholangiocarcinoma (GSE32879), Colorectal Cancer (GSE110224), Esophageal Squamous Cell Carcinoma (GSE38129), Pancreatic Ductal Adenocarcinoma (GSE71989), Liver Cancer (GSE29721), and Gastric Cancer (GSE54129), the box plot indicated that INHBA expression was upregulated in the tumor tissues of Intrahepatic Cholangiocarcinoma, Colorectal Cancer, Pancreatic Ductal Adenocarcinoma, Esophageal Squamous Cell Carcinoma, and Gastric Cancer compared to normal tissues and found to be significant (p -value < 0.05). However, no significant difference was observed in the case of liver cancer (Fig. 4). The number of samples of each case has been reported in **Supplementary Table 1**. Next, we plotted a receiver operating characteristic (ROC) curve to evaluate the diagnostic value of INHBA expression in several cancers: Intrahepatic Cholangiocarcinoma (GSE32879), Colorectal Cancer (GSE110224), Esophageal Squamous Cell Carcinoma (GSE38129), Pancreatic Ductal Adenocarcinoma (GSE71989), Liver Cancer (GSE29721), and Gastric Cancer (GSE54129). This analysis revealed high AUC values for Intrahepatic Cholangiocarcinoma (AUC=0.96), Colorectal Cancer (AUC=0.79), Esophageal Squamous Cell Carcinoma (AUC=0.89), Pancreatic Ductal Adenocarcinoma (AUC=0.82), and Gastric Cancer (AUC=0.81), indicating strong diagnostic potential for these cancers. However, a lower AUC value was observed in Liver Cancer (AUC=0.49). These findings suggest that INHBA expression could serve as a prognostic indicator for GI-tract cancers, except for liver cancer, as shown in **Supplementary Figure S2(A-F)**.

Further, we also calculated the AUC value for pan-cancer to assess the accuracy of INHBA in tumors. Table 1 represents the AUC value in different cancers. Among them, the AUC value is higher than 0.7 in CHOL

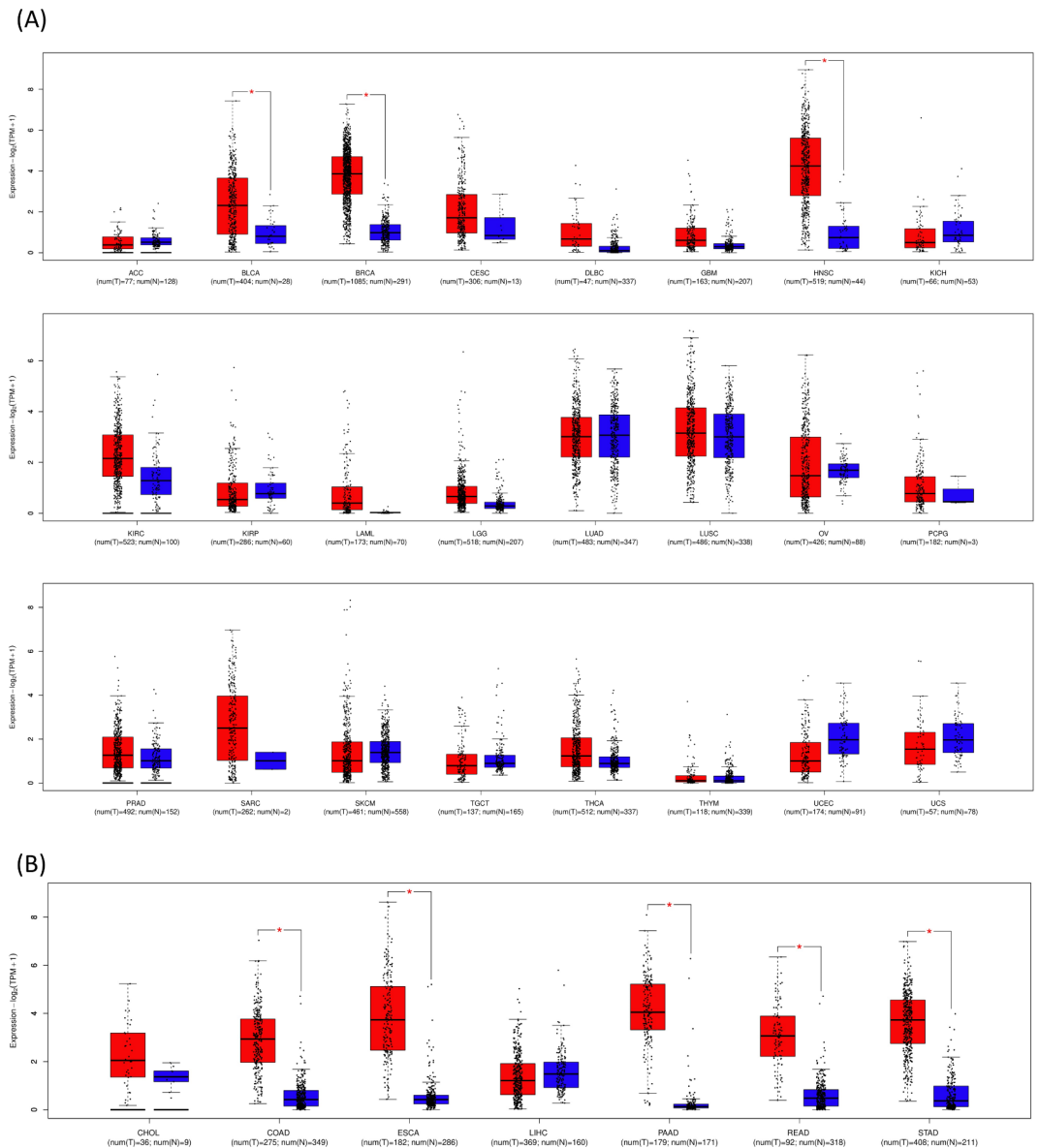


Fig. 2. INHBA expression across diverse cancers. (A) The expression of mRNA level (INHBA) in various TCGA cancers and normal tissues along with the GTex data was used as a control. (B) INHBA expression in GI cancers (CHOL, COAD, ESCA, LIHC, PAAD, READ, and STAD) and expression in normal tissues and GTex dataset with p -value < 0.05 .

(AUC=0.71), COAD (AUC=0.99), ESCA (AUC=0.86), READ (AUC=0.98), and STAD (AUC=0.96). The AUC value for LIHC (AUC=0.61) and PAAD (AUC=0.52) was observed to be lower than other GI-tract cancers. However, this suggests that INHBA could be used as a potential predictor of tumor prognosis. Using UALCAN, we also accessed the expression of INHBA in TCGA normal samples according to pathological stages of GI tract cancers Fig. 5(A-G). Additionally, **Supplementary Figure S3** shows the expression of INHBA according to nodal metastasis status, weight, age, gender, tumor grade, and other patient characteristics. We observed a significant difference in INHBA expression between normal tissues and various stages of COAD, and READ in the early stages, suggesting INHBA's involvement in cancer initiation. We also observed a notable difference in late-stage cancer when compared to the early-stage in STAD, which might suggest the role of INHBA in cancer progression and invasion. We haven't observed any significant difference in CHOL, PAAD, and LIHC in either the early or late stage of cancer when compared to the different stages of cancer (Fig. 5, p -value < 0.05) except for CHOL. We found a significant difference between Normal and Stage 1; for PAAD, a significant difference was observed between Stage 1 and Stage 2. Abutalebi et al., in their study, identified dysregulation of INHBA and PELATON in colorectal cancer (CRC) using bioinformatics and validated the upregulation of PELATON in CRC tissues through RT-qPCR and also checked for tumor purity, and response to immune checkpoint inhibitors in CRC, explored associations between INHBA expression focussing on COAD and READ²⁴. In our study, we also identified higher INHBA expression in COAD and READ. Furthermore, we

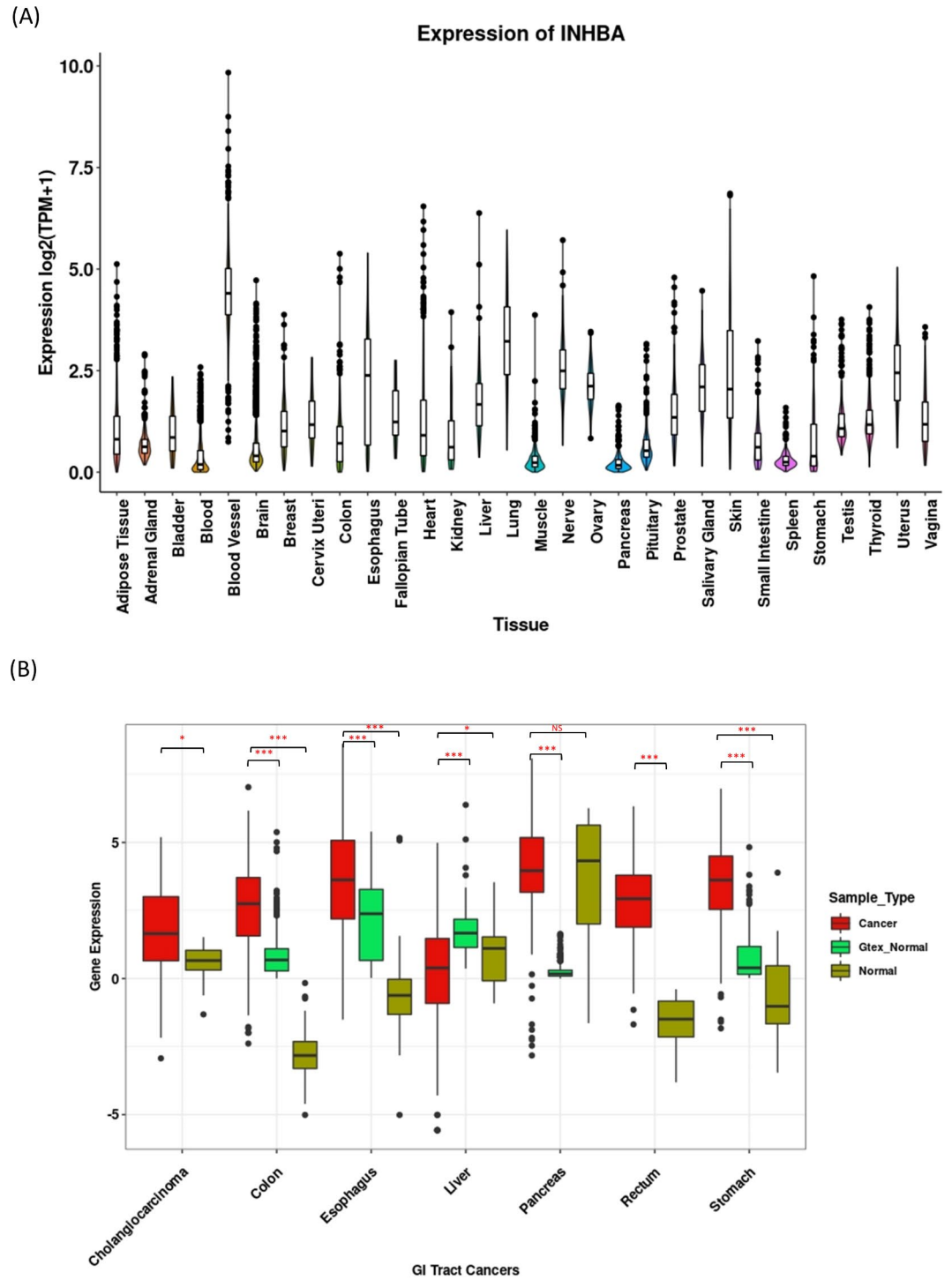


Fig. 3. A measure of INHBA expression in GTEx normal tissue. (A) GTEx data showing INHBA expression in different tissues. (B) Comparison of TCGA normal and GTEx normal tissue with GI tract cancer tissue (p-value < 0.001 (***), p-value < 0.01 (**), and p-value < 0.05 (*)).

conducted a comprehensive analysis across all GI tract cancers to uncover the broader mechanisms involving INHBA in the GI-tract cancers.

Survival analysis

To further examine the relationship between the INHBA expression and disease prognosis in patients with GI tract cancers, we took a median of INHBA expression as a cutoff for higher or lower expression of the patients (Cutoff-High(50%) and Cutoff-Low(50%)). This cutoff value was chosen because using the median is a common and statistical approach. It also works on equal group sizes, allowing for balanced data comparisons.

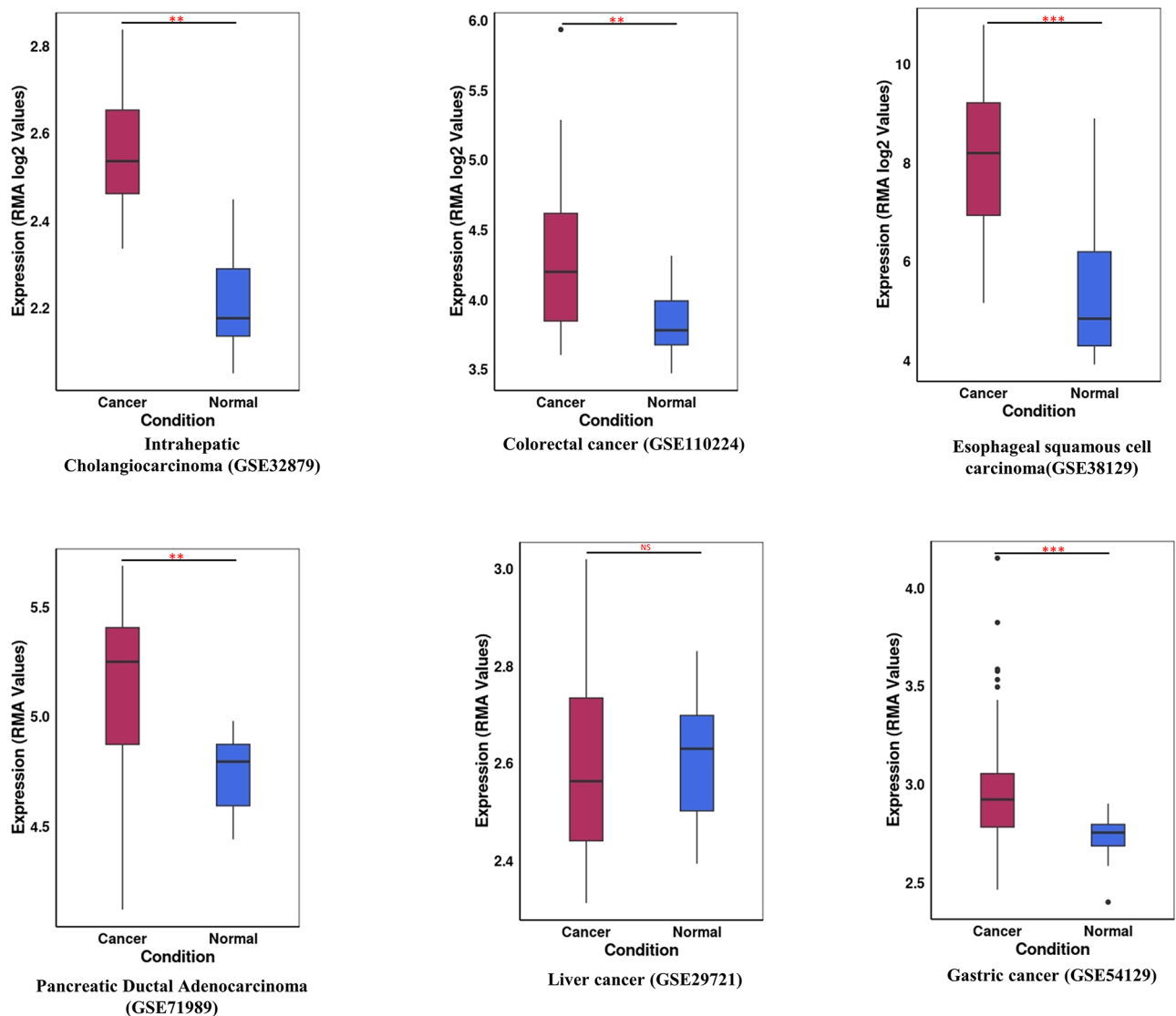


Fig. 4. Expression levels of INHBA in various microarray datasets for different cancers, including Intrahepatic Cholangiocarcinoma (GSE32879), Colorectal Cancer (GSE110224), Esophageal Squamous Cell Carcinoma (GSE38129), Pancreatic Ductal Adenocarcinoma (GSE71989), Liver Cancer (GSE29721), and Gastric Cancer (GSE54129), compared to normal tissues.

We observed that the cancer cases associated with higher expression of INHBA have worse overall survival (OS) in COAD (p-value=0.045, HR=1.6) and STAD (p-value=0.028, HR=1.4) in Fig. 6(A). When we did disease free survival (DFS) analysis, we observed a poor DFS for the patients with higher expression of INHBA in COAD (p-value=0.04, HR=1.6) and PAAD (p-value=0.049, HR=1.6) in Fig. 6(B). This indicates that the expression of INHBA is strongly associated with COAD, STAD, and PAAD, suggesting that INHBA expression correlated in these GI tract cancer prognosis. We also plotted the overall survival (OS) of other GI-tract cancers, including CHOL, ESCA, LIHC, PAAD, and READ, but did not observe any statistically significant p-values for these cancer types using GEPIA2, as shown in **Supplementary Figure S4 (A)**. Additionally, we compared these findings with the UALCAN tool, which indicated a significant (p-value<0.05) for COAD and PAAD. We also observed a p-value of 0.06 for STAD in the survival analysis, indicating that higher expression of INHBA has prognostic significance in predicting survival outcomes for patients with COAD, STAD, and PAAD, as shown in **Supplementary Figure S4 (B)**.

Gene alteration analysis of INHBA in GI-tract cancers

Here, we analyzed gene alteration of INHBA in Cholangiocarcinoma, Liver Hepatocellular Carcinoma, Colorectal, Esophageal, Pancreatic, and Stomach Adenocarcinoma for available TCGA samples using cBioPortal. We noted 3% of genetic alteration in all the GI tracts mentioned above in cancer, as shown in Fig. 7(A), with genetic alteration in cancer type in detail. Further, we found that when selecting the “cancer type detailed” type (>10%) of highest alteration of INHBA in (>10%) of Mucinous adenocarcinoma of the colon and rectum followed by

TCGA-Cancer	AUC	TCGA-Cancer	AUC
BLCA	0.71	LUSC	0.63
BRCA	0.94	MESO	0.63
CESC	0.45	PAAD	0.52
CHOL	0.71	PCPG	0.57
COAD	0.99	PRAD	0.57
ESCA	0.86	READ	0.98
GBM	0.64	SARC	0.76
HNSC	0.93	SKCM	0.67
KICH	0.78	STAD	0.96
KIRC	0.66	THYM	0.68
KIRP	0.70	THCA	0.54
LIHC	0.61	UCEC	0.59
LUAD	0.69	UVM	0.59

Table 1. The AUC values for INHBA in TCGA tumor expression and TCGA normal tissue expression in pan-cancer.

(>6%) in esophageal squamous cell carcinoma mainly of the “Mutation” type as shown in Fig. 7(B–C). The highest alteration was found in esophageal (>5%) and in stomach adenocarcinoma (>4%) when selected as a “cancer study” in cBioPortal. In esophageal adenocarcinoma, we found mutation, amplification, and deep deletion, whereas we found only mutation and amplification type in stomach adenocarcinoma. In colorectal adenocarcinoma, we found mutations and very low frequency of amplification types. In cholangiocarcinoma, we observed only mutation type. Additionally, we analyzed the association between INHBA expression and clinical survival across all available TCGA GI-tract cancers from cBioPortal. We didn’t observe any significant difference in overall survival (OS), disease specific survival (DSS), progression free survival (PFS), and disease free survival (DFS) with or without alteration of INHBA, as shown in Fig. 7(D–G) using the Kaplan–Meier plotter. A significant difference was observed for the Esophageal adenocarcinoma (p-value=0.05) case in OS with or without alteration of INHBA, as shown in Fig. 7(I), indicating poorer prognosis and shorter survival time with the altered group. We also showed the gene alteration of INHBA in all the GI-tract cancer studies other than TCGA from cBioPortal, as shown in **Supplementary Figure S5** (A). No significant statistical difference was observed in the survival analysis of other GI-tract cancers with or without alteration of INHBA, as shown in **Supplementary Figure S5** (B). Figure 7(H) depicts the different types of cancer in which the mutation loci correspond to the cases in INHBA as lollipop plots. The missense mutation was found to be associated mainly with the INHBA. We observed R308C alteration in INHBA, which was found to be associated with one case of each: esophageal adenocarcinoma, hepatocellular carcinoma, and colon adenocarcinoma. This site of alteration induced mutation in INHBA from Arginine (R) to Cysteine (C) at 308 sites in Fig. 7(H). We also observed that in the transforming growth factor beta-like domain (319–426 aa) of INHBA there were seven sites mainly missense mutation associated with one case in C350R (Mucinous adenocarcinoma of colon and rectum), one case in S356N (Stomach adenocarcinoma), two cases in R377C/H (Mucinous adenocarcinoma of colon and rectum), one case in R379W (Esophageal adenocarcinoma), one case in A385S (Colon Adenocarcinoma), one case in K395R (Mucinous adenocarcinoma of colon and rectum) and one case in N417S (Intestinal Type Stomach Adenocarcinoma) sites. Figure 7(J) shows the most frequent site (R308C) for genetic alteration using the alpha fold structure available for INHBA. Further studies into the functional consequences of INHBA gene alterations and their interactions with other molecular pathways could offer valuable insights for developing targeted cancer therapies.

Analysis of immune infiltrates with INHBA in GI-tract cancers

Tumor growth, progression, and metastasis are all impacted by the presence and activity of immune cells infiltrating the tumor microenvironment. We have used TIMER2 and explored the Cancer-associated fibroblasts (CAF) for a detailed study. CAFs have an essential role in tumor microenvironment, influencing immune infiltration and tumor progression. CAF plays a vital role in tumor progression by secreting inflammatory ligands, growth factors, and matrix proteins. Different algorithms have been used for CAF analysis, like EPIC, MCPOUNTER, XCELL, and TIDE, to find further the correlation between the expression of INHBA and the infiltrates of CAF in different TCGA cancers, emphasizing GI-tract cancer. We observed a positive correlation of INHBA expression in all GI tract cancers. Interestingly, most of the other TCGA cancers were found to be positively correlated with the expression of INHBA in CAFs, as shown in Fig. 8(A). We have also shown correlations between the infiltration level of different immune cells and the INHBA expression level in **Supplementary Figure S6**. In Fig. 8(B–H), we have demonstrated the results of the correlation plot between the infiltration level and INHBA expression level in CAF with the algorithm, which has the greatest correlation and significant p-value. We also used the old version of TIMER to analyze the correlation of infiltration levels in different immune systems and expression of INHBA, shown in Fig. 9(A), where we observed that in CHOL, the INHBA gene is positively correlated with B cells, CD8+ cells, CD4+ cells, macrophages, neutrophils, and dendritic cells while in COAD we observed INHBA is negatively correlation with B cells. Similar to COAD

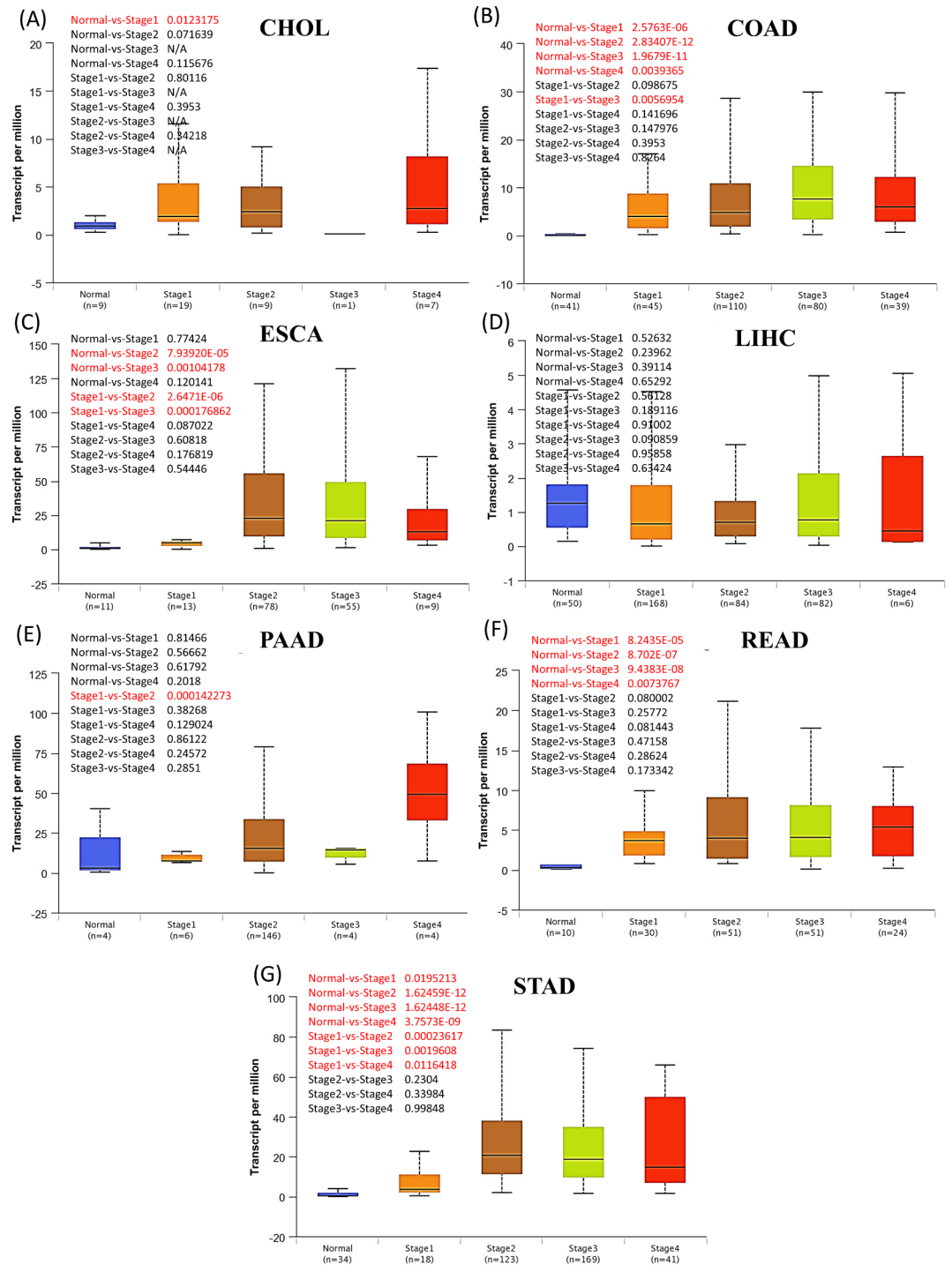


Fig. 5. Expression of INHBA across different stages of GI-tract cancers (A-G) where the Y-axis represents the transcript per million and the X-axis represents the cancer stages with the total number of samples in the individual stages.

observation, B cells were found to be negatively correlated in ESCA and STAD while positively correlated with other immune cell infiltrates. In LIHC and READ, we observed a similar correlation pattern as observed in CHOL, where all the infiltrates were found to be positively correlated, while in PAAD, we observed a negative correlation in T cells CD4+, while others were positively correlated. The higher expression of INHBA was found to be positively correlated with an increased immune cell infiltration, particularly macrophages, in all the GI-tract cancers, further leading to cancer progression. The significant p-value and partial correlation values have been provided in **Supplementary Table 2**. These findings imply that INHBA might contribute to the immune response during the development and progression of cancer. According to the cumulative survival analysis,

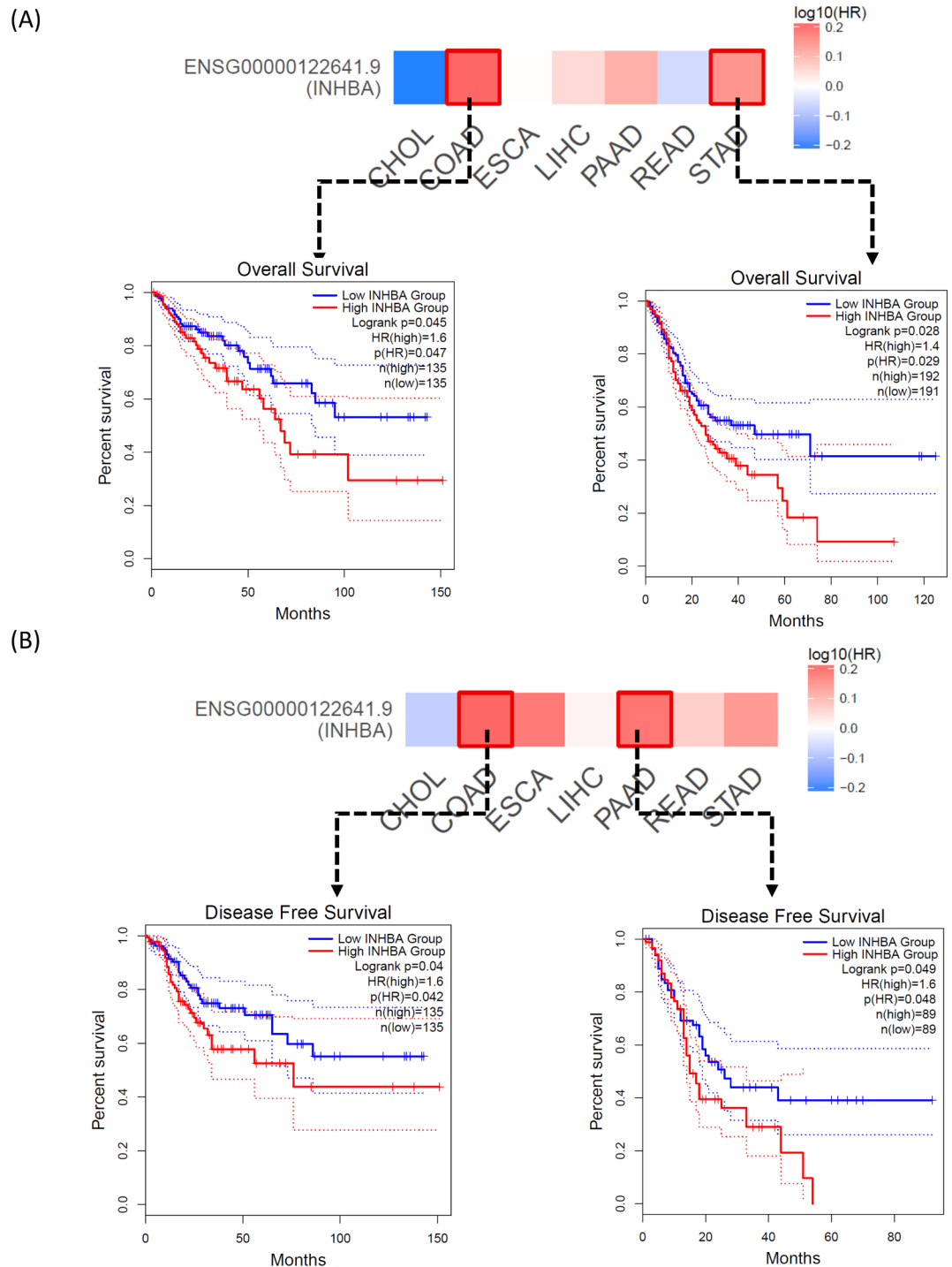


Fig. 6. Relationship between the gene expression of INHBA and survival prognosis for patients with GI tract cancer. Here, we have performed (A) overall survival analysis and (B) disease-free analysis.

neutrophils are strongly related to INHBA in CHOL (p -value < 0.05), while macrophages are associated with INHBA in STAD (p -value < 0.05), as shown in Fig. 9(B). We didn't observe any statistical difference in other GI-tract cancers while analyzing cumulative survival analysis. As shown in **Supplementary Figure S7**, we performed a somatic copy number alterations analysis to understand the distribution of each immunological subgroup in GI-tract cancers based on INHBA copy number status (deep deletion, arm-level deletion, diploid/normal, arm-level gain, and high amplification) using TIMER. This analysis was performed to understand how changes in the number of copies of certain genomic regions are associated with different immune cell populations within the tumor microenvironment. In COAD, significant differences were observed in B cells with arm-level deletion ($p < 0.001$), arm-level gain ($p < 0.01$), and high amplification ($p < 0.01$); in CD8+ T cells with arm-level gain

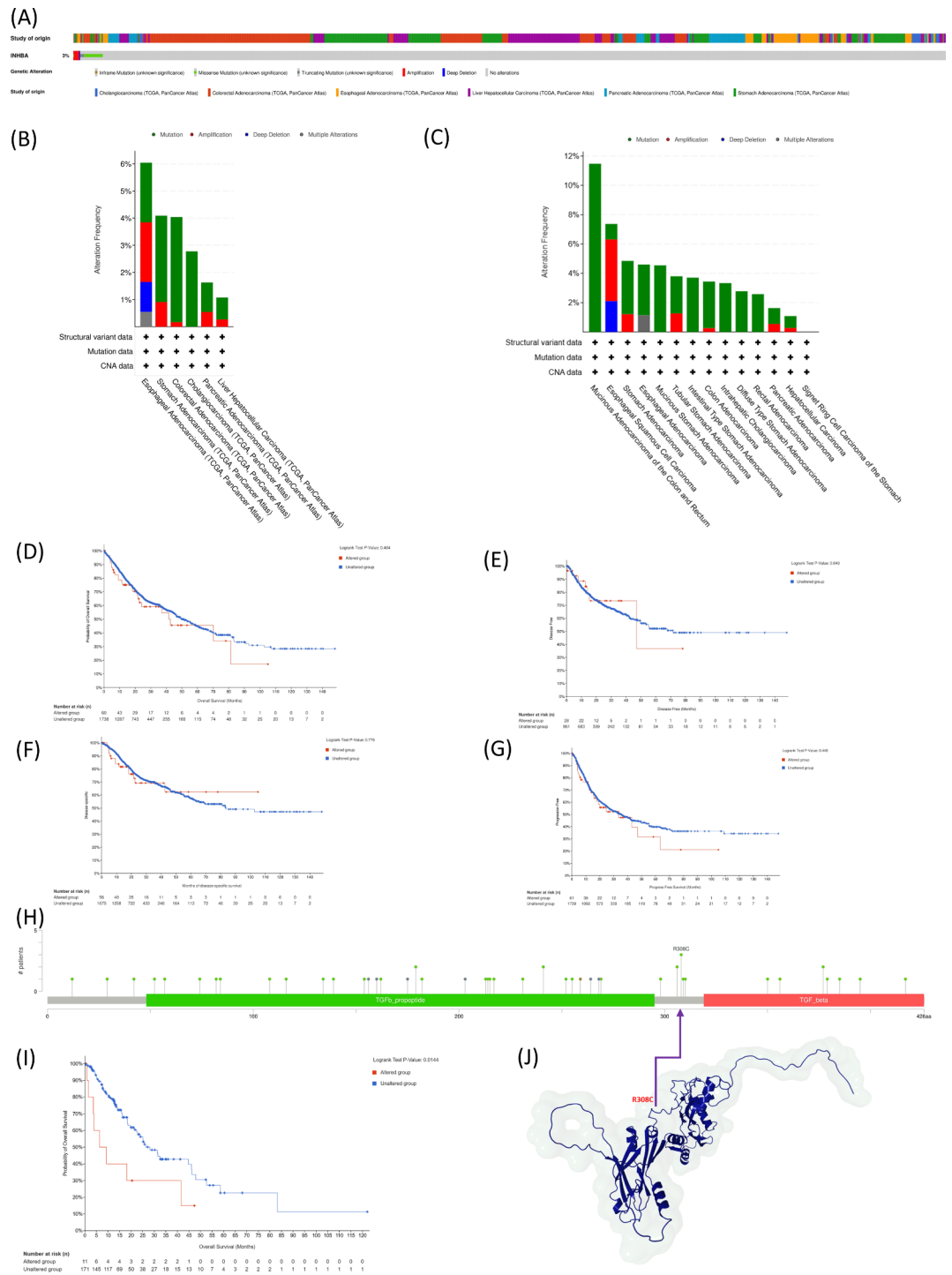


Fig. 7. Mutation analysis of INHBA in GI-tract cancers (A) Genetic alteration of INHBA in detailed cancer type (B-C) mutational frequency (D-G) The overall, disease-specific, progression-free and disease-free survival in patients with GI-tract cancers. (H) Lollipop plot showing the mutation site in GI-tract cancers (I) Overall survival in ESCA with correlation with mutation status (J) The alpha fold structure of INHBA and the mutational site showing which is common and present in COAD, ESCA, and LIHC.

($p < 0.001$); in neutrophils with arm-level gain ($p < 0.01$); and in dendritic cells with arm-level gain ($p < 0.001$). In ESCA, significant differences were found in B cells with high amplification ($p < 0.05$), CD8 + T cells with arm-level deletion ($p < 0.05$), and CD4 + T cells with both arm-level deletion and arm-level gain ($p < 0.05$). In PAAD, significant differences were seen in B cells with arm-level gain ($p < 0.01$) and CD4 + T cells with arm-level gain ($p < 0.01$). In READ, significant differences were observed in macrophages with arm-level gain ($p < 0.05$). In STAD, significant differences were found in B cells with arm-level deletion ($p < 0.01$), arm-level gain ($p < 0.01$),

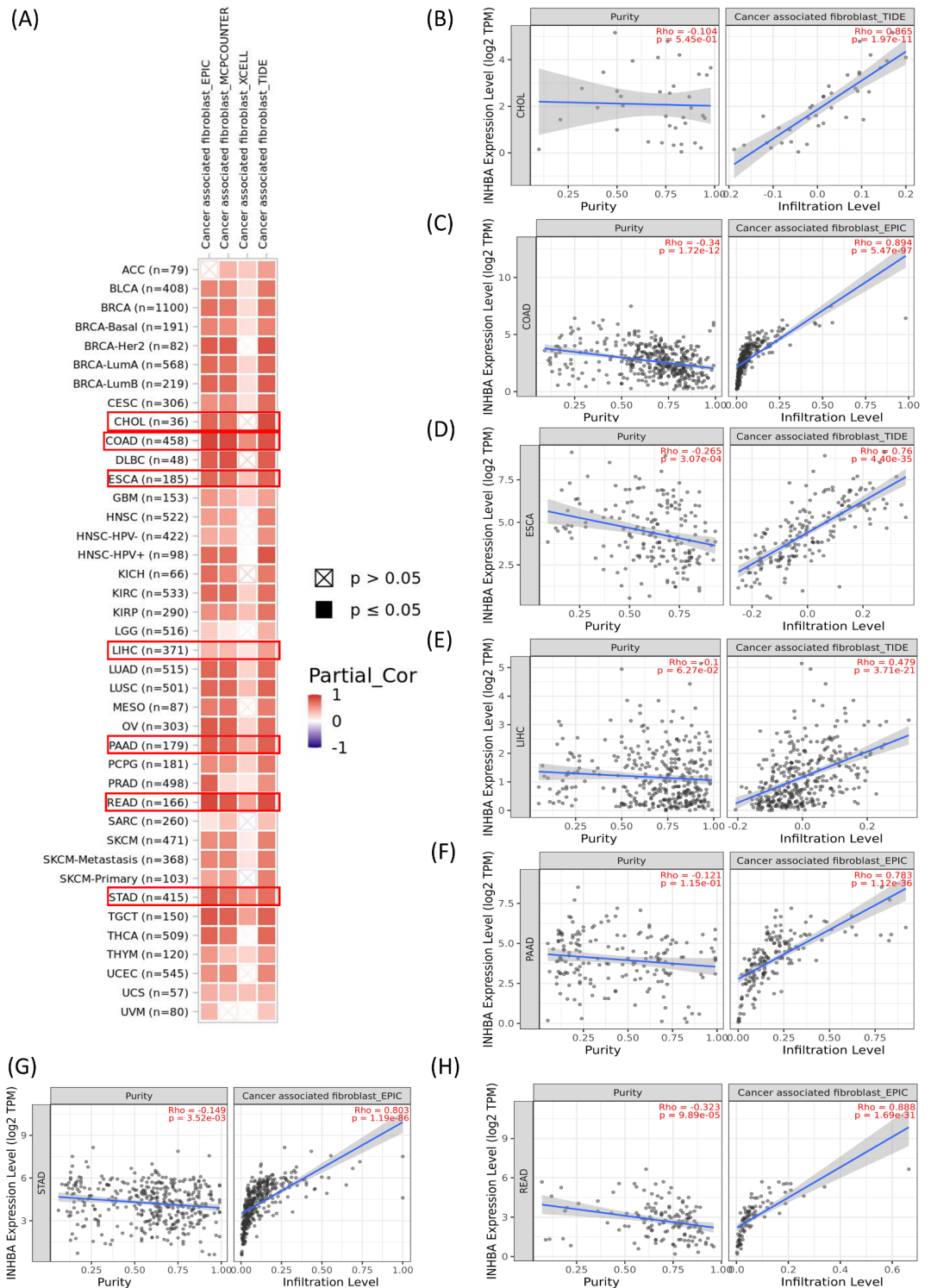


Fig. 8. Correlation of INHBA and immune infiltration of Cancer-associated fibroblast in GI tract cancers (A) Heatmap (B-H) Correlation with CHOL, COAD, ESCA, LIHC, PAAD, STAD and READ.

and high amplification ($p < 0.05$); in CD8+ T cells with arm-level gain ($p < 0.001$); in CD4+ T cells with arm-level deletion ($p < 0.01$), arm-level gain ($p < 0.01$), and high amplification ($p < 0.05$); in macrophages with arm-level gain ($p < 0.001$); in neutrophils with arm-level deletion ($p < 0.05$) and arm-level gain ($p < 0.001$); and in dendritic cells with arm-level gain ($p < 0.001$) as also reported by Liu et al. in gastric cancer study¹⁹. However, we have addressed a systematic analysis of somatic mutation study in all the GI tract cancers to understand the underlying mechanism of INHBA in all other GI tract cancers.

(A)

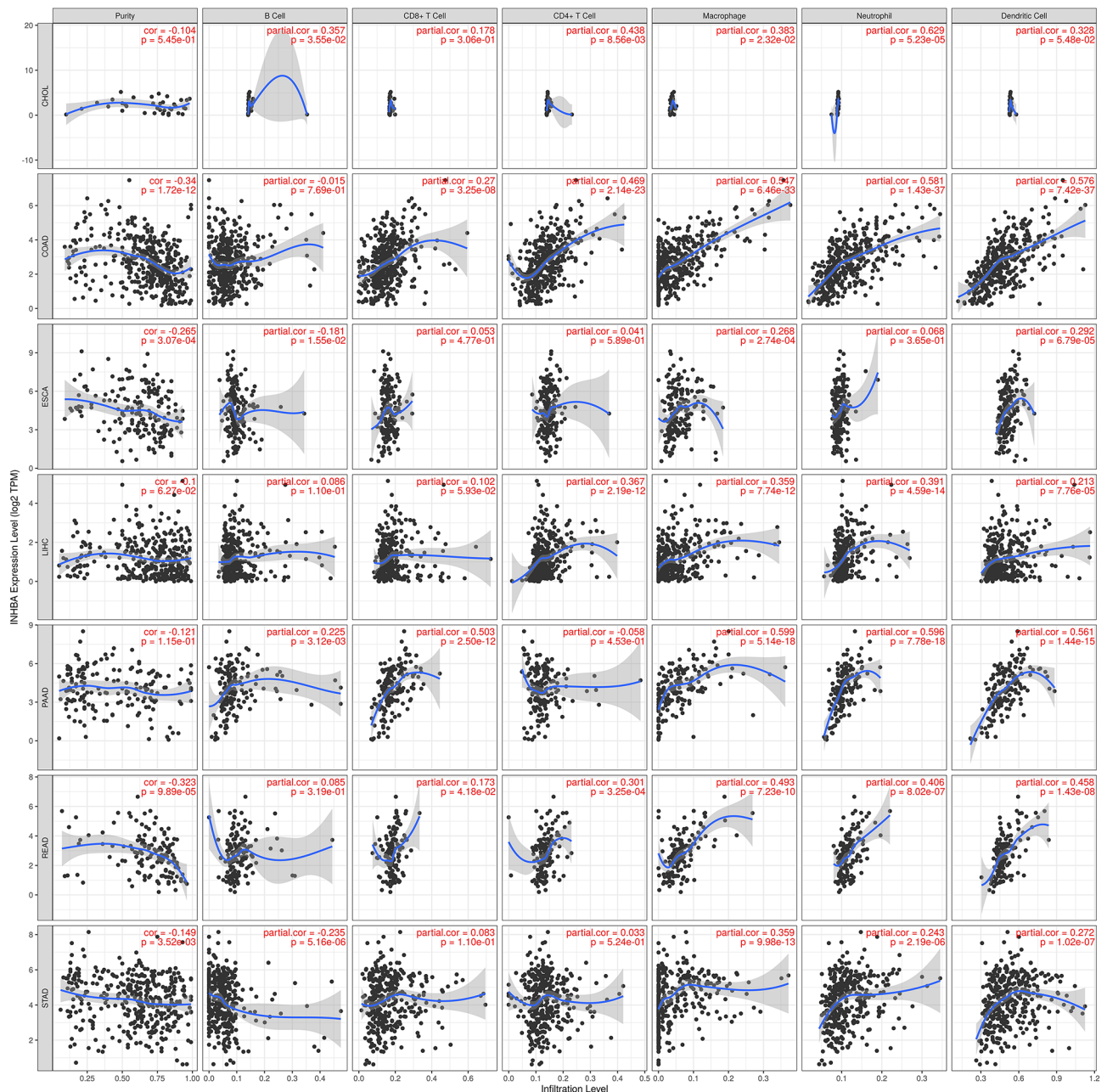


Fig. 9. Correlation analysis with immune cells (A) Correlation of INHBA and immune infiltration of B cells, T cells CD8+, CD4+, Macrophages, Neutrophils and Dendritic cells in GI tract cancers (B) Survival plot of immune infiltrates of INHBA expression.

DNA methylation analysis

Gene methylation has been extensively studied in various human cancers and therefore we conducted DNA methylation of INHBA using the SMART app tool to study the genomic information of INHBA along with the CpG probe associated with INHBA, and we observed the chromosome distribution at 1000 genomic length scale is mostly in N shore and N shelf as shown in Fig. 10(A) and (B). We also observed a significant difference in tumor methylation levels and normal levels in GI tract cancers when CpG aggregation methylation was taken for CHOL, COAD, ESCA, LIHC, and READ, but no significant difference was observed for PAAD and STAD, as shown in Fig. 10(C). **Supplementary Figure S8(A-G)** shows the INHBA promoter methylation level expression in tumor and normal TCGA samples in which the methylation level of the tumor is upregulated in CHOL, whereas it is downregulated in other GI-tract cancers. The significant difference in p-value is shown in red color ($p < 0.05$). Using “Methylation DIY,” we analysed the difference between methylation in tumors and normal for each probe ID for each GI tract cancer with methylation by pathological stages based on the clinical data by

(B)

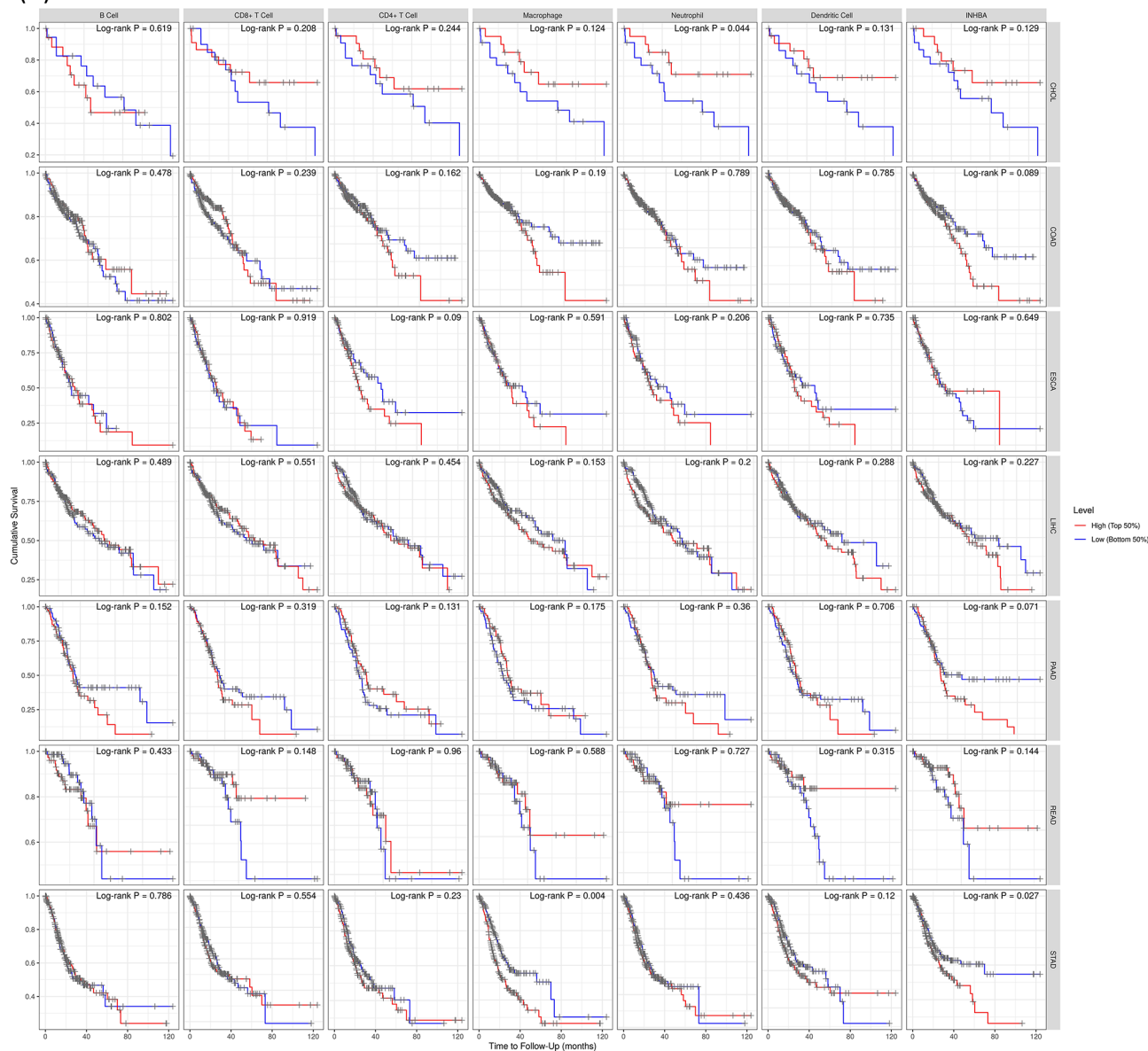


Figure 9. (continued)

TCGA in **Supplementary Figure S9**. Among the 18 CpG probes, cg18413237 was found to be downregulated in tumors and showed a statistically significant difference ($p < 0.05$) in all gastrointestinal tract cancers, as shown in **Supplementary Figure S9(A)**. The above observation from the methylation study suggests that INHBA might play different roles in GI tract cancer.

Gene enrichment analysis of INHBA-associated genes

To further study the gene INHBA's role in tumorigenesis, we have used a string database to find INHBA-binding proteins. We employed the STRING database to identify 50 genes interacting with INHBA, setting the “maximum number of interactors to show” and “minimum required interaction score” to 0.4 parameters. These interactions, depicted in Fig. 11(A), were determined based on text mining, experiments, databases, co-expression, gene fusion, and co-occurrence. The analysis identified 51 interacting protein nodes, including INHBA, and 377 edges. The network exhibited an average node degree of 14.8, an average local clustering coefficient of 0.716, and significant protein-protein interaction (PPI) enrichment (p -value $< 1.0e-16$). We also investigated INHBA-interacting proteins by configuring the parameters in the STRING database. We set “active interaction sources” to include STRING experiments, textmining, databases, co-expression, neighborhood, gene Fusion, and co-occurrence. We also set the “minimum required interaction score” to 0.7 and the “maximum number of interactors to show” to 50. The analysis identified 17 interacting protein nodes and 64 edges, with an average node degree of 7.53, an average local clustering coefficient of 0.821, and significant PPI enrichment (p -value $< 1.0e-16$). Further, to find

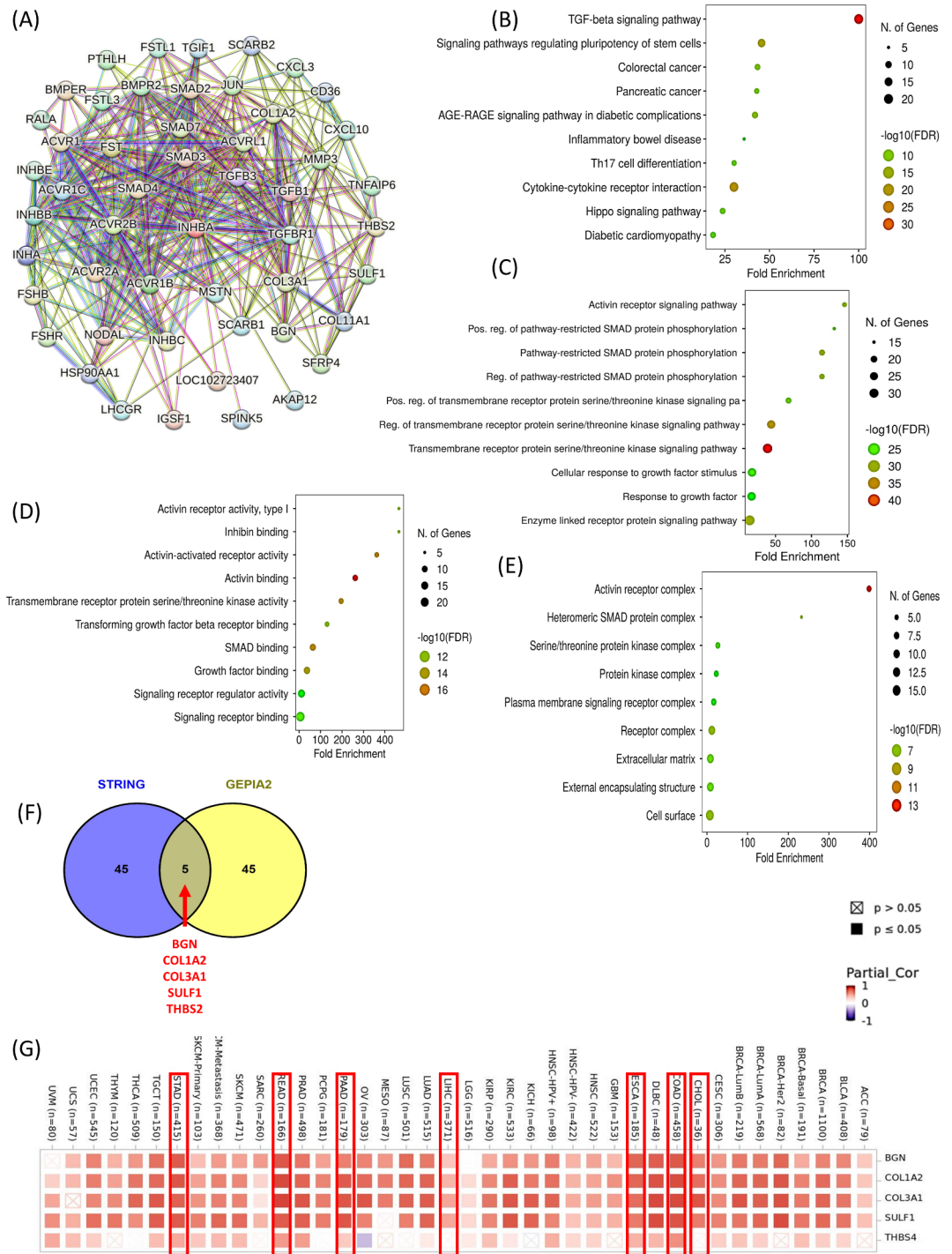


Fig. 11. INHBA-related genes enrichment analysis (A) top 50 genes associated with INHBA PPI network (B-E) KEGG pathway analysis, BP, CC, MF of top 50 INHBA associated proteins using ShinyGo (version 0.77) (<http://bioinformatics.sdstate.edu/go77/>) (F) Common genes from string database and GEPIA2 (G) Heatmap of correlation between INHBA and the shared genes.

mechanisms. This pathway is essential for cell growth, differentiation, and programmed cell death. The observed positive regulation and phosphorylation events indicate active signaling, which can influence tumor progression. Signaling through transmembrane receptor protein serine/threonine kinases, which carry signals from outside the cell to the inside, affects cellular responses like proliferation and survival. When this signaling is disrupted, it can lead to tumor growth and metastasis. The involvement of the extracellular matrix and cell surface receptor complexes suggests that INHBA expression may interact with the tumor microenvironment and cell surface receptors, impacting cell communication and growth. Functions such as Activin receptor activity, SMAD binding, and growth factor binding highlight INHBA's role in interacting with key signaling molecules and

receptors, as shown in Fig. 11(C-E). Furthermore, based on the expression similarity of genes, we screened 50 genes from GEPIA2 from all the tumor expression data related to INHBA, and after overlapping the 51 screened genes from STRING database, we identified five common genes based on expression and binding with INHBA as shown in Fig. 11(F) and found all five genes were positively correlated with INHBA in CHOL, COADREAD (Colorectal), LIHC, PAAD, ESCA and STAD as shown in Fig. 11(G). The five overlapping genes, namely Biglycan (BGN)^{25–28}, Sulfatase 1 (SULF1)^{29–31}, Thrombospondin 2 (THBS2)^{32,33}, Collagen Type I Alpha 2 Chain (COL1A2)^{34,35} and Collagen Type III Alpha 1 Chain (COL3A1)^{36–38} have been previously studied in various GI-tract cancers.

We utilized LinkedOmics to investigate gene correlations in CHOL, COADREAD, LIHC, ESCA, PAAD, and STAD cohorts. Employing the Pearson correlation method via the “LinkFinder” module, we analyzed associations between INHBA and differentially expressed genes. Heat maps in **Supplementary Figure S11** illustrate the top 50 genes positively and negatively correlated with INHBA. Additionally, using the “LinkInterpreter” module, we conducted Gene Set Enrichment Analysis (GSEA) for KEGG pathways, biological processes, cellular components, and molecular functions, detailed in **Supplementary Figure S11**. In the CHOL cohort, the top positively and negatively correlated genes were COL8A1 and BLOC1S1. While in COADREAD, LIHC, ESCA, PAAD, and STAD, the top positively correlated genes COL5A2, KIAA0754, TNFAIP6, ADAMTS12, COL10A1, and negatively correlated genes are C6orf136, ATP5G2, TJP3, TRMU, and ALDH9A1 respectively.

When we compared the GSEA among all the GI-tract cancers, these were found to be mainly the common biological processes involved, such as extracellular structure organization (in all GI-tract cancers), muscle cell migration (CHOL, COADREAD, PAAD, STAD, and LIHC), collagen metabolic process (ESCA and STAD). The negatively enriched pathways were mitochondrial and its related pathways. In the KEGG pathway analysis, positively enriched pathways included ECM-receptor interaction and focal adhesion, while negatively enriched pathways mainly fatty acid degradation, which was found to be commonly regulated. For cellular components, these genes were enriched in collagen trimers and extracellular matrix and negatively enriched NADH dehydrogenase complex. In terms of molecular function, they were mainly involved in extracellular matrix structural constituents, extracellular matrix binding, collagen binding, and so on. The GO and KEGG pathways were shown individually for each GI-tract cancer in **Supplementary Figure S11**.

Discussion

Cancers of the gastrointestinal tract have a significant risk of recurrence and have a poor prognosis. Despite different treatments, gastrointestinal tract cancer remains a leading cause of death worldwide³⁹. Therefore, early detection can be monitored using accurate biomarkers in disease progression. INHBA has been previously investigated across various cancers. However, no systematic study has yet been made on the role of INHBA in cancers related to the GI tract. In this study, we did bioinformatics analysis to find the role and mechanism of INHBA in GI tract cancers that are available from TCGA available data. Several biomarkers are available for this particular cancer, but it would be promising to identify a biomarker that could be targeted for GI-associated cancers and their therapeutic targets.

INHBA, a member of TGF Beta superfamily, encodes subunit of activin and inhibin, which plays a role in cancer progression⁵. In our study, we have seen that there is significant upregulation of INHBA in many cancers, and previous studies have been reported, including gastric cancer⁷, esophageal squamous cell carcinoma⁴⁰, colon adenocarcinoma⁴¹, breast cancer¹², which matched our findings. In our study, we focused mainly on GI tract cancers and the expression of INHBA in these cancers, and we observed the overexpression of INHBA in COAD, ESCA, PAAD, READ, and STAD. In one previous study, INHBA was found to act as a senescence-related regulator, prognostic predictor in CRC, and a mediator of immune evasion⁴². In pancreatic cancer, the knockdown of two genes, INHBA and MMP14, inhibited the invasion of cancer, and it is reported that the overexpression of INHBA has a role in cancers¹³. In vitro studies have shown that INHBA plays a crucial role in enhancing breast cancer cell proliferation as well as invasion and INHBA induced epithelial-mesenchymal transition (EMT), thereby hastening cell motility by activating TGF- β -regulated genes. These findings showed that INHBA might be employed as a diagnostic marker in GI tract cancers. A study on colon adenocarcinoma examined the role of INHBA in predicting the prognosis of patients with COAD. A model was developed using a nomogram to predict survival and recurrence in COAD patients by combining genetic information with clinical data. Using data from TCGA, differentially expressed genes (DEGs) were identified, and created a new model that incorporates INHBA expression alongside clinical factors. This model was compared to a TNM nomogram based on the pathologic TNM stage, and their results suggest that the combined model outperforms the TNM model in predicting survival and recurrence, highlighting the clinical utility of integrating genetic and clinical data⁴³. In our study, we found that higher expression of INHBA correlates with poorer overall survival (OS) in COAD. In another study on colorectal cancer (CRC), two genes, CTHRC1 and INHBA, were identified as key drivers of disease progression. These genes are highly expressed in cancer-associated fibroblasts (CAFs) within the tumor microenvironment. Single-cell sequencing analysis revealed that CTHRC1 and INHBA are significantly upregulated in fibroblasts within CRC tumor tissues compared to normal tissues. The study also found that fibroblasts are significantly more abundant in CRC tissues than in normal tissues, and a high proportion of fibroblasts is associated with poor prognosis in CRC patients. Consequently, CTHRC1, INHBA, and CAFs are identified as crucial contributors to CRC progression and could serve as promising therapeutic targets⁴⁴. In our study, we observed a positive correlation between INHBA expression and CAFs in all GI tract cancers.

Here in our study, the prognostic analysis in the GI-tract tumors investigation was visualized using the TCGA dataset “Survival heatmap” and the “Kaplan-Meier plotter.” Using GEPIA2. The study utilized two primary survival analysis criteria: (1) overall survival (OS) and (2) disease-free survival (DFS). The Kaplan-Meier plots for OS in COAD and STAD and for DFS in COAD and PAAD indicated a correlation between

INHBA expression and poorer prognosis. INHBA showed higher expression in several gastrointestinal cancers—COAD, ESCA, PAAD, READ, and STAD—but not in CHOL or LIHC. High INHBA expression correlated with adverse outcomes, specifically in COAD, PAAD, and STAD, while its role varied across different GI cancer types, suggesting a complex involvement in cancer progression. We observed the R308C mutation occurring frequently in COAD, ESCA, and LIHC using cBioPortal. We also found R377C/H mutation at the TGF- β domain for Mucinous adenocarcinoma of the colon and rectum. Further detailed investigations are required to assess the impact of the INHBA mutation using molecular dynamics analysis, providing insights into its atomic-level implications. While analyzing genetic mutation, using cBioPortal, the KM-plot for the TCGA GI-tract cancers, we didn't find any significant statistical differences in the disease-specific interval, disease-specific survival, overall survival, and progression-free survival of genetic alteration of INHBA and the clinical survival prognosis of patients. Meanwhile, genetic alterations of INHBA in esophageal adenocarcinoma are associated with a poor prognosis in overall survival.

The tumor microenvironment (TME) consists of various cellular and non-cellular components, including non-cancerous cells, nerves, lymphoid tissues, blood vessels, and metabolites, all located within or near the tumor site, whether at its core, edges, or adjacent areas. It is an environment that has an impact on all aspects of tumor biology⁴⁵. TME relies heavily on fibroblasts and immune cells. Our findings show that INHBA expression in GI tract cancer is positively correlated with cancer-associated fibroblasts, which suggests that using cancer-associated fibroblasts might be used as a potential approach that can be employed for immune treatment for GI tract-associated cancers.

The DNA methylation study of INHBA CpG aggregate methylation showed significant differences when comparing tumor and normal samples in the different GI tract cancers, which further shed light on the novel approach that can be used to study the different CpG probe and their correlation with the INHBA expression. The KEGG pathways associated with INHBA proteins are mainly enriched in the “TGF- β signalling pathway”, “Signalling pathways regulating pluripotency of stem cells”, “Colorectal cancer”, “Pancreatic cancer” while GO enrichment analysis revealed INHBA mainly functions as “Activin receptor signalling pathway”, “SMAD associated pathways”, “SMAD protein complex”, “Activin receptor activity” and “Activin receptor complex”. In addition, we found Biglycan (BGN), Sulfatase 1 (SULF1), Collagen Type III Alpha 1 Chain (COL3A1), Thrombospondin 2 (THBS2), and Collagen Type I Alpha 2 Chain (COL1A2) were positively correlated with INHBA which suggests that relationship between these genes might regulate pathways that are regulated in GI-tract cancer progression.

Conclusion

We conducted and did a comprehensive bioinformatics analysis of INHBA in GI-tract cancers (CHOL, COAD, READ, ESCA, PAAD, LIHC, and STAD) and explored to find a relationship between INHBA expression level, disease prognosis, mutations (gene alteration), immune cell infiltration, genes integrative analysis, INHBA -DNA methylation, and associated pathways in this study. INHBA is overexpressed in COAD, ESCA, PAAD, READ, and STAD and is found to be significantly associated with poor overall survival in COAD and STAD. The results also suggest the role of INHBA in immune infiltration for GI tract cancers. The INHBA also significantly differs in DNA methylation in all GI tract cancers. Therefore, INHBA might be used as a promising biomarker for GI tract cancers. Using molecular dynamic study of mutation sites might provide insight into the TGF- β domain and the structural impact and stability of INHBA. The above findings indicate that INHBA might play different biological functions in different GI tract cancers, but it still needs clinical validation from more cancer patients.

Finally, further research is needed for INHBA, which might yield a more clinically significant and potentially therapeutic gene target for GI tract cancers as our study is based on publicly available data and bioinformatics analysis.

Materials and methods

For the current study, we have used different bioinformatics approaches, online web servers, and datasets, including GEPIA2 (<http://gepia2.cancer-pku.cn/#analysis>), a web server that collects data from the TCGA data portal, and the GTEx normal tissue database. TIMER2 (<http://timer.cistrome.org/>) (a web resource for methodical analysis of gene INHBA through immune infiltrates across a wide range of cancer types), SMART (<http://www.bioinfo-zs.com/smartapp/>) (an interactive web tool for analysis of INHBA in DNA methylation analysis and visualisation), UALCAN (<https://ualcan.path.uab.edu/>) (a web server used for the stage analysis of INHBA in various GI tract cancers), cBioPortal (<https://www.cbioportal.org/>) (an online web server for Exploring INHBA in different GI tract cancers data and its mutation), ShinyGO (version 0.77) (<http://bioinformatics.sdstate.edu/go77/>) (gene ontology and pathway analysis) and (<https://xena.ucsc.edu/>).

Gene expression data

The gene expression of INHBA was studied across all TCGA GI-tract tumors using the GEPIA2 tool⁴⁶ with a log₂FC threshold of 1 and a significance level set at 0.01 for the p-value. We have used GEPIA2 to examine the differential expression pattern of INHBA in different cancer samples over TCGA tissue normal and GTEx for 31 (24 other cancer types and seven gastrointestinal cancers) and data for pan-cancer was utilized using UCSC XENA⁴⁷.

We used UALCAN⁴⁸ for pathological stage analysis to examine INHBA expression in different stages of GI tract cancer. All the box plots in the figure were created using log₂(TPM + 1) converted expression data. Using the UALCAN, we also explored the changes in protein expression using the module “clinical proteomic tumor analysis consortium” (CPTAC) to investigate changes in INHBA for the total protein levels across

different GI-tract cancers and normal tissues. We only found a dataset for this investigation for the expression of PAAD among all GI tract tumors.

Survival analysis

Here, we have used the GEPIA2 online tool to get the survival heat map. Further, we also examined the survival analysis⁴⁹ for both (1) overall survival (OS) and (2) disease-free survival (DFS) for INHBA expression in various samples present only for GI tract-related cancer. Here, as a threshold, we have used months instead of days as the survival, higher cutoff (50%) values for higher expression and lower cutoff (50%) values of lower expression samples, and p -value < 0.05 as the level of significance for the Kaplan-Meier⁵⁰ plot. Then, using the GEPIA2 survival analysis module, we reported cancers with statistically significant variability in patients' survival using a heatmap and obtained the corresponding survival plots.

Gene mutation analysis

We utilised the cBioPortal tool⁵¹ to study the prevalence, type, and location of INHBA mutations in all GI tract tumors. We also accessed the survival analysis of INHBA genetic mutations. The INHBA mutation site information is presented in the 'Mutations' segment. Then, using the survival/comparison analysis module, we analysed the (1) overall survival (OS), (2) progression-free survival (PFS), (3) disease-specific survival (DSS), and (4) disease-free survival (DFS) data from TCGA cancer patients for INHBA gene changes and got the relevant KM plots with log-rank test p -value for cumulative GI tract cancers. The structure of INHBA was generated using a 3D protein imaging online tool⁵².

Analysis of the correlation between INHBA expression and immune cell infiltration

For immune infiltration analysis, we have used TIMER⁵³ and TIMER2 websites⁵⁴ and accessed immune cells to find the relationship between INHBA and the expression of immune infiltration across all GI tract cancers. For further investigation, we chose (1) B-cells, (2) macrophages, (3) natural killer cells, (4) neutrophils, (5) CD8 + T cells, and (6) CD4 + T cells from TIMER, and we also accessed cancer-associated fibroblasts (CAF) from TIMER2. The p -values and partial correlation values of 4 algorithms, namely EPIC, MCPCOUNTER, TIDE, and XCELL, were accessed using Spearman's rank (purity-adjusted) correlation analysis from TIMER2 for CAF analysis. After the data was visualized, a correlation plot, survival plot, and mutation plot were generated using TIMER for different GI-tract cancers.

INHBA-related gene enrichment analysis

The online STRING tool⁵⁵ was used to analyse the molecular interaction network for INHBA and related genes. We then defined the network for a minimum needed interaction score with low confidence of 0.400 and a maximum number of interactors to show not more than 50 interactors. Finally, we are able to obtain the top 50 INHBA interacting proteins as well as the accompanying network diagrams. We then utilised ShinyGO (version 0.77)^{56–61} to do an enrichment analysis of INHBA and associated genes. In this study, we have provided insights into gene ontology (biological processes (BP), cellular components (CC), molecular functions (MF), and KEGG pathways). Further, GEPIA2 was accessed to find only the top 50 genes that are similar to INHBA, and Venny2.0 was accessed to find the shared genes between STRING and GEPIA2. Finally, we used the TIMER2 gene correlation module to acquire heatmap data from 5 common genes in different cancer types. Finally, the degree of association was quantified by presenting the Spearman rho value (adjusted for purity).

LinkedOmics analysis

The LinkedOmics⁶² was accessed, which is a unique online platform for analyzing and evaluating cancer data across all cancer projects. We utilised LinkedOmics to identify positively and negatively correlated significant genes analytical modules for each GI tract cancer. A functional enrichment analysis and integrated association results were compared using Gene Set Enrichment Analysis (GSEA)⁶³. In this investigation, we also used Pearson correlation and a substantial p -value to highlight the top positively and negatively linked genes in distinct GI tract tumors.

SMART app

We employed the SMART (Shiny Methylation Analysis Resource Tool), which is a web App/Tool⁶⁴, to ease the DNA methylation analysis study for INHBA. It is a user-accessible web tool designed to study DNA methylation data from various cancer projects within TCGA. The SMART app was used to investigate the methylation probes associated with INHBA. We analysed the different probe IDs associated with INHBA in normal and tumor tissue from several GI-tract malignancies. Further methylation among distinct clinical stages of cancer was also investigated.

Data availability

In this study, we primarily utilized datasets from the publicly accessible TCGA projects (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>) and various online bioinformatics tools detailed in the materials and methods. The datasets analyzed, accessible through the Gene Expression Omnibus (GEO) repository, include accession numbers GSE32879, GSE110224, GSE38129, GSE71989, GSE29721, and GSE54129 (<https://www.ncbi.nlm.nih.gov/gds>).

Received: 2 May 2024; Accepted: 10 September 2024

Published online: 07 January 2025

References

- Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263 (2024).
- Tabibzadeh, A. et al. Signal transduction pathway mutations in gastrointestinal (GI) cancers: A systematic review and meta-analysis. *Sci. Rep.* **10**, 18713 (2020).
- Peng, C., Ouyang, Y., Lu, N. & Li, N. The NF- κ B signaling pathway, the microbiota, and gastrointestinal tumorigenesis: Recent advances. *Front. Immunol.* **11**, 1387 (2020).
- Macarthur, M., Hold, G. L. & El-Omar, E. M. Inflammation and Cancer II. Role of chronic inflammation and cytokine gene polymorphisms in the pathogenesis of gastrointestinal malignancy. *Am. J. Physiol. Gastrointest. Liver Physiol.* **286**, G515–G520 (2004).
- Chen, Z. L., Qin, L., Peng, X. B., Hu, Y. & Liu, B. INHBA gene silencing inhibits gastric cancer cell migration and invasion by impeding activation of the TGF- β signaling pathway. *J. Cell. Physiol.* **234**, 18065–18074 (2019).
- Gold, L. I. The role for transforming growth factor-beta (TGF-beta) in human cancer. *Crit. Rev. Oncog.* **10**, 303–360 (1999).
- Wang, Q. et al. Upregulated INHBA expression is associated with poor survival in gastric cancer. *Med. Oncol.* **29**, 77–83 (2012).
- Oshima, T. et al. Relation of INHBA gene expression to outcomes in gastric cancer after curative surgery. *Anticancer Res.* **34**, 2303–2309 (2014).
- Zhao, K., Yi, Y., Ma, Z. & Zhang, W. INHBA is a prognostic biomarker and correlated with immune cell infiltration in cervical cancer. *Front. Genet.* **12**, 705512 (2021).
- Li, X. et al. Targeting INHBA in ovarian cancer cells suppresses cancer xenograft growth by attenuating stromal fibroblast activation. *Dis. Markers* 7275289 (2019) (2019).
- Jayaraman, H., Anandhapadman, A. & Ghone, N. V. In vitro and in vivo comparative analysis of differentially expressed genes and signaling pathways in breast cancer cells on interaction with mesenchymal stem cells. *Appl. Biochem. Biotechnol.* **195**, 401–431 (2023).
- Yu, Y., Wang, W., Lu, W., Chen, W. & Shang, A. Inhibin β -A (INHBA) induces epithelial–mesenchymal transition and accelerates the motility of breast cancer cells by activating the TGF- β signaling pathway. *Bioengineered* **12**, 4681–4696 (2021).
- Liang, Z. W. et al. M2-phenotype tumour-associated macrophages upregulate the expression of prognostic predictors MMP14 and INHBA in pancreatic cancer. *J. Cell. Mol. Med.* **26**, 1540–1555 (2022).
- Seder, C. W. et al. Upregulated INHBA expression may promote cell proliferation and is associated with poor survival in lung adenocarcinoma. *Neoplasia* **11**, 388–396 (2009).
- Zhang, S., Jin, K., Li, T., Zhou, M. & Yang, W. Comprehensive analysis of INHBA: A biomarker for anti-TGF β treatment in head and neck cancer. *Exp. Biol. Med.* **247**, 1317–1329 (2022).
- Okano, M. et al. Significance of INHBA expression in human colorectal cancer. *Oncol. Rep.* **30**, 2903–2908 (2013).
- Massagué, J. TGF β in cancer. *Cell* **134**, 215–230 (2008).
- Staudacher, J. J. et al. Prognostic impact of activin subunit inhibin beta A in gastric and esophageal adenocarcinomas. *BMC Cancer* **22**, 953 (2022).
- Vizeacoumar, F. S. et al. Mining the plasma-proteome associated genes in patients with gastro-esophageal cancers for biomarker discovery. *Sci. Rep.* **11**, 7590 (2021).
- Qiu, S. et al. CircTHBS1 drives gastric cancer progression by increasing INHBA mRNA expression and stability in a ceRNA- and RBP-dependent manner. *Cell. Death Dis.* **13**, 266 (2022).
- Wang, J. et al. Upregulation of INHBA mediated by the transcription factor BHLHE40 promotes colon cancer cell proliferation and migration. *J. Clin. Lab. Anal.* **36**, e24539 (2022).
- Liu, F. et al. Identification of INHBA as a potential biomarker for gastric cancer through a comprehensive analysis. *Sci. Rep.* **13**, 12494 (2023).
- Zhang, X. et al. The new role of HNF1A-NAS1/miR-214/INHBA signaling axis in colorectal cancer. *Front. Biosci.* **28**, 301 (2023).
- Abutalebi, M. et al. Discovery of PELATON links to the INHBA gene in the TGF- β pathway in colorectal cancer using a combination of bioinformatics and experimental investigations. *Int. J. Biol. Macromol.* **270**(Pt 1) (2024).
- Gu, X. et al. Up-regulated biglycan expression correlates with the malignancy in human colorectal cancers. *Clin. Exp. Med.* **12**, 195–199 (2012).
- Weber, C. K. et al. Biglycan is overexpressed in pancreatic cancer and induces G1-arrest in pancreatic cancer cell lines. *Gastroenterology* **121**, 657–667 (2001).
- Pinto, F., Santos-Ferreira, L., Pinto, M. T., Gomes, C. & Reis, C. A. The extracellular small leucine-rich proteoglycan biglycan is a key player in gastric cancer aggressiveness. *Cancers* **13**, (2021).
- Zhao, S. F., Yin, X. J., Zhao, W. J., Liu, L. C. & Wang, Z.-P. Biglycan as a potential diagnostic and prognostic biomarker in multiple human cancers. *Oncol. Lett.* **19**, 1673–1682 (2020).
- Matusiewicz, M. et al. Serum sulfatase activity is more elevated in colonic adenomas than cancers. *Int. J. Colorectal Dis.* **23**, 383–387 (2008).
- Lyu, Y., Cheng, Y., Wang, B., Chen, L. & Zhao, S. Sulfatase 1 expression in pancreatic cancer and its correlation with clinicopathological features and postoperative prognosis. *Cancer Biomark.* **22**, 701–707 (2018).
- Lai, J. P. et al. hSulf1 sulfatase promotes apoptosis of hepatocellular cancer cells by decreasing heparin-binding growth factor signaling. *Gastroenterology* **126**, 231–248 (2004).
- Shi, H. et al. Bone marrow-derived mesenchymal stem cells promote Helicobacter pylori-associated gastric cancer progression by secreting thrombospondin-2. *Cell. Prolif.* **54**, e13114 (2021).
- Tokunaga, T. et al. Thrombospondin 2 expression is correlated with inhibition of angiogenesis and metastasis of colon cancer. *Br. J. Cancer* **79**, 354–359 (1999).
- Li, J., Ding, Y. & Li, A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J. Surg. Oncol.* **14**, 297 (2016).
- Sadler, T., Scarpa, M., Rieder, F., West, G. & Stylianou, E. Cytokine-induced chromatin modifications of the type I collagen alpha 2 gene during intestinal endothelial-to-mesenchymal transition. *Inflamm. Bowel Dis.* **19**, 1354–1364 (2013).
- Zhang, H. et al. Data mining-based study of collagen type III alpha 1 (COL3A1) prognostic value and immune exploration in pancreatic cancer. *Bioengineered* **12**, 3634–3646 (2021).
- Zhang, S. W., Zhang, N. & Wang, N. Role of COL3A1 and POSTN on pathologic stages of esophageal cancer. *Technol. Cancer Res. Treat.* **19**, 1533033820977489 (2020).
- Shen, H. et al. The prognostic value of COL3A1/FBN1/COL5A2/SPARC-mir-29a-3p-H19 associated ceRNA network in gastric cancer through bioinformatic exploration. *J. Cancer* **11**, 4933–4946 (2020).
- Fernández-Aceñero, M. J., Hernández, D. & del Arco, C. D. A review and update on therapy of gastrointestinal tract tumors: From the bench to clinical practice. *J. Clin. Transl. Pathol.* (2024).
- Lyu, S., Jiang, C., Xu, R., Huang, Y. & Yan, S. INHBA upregulation correlates with poorer prognosis in patients with esophageal squamous cell carcinoma. *Cancer Manag Res.* **10**, 1585–1596 (2018).
- Xiao, Q. et al. Metformin suppresses the growth of colorectal cancer by targeting INHBA to inhibit TGF- β /PI3K/AKT signaling transduction. *Cell. Death Dis.* **13**, 202 (2022).
- Chen, S. et al. INHBA is a novel mediator regulating cellular senescence and immune evasion in colorectal cancer. *J. Cancer* **12**, 5938–5949 (2021).

43. Li, X. et al. INHBA is a prognostic predictor for patients with colon adenocarcinoma. *BMC Cancer* **20**, 305 (2020).
44. Hu, S. et al. Integrated analysis of single cell and bulk RNA sequencing identifies CTHRC1 + INHBA + CAF as drivers of colorectal cancer progression. *Mol. Carcinog.* **62**, 1787–1802 (2023).
45. Jin, M. Z. & Jin, W. L. The updated landscape of tumor microenvironment and drug repurposing. *Signal. Transduct. Target. Ther.* **5**, 166 (2020).
46. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560 (2019).
47. Goldman, M. J. et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).
48. Chandrashekar, D. S. et al. An update to the integrated cancer data analysis platform. *Neoplasia* **25**, 18–27 (2022).
49. George, B., Seals, S. & Aban, I. Survival analysis and regression models. *J. Nucl. Cardiol.* **21**, 686–694 (2014).
50. Rich, J. T. et al. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol. Head Neck Surg.* **143**, 331–336 (2010).
51. Cerami, E. et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
52. Tomasello, G., Armenia, I. & Molla, G. The protein imager: A full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics* **36**, 2909–2911 (2020).
53. Li, T. et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* **77**, e108–e110 (2017).
54. Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514 (2020).
55. Szklarczyk, D. et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
56. Ge, S. X., Jung, D. & Yao, R. ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).
57. Luo, W. & Brouwer, C. Pathview: An R/bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
58. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
59. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
60. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
61. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
62. Vasaikar, S. V., Straub, P., Wang, J. & Zhang, B. LinkedOmics: Analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* **46**, D956–D963 (2018).
63. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: A desktop application for gene set enrichment analysis. *Bioinformatics* **23**, 3251–3253 (2007).
64. Li, Y., Ge, D. & Lu, C. The SMART app: An interactive web application for comprehensive DNA methylation analysis and visualization. *Epigenetics Chromatin* **12**, 71 (2019).

Acknowledgements

The authors extend their gratitude to the Shiv Nadar Institution of Eminence (SNiOE) for furnishing the essential research framework and infrastructure vital to the successful execution of this study.

Author contributions

AS, PKS, and RV have designed and conceptualized the study. RV performed the analysis and wrote the manuscript. AS and PKS approved further analysis and edited and reviewed the manuscript. All authors approved the manuscript. All authors contributed to the research article and approved for the submitted version.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-72679-3>.

Correspondence and requests for materials should be addressed to P.K.S. or A.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025