

# CpG Island Methylator Phenotype With Distinctly Different Prognosis And Molecular Features In Pancreatic Cancer Patient

Gang Ning (✉ [eyninggang@scut.edu.cn](mailto:eyninggang@scut.edu.cn))

Guangzhou First People's Hospital <https://orcid.org/0000-0002-7691-2837>

Yongqiang Li

Guangzhou First People's Hospital

Wenji Chen

Guangzhou First People's Hospital

Wenjuan Tang

Guangzhou First People's Hospital

Diwen Shou

Guangzhou First People's Hospital

Qingling Luo

Guangzhou First People's Hospital

Huiting Chen

Guangzhou First People's Hospital

Yongjian Zhou

Guangzhou First People's Hospital

---

## Research article

**Keywords:** pancreatic cancer, CpG island methylator phenotype, somatic mutation, tumor microenvironment, cancer stemness indices.

**Posted Date:** January 25th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-153446/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

## Background

CpG island methylator phenotype (CIMP), featured with concurrent and widespread hypermethylation of a cluster of CpGs, has been reported to play an important role in carcinogenesis. Limited studies have investigated the CIMP in pancreatic cancer (PC). The aim of this study was to explore the CIMP in PC patients and its clinical-pathological and genomic characteristics.

## Methods:

DNA methylation, somatic mutation, mRNA and corresponding clinical data of PC patients were downloaded from TCGA (184 patients) and ICGC (264 patients). Univariate and multivariate regression analysis were used to identify prognosis related CpGs. Consensus clustering analysis was used for identification of the CIMP in PC patients. ESTIMATE and CIBORORT were used for estimation of tumor microenvironment (TME) in PC patients.

## Results

In TCGA PC cohort, 22,450 differential CpGs, including 12,937 hypermethylated CpGs and 9,513 hypomethylated CpGs were identified between 185 PC patients and 10 normal controls. Univariate and multivariate Cox analysis further screened out 72 OS related CpGs and three distinct CIMP groups with distinctly different prognosis and molecular features, including CIMP-L subgroup, CIMP-M subgroup and CIMP-H subgroup were identified based on unsupervised consensus clustering analysis of these CpGs. Patients of the CIMP-H subgroup had poorer OS and RFS, while patients of CIMP-L subgroup had better OS and RFS. The CIMP status were also an independent prognostic factors for OS and PFS. In molecular features, significantly higher somatic mutation burdens and tumor mutational burden were found at patients of CIMP-H subgroup compared to that in patients of CIMP-L subgroup. Besides, lower stromal score, immune score and higher cancer stemness indices and tumor purity were also found at patients of CIMP-H subgroup compared to that in patients of CIMP-L subgroup. Correspondingly, significantly total T cells, total B cells, CD8 T cells, memory CD4 T cells and higher regulatory T cells were found at patients of CIMP-H subgroup. Moreover, significantly lower expression of immune checkpoint genes, such as PD-1, CTLA4, CD86, VTCN1 and LAG-3 were also found at patients of CIMP-H subgroup compared to that in patients of CIMP-L subgroup. In the end, we validated the CIMP status in PC patients of ICGC dataset.

## Conclusion

Three distinct CIMP subgroups of PC patients with distinct clinical characteristic, prognosis, gene mutation landscape and TME were identified and validated. The CIMP may help to make an assertion to provide specific and efficient treatment options for patients of different subtypes.

## Introduction

Pancreatic cancer (PC) is the fifteenth most common cancers and the seventh leading cause of cancer-related death worldwide, which is reported to have poor prognosis and high mortality rate(1). Lots of efforts were put in understanding the underlying mechanisms of PC and new treatments were investigated, but a marginal improvement of survival was observed in PC (2). Neureiter et al. reported the median survival of PC patients to be 6 months and the 5-year survival rate to be as low as 6%, attributing the cause to be lack of reliable early diagnostic markers and the nature of high aggressiveness and drug resistance of the disease itself (3). Of note, only about 10% of PC patients were report with family histories (4), indicating the etiology of PC to be more likely sporadic than genetic and suggesting epigenetic alterations may take in part of the development and progression of PC.

DNA methylation is one of the main epigenetic modifications and plays an important role in the development and progression of sorts of malignant tumors (5). During the process of oncogenesis, aberrant DNA methylation of CpG islands are located in the promoter regions of genes, such as hypermethylation surrounding the promoters of tumor suppressor genes (TSGs) and hypomethylation of promoters of oncogenes, resulting in transcriptional silence of TSGs and over-expression of oncogenes (6). Aberrant DNA methylation of non-promoter of genes also contributes to the formation of intra-tumoral heterogeneity (7). To date, many gene mutations with deregulated DNA methylation have been identified in PC patients. Treatments targeting these genes were reported to significantly improved the overall survival in PC mice models (8, 9).

CpG island methylator phenotype (CIMP), first discovered and validated in colorectal cancer, is featured with concurrent and widespread hypermethylation of a cluster of CpGs in distinct cancer subtypes. CIMP affects the intergenic regions of the whole genome and leads to chromosomal instability, which is thought to play an important role in carcinogenesis (10). To date, CIMP phenotype has been identified in many kinds of tumors, including PC (11). Subtypes with different CIMP pattern showed distinct epidemiological, clinical-pathological and genomic characteristics (12, 13). However, the limitation of earlier studies on CIMP of PC patients was obvious for the limited patient samples or CpG sites used in these study (5, 14). Little of these studies validated their findings in independent cohorts and further explored the association of CIMP status with tumor microenvironment (TME) in PC patients (11). In the present study, we identified and validated three distinct methylation subgroups of PC patients, termed CIMP-L subgroup, CIMP-M subgroup, and CIMP-H subgroup, with the data downloaded from The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC). Patients of different CIMP status showed distinct clinical characteristic, prognosis, gene mutation landscape and TME. As identification of clinically relevant cancer subtypes based on DNA methylation pattern is an important computational problem in medicine, CIMP may help to make an assertion to provide specific and efficient treatment options for patients of different subtypes.

## Methods

### Ethics statement

All the data analyzed in the present study were got from TCGA and ICGC. Informed consents had already been obtained from the patients before the present study.

### Data acquisition from TCGA

Level-3 DNA methylation data of 184 PC patients and 10 normal controls were downloaded from the TCGA (<https://cancergenome.nih.gov/>, 2020-04-20). The methylation status of each CpG was expressed as a  $\beta$  value, which ranged from 0 to 1, where higher  $\beta$  values indicated higher methylation. We removed  $\beta$  values for those CpG probes that were either mapped against chromosomes X and Y, or missing in more than 20% of the samples. Besides, those probes mapped to SNP within 10bp of interrogated CpG sites were also removed (15).

The somatic mutation data of 177 PC patients measured by the whole exome sequencing were downloaded from TCGA and maftools package was used to analyze these data (16). Patients harboring missense mutations, nonsense mutations, multiple hits, splice-site mutations, frame shift insertions, frame shift deletions, in-frame insertions or in-frame deletions were considered as positive for mutation. Moreover, tumor mutational burden (TMB), regarded as a promising biomarker for immunotherapy responses, was also calculated with the method described in previous study (17).

mRNA expression data of 178 PC patients was also obtained from TCGA. The averaged gene expression was adopted in case of duplicates and the gene expression data was normalized with the scale method (18). Besides, gene expression missed in more than 20% of the samples were removed.

Meanwhile, corresponding clinical-pathological data of 184 PC patients, including gender, age, T stage (T), N stage (N), M stage (M), TNM stage, overall survival (OS) status and time, progression free status (PFS) status and time were also downloaded.

### Identification of the CIMP in PC patients

To evaluate the CIMP phenomenon in PC patients, first, the differential CpGs between 184 PC patients and 10 normal samples were screened out. Then, univariate and multivariate Cox regression analysis were performed to further identify the OS-related CpGs. After that, consensus clustering analysis for unsupervised class of 184 PC patients based on expression similarity of OS-related CpGs was performed with consensusclusterplus packages. Besides, principal component analysis (PCA) was also performed to examine whether the clusters of PC patients were suitable with 'limma' package.

### Estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE)

ESTIMATE was a tool used for predicting tumor purity, the presence of infiltrating stromal/immune cells in tumor tissues with mRNA expression data. Based on single sample Gene Set Enrichment Analysis (ssGSEA), ESTIMATE could generate three scores: stromal

score (that captured the presence of stroma in tumor tissue), immune score (that represented the infiltration of immune cells in tumor tissue), and the estimate score (that negatively correlated with tumor purity) (19).

### **Cancer stemness indices of PC patients**

Previously, Malta et al. extracted transcriptomic and epigenetic feature sets derived from non-transformed pluripotent stem cells and their differentiated progeny using an innovative one-class logistic regression machine learning algorithm (OCLR). As a result, they identified four cancer stemness indices, including mRNasi, epigenetically regulated mRNasi (EREG-mRNasi), mDNasi and EREG-mDNasi for assessing the degree of oncogenic dedifferentiation (20). Based on their study, we could attain the four stemness indices of each PC patients in the TCGA database.

### **CIBORORT**

22 kinds of tumor-infiltrating immune cells of each PC patients was calculated with CIBERSORT (<https://cibersort.stanford.edu>), an online tool designed for estimating the abundances of tumor-infiltrating immune cells with transcriptomic data (21).

### **Gene set enrichment analysis (GSEA)**

To explore the underlying mechanism exploited by the CIMP status to influence the prognosis of PC patients, GSEA analysis (Version: 4.2; <http://software.broadinstitute.org/gsea/index.jsp>) was performed. As a result, the difference of KEGG pathways between PC patients with distinct CIMP status were identified (22).

### **Validation of CIMP status in PC patients of ICGC dataset**

To independently test the CIMP status in PC patients, DNA methylation data of 264 PC patients, somatic mutation data of 264 PC patients, mRNA expression data of 175 PC patients and clinical-pathological parameter of 264 PC patients were downloaded from ICGC (<https://dcc.icgc.org/>, 2020-04-20). Processing of methylation data, somatic mutation data and mRNA expression data were similar to that in TCGA.

### **Data analysis flow chart**

To make our study to be better understood, a workflow of the study was depicted and was shown at **Figure 1**.

### **Statistical analysis**

GraphPad Prism 6 (GraphPad Software) and R software (version 3.5.1) were used for statistical analysis and plotting graph. The association between CIMP status with clinical-pathological features were analyzed with the chi-square test. One-way ANOVA analysis was carried out to compare difference of TMB, stromal score, immune score, estimate score, tumor-infiltrating immune cells, expression of chemokines and immune checkpoint genes among PC patients with different CIMP status. Univariate and multivariate Cox regression analysis were performed to analyze the prognostic value of CIMP status. Kaplan–Meier analysis with log-rank test was performed to analyze the difference of OS or PFS among patients with different CIMP status.  $P < 0.05$  was considered as statistically significant.

## **Results**

### **Identification of the CIMP in PC patients**

To identify the CIMP in PC patients, we first screened out the differential CpGs between PC samples and normal samples with the DNA methylation data of 185 PC patients and 10 normal controls downloaded from TCGA. In total, 22,450 differential CpGs were identified, and among these CpGs, 12,937 were hypermethylated CpGs and 9,513 were hypomethylated CpGs. Next, univariate Cox analysis was performed to identify OS-related CpGs and 3102 CpGs were selected with  $p < 0.05$ . Then, 1037 out of 3102 CpGs with  $p < 0.01$  were further used to identify the most OS-related CpGs by multivariate Cox analysis. Seventy-two CpGs were finally selected for unsupervised consensus clustering analysis with  $p < 0.05$  (**Supplementary materials**). Based on the unsupervised consensus clustering analysis, 184 PC patients were clustered into three distinct groups (**Figure 2A-2C**), namely CIMP-L subgroup ( $n=46$ ), CIMP-M subgroup ( $n=82$ ) and CIMP-H subgroup ( $n=56$ ). The methylation level of CIMP-L subgroup was low while patients of the CIMP-H subgroup had widespread hypermethylated CpGs.

Next, the associations between CIMP status and clinical characteristics were analyzed. As were shown at **Table 1**, there were more patients with advanced T stage and TNM stage in the CIMP-M subgroup and CIMP-H subgroup compared to that in CIMP-L subgroup (all  $p < 0.05$ , **Table 1**). Besides, Kaplan-Meier analysis showed that there were significant differences in OS and PFS among PC patients from different CIMP status. Patients of the CIMP-H subgroup had poorer OS and RFS while the patients of CIMP-L subgroup had better OS and RFS (all  $p < 0.05$ , **Figure 2D, 2E**). Moreover, univariate Cox analysis indicated that CIMP status were significantly related with OS and PFS, and multivariate Cox analysis also suggested that CIMP status were independent prognostic factors for OS and PFS of PC patients after adjusting for gender, age, T state, N stage, M stage and TNM stage (all  $p < 0.05$ , **Table 2**).

### **Mutational landscapes of PC patients with different CIMP status**

A number of mutated genes with deregulated DNA methylation had been identified to be played important roles in the development and progression of PC (8). The association of CIMP status with gene mutations was analyzed with somatic mutation data of 177 PC patients downloaded from TCGA. As were shown at **Figure 3**, there were significantly higher somatic mutation burdens among patients with different CIMP status. All the PC patients of the CIMP-H subgroup ( $n=54$ ) had gene mutation, and 79 out of 81 PC patients of the CIMP-M subgroup had gene mutation, while only 31 out of 41 PC patients of the CIMP-L subgroup had gene mutation. Obviously, significantly higher somatic mutation burdens in KRAS, TP53, SMAD4, CDKN2A and TTN were observed in patients of CIMP-H group (**Figure 3A-3C**), which had been shown to be major driver genes in PC (23). Similarly, significant difference of tumor mutational burden (TMB), served as a biomarker of immunotherapy responses, were also found among patients with different CIMP status. Higher TMB were found at patients of CIMP-H subgroup, while lower TMB were observed at patients with CIMP-L status (**Figure 3D**).

### **Landscape of TME in PC patients with distinct CIMP status**

Consisted of cancer cells, stromal cells and extracellular components, the TME had been demonstrated to play indispensable roles in tumorigenesis, progression, metastasis, recurrence, and drug resistance of PC (24). The difference of TME in patients with distinct CIMP status was also analyzed. As were shown at **Figure 4**, significantly lower stromal score, immune score and estimate score were found at patients of CIMP-H subgroup, while significantly higher stromal score, immune score and estimate score were found at patients of CIMP-L subgroup (**Figure 4A-4C**). Similarly, significantly higher tumor purity score was also observed in patients of CIMP-H subgroup (**Figure 4D**).

Cancer stem cells (CSCs) were cancer cells that possessed the ability to give rise to all tumor cell types and CSCs were considered to be responsible for tumor growth, metastasis and recurrence, and resistant to chemotherapy and radiation therapy. The association of CIMP status with cancer stemness indices was explored. As expected, significantly higher tumor stemness indices, including mRNasi score, mDNasi score EREG- mDNasi score were found at patients of CIMP-H subgroup compared to that in patients of CIMP-L subgroup (**Figure 5A, 5C, 5D**).

Next, we analyzed the difference of tumor-infiltrating immune cells among patients with distinct CIMP status. As were shown at **Figure 6**, significantly lower total T cells, total B cells, naive B cells, CD8 T cells, CD4 T cells, resting memory CD4 T cells and activated memory CD4 T cells were found in patients of CIMP-H subgroup compared to that in patients of CIMP-L subgroup (**Figure 6A-6F**). Besides, significantly higher M0 macrophages was found in patients of CIMP-H subgroup and significantly higher regulatory T cells was found at patients of CIMP-M subgroup compared to that in patients of CIMP-L subgroup (**Figure 6G-6H**). It had been reported that different kinds of immune cell subsets were recruited into the TME via interactions between chemokines and their chemokine receptors (25). We furthered analyzed the difference of expression of 58 kinds of chemokines among patients with different CIMP status. As were summarized at **Table 3**, in line with the results of tumor-infiltrating immune cells, 31 kinds of chemokines, such as CCL2, XCL2, CCR2, CCL5 and CCR5, were found to be overexpressed in patients of CIMP-L subgroup, while only 7 kinds of chemokines, such as CXCL14 and CXCL16, were found to be increased in patients of CIMP-M and CIMP-H subgroup. Taken together, these results suggested that patients in CIMP-H subgroup have a distinct TME, characterized by higher tumor purity and tumor stemness, and lower immune activation and immune infiltration.

### **Expression of immune checkpoint genes in PC patients with different CIMP status**

The advent of immunotherapy, especially checkpoint inhibitor-based immunotherapy, had revolutionized cancer treatments, especially for patients with advanced tumor. These treatments functioned through blockade of immunosuppressive checkpoints, so the expression of these immune checkpoint genes were necessary for checkpoint inhibitor immunotherapy (26). The difference of expression in 10 immune checkpoint genes (including PD-1, PD-L1, CTLA4, PD-L2, CD86, CD80, CD276, VTCN1, Tim-3 and LAG-3) in PC

patients with different CIMP status were further analyzed. As were shown at **Figure 7**, the expression of PD-1, CTLA4, CD86, VTCN1 and LAG-3 of PC patients of CIMP-H subgroup were significantly lower than that of patients of CIMP-L subgroup (**Figure 7A, 7C, 7E, 7H-J**). These results may indicate that checkpoint inhibitor immunotherapy to be less effective in patients of CIMP-H subgroup as they showed less expression of immune checkpoint genes.

### Potential mechanism by which CIMP status to influence the prognosis of PC patients

GSEA analysis was performed to explore the underlying biological mechanism by which CIMP influenced prognosis of PC patients. As was shown at **Figure 8**, KEGG pathways, such as “P53 signaling pathway”, “notch signaling pathway”, “calcium signaling pathway”, “DNA replication” and “base excision repair” were found to be significantly enriched in patients of CIMP-H subgroup compared that in patients of CIMP-L patients (**Figure 8A**). Similarly, “P53 signaling pathway”, “base excision repair” and “proteasome” were found to be significantly enriched in patients of CIMP-M subgroup compared that in patients of CIMP-L subgroup (**Figure 8B**). These results may suggest that the CIMP status may influence the prognosis of PC patients by regulation of the aforementioned biological process.

### Validation of CIMP status in PC patients of ICGC dataset

To independently test the CIMP status in PC patients, DNA methylation data of 264 PC patients, somatic mutation data of 264 PC patients, mRNA expression of 175 PC patients and clinical-pathological parameter of 264 PC patients were downloaded from ICGC (<https://dcc.icgc.org/>). Unsupervised consensus clustering analysis was also performed for 264 PC patients based on the expression of these 72 OS-related CpGs. Similarly, these 264 PC patients were also clustered into three distinct groups (**Figure 9A**), there were 58 PC patients in CIMP-L subgroup, 171 PC patients in CIMP-M subgroup, and 35 PC patients in the CIMP-H subgroup. Besides, significant associations between CIMP status and clinical-pathological characteristics, including age, T state, N stage, M stage and TNM stage, were also observed (**Table 4**). Moreover, Kaplan-Meier analysis showed that there were significant differences in OS among PC patients with different CIMP status. Patients of the CIMP-H subgroup had poorer OS while the patients of CIMP-L subgroup had better OS ( $p=0.003$ , **Figure 9A**). Univariate and multivariate Cox analysis also suggested that the CIMP status were significantly related with OS, and was also an independent prognostic factor for OS of PC patients after adjusting for gender, age, T state, N stage, M stage and TNM stage ( $p=0.012$ , **Table 5**).

The association of CIMP status with gene mutations was further analyzed. In line with the results in TCGA, all the 35 PC patients of the CIMP-H subgroup (100%) had gene mutation; 168 out of 171 PC patients of the CIMP-M subgroup (98%) had gene mutation, while only 35 out of 58 PC patients of the CIMP-L subgroup (58.66%) had gene mutation. Obviously, higher somatic mutation burdens in KRAS, TP53, CDKN2A and TTN were found at patients of CIMP-H group compared to that in patients of CIMP-L group (**Figure 9B**). Besides, higher TMB were also found at patients of CIMP-H subgroup, while lower TMB were observed at patients with CIMP-L status (**Figure 9B**).

The landscape of TME among PC patients with distinct CIMP status was also explored. As expected, significantly lower stromal score, immune score and estimate score were found at patients of CIMP-L subgroup. Meanwhile, significantly lower total T cells, total B cells, CD8 T cells, memory CD4 T cells were found in patients of CIMP-H subgroup, but significantly higher follicular helper T cells were found in patients of CIMP-H subgroup (**Figure 9C**).

Finally, the association of CIMP status with immune checkpoint genes was also analyzed. Only expression of 4 immune checkpoint genes, including PD-1, PD-L1, CD86 and CD276 were available. Similarly, lower expression of PD-1, PD-L1 and CD86 were found at patients of CIMP-H subgroup, but the difference was not statistically significant (**Figure 9D**).

## Discussion

Identification of clinically relevant cancer subtypes based on DNA methylation pattern is an important computational problem in medicine, which may help to provide specific and effective treatment options for patients with different subtypes. Previously, DNA methylation pattern analysis of PC patients have been performed but were limited to low sample size and small number of CpGs sites (5, 11, 14). Sato et al. analyzed the genome-scale DNA methylation patterns in PC patients, but only 8 genes of methylation specific site were used (14). Thompson et al. explored the association of DNA methylation patterns with survival of PC patients, but number of samples used for analysis was small, in which only 11 PC patients, 2 normal controls and 3 chronic pancreatitis patients were included (2). Recently, Nitish et al. identified three CIMP subtypes of PC patients with distinct clinical characteristic and gene mutation landscape by clustering of differentially methylated sites using the genome-scale methylome data of PC patients from TCGA; however,

they did not further explore the relationship among CIMP status, prognosis of PC patients and TME. They also did not validate the three CIMP subtypes in independent PC cohorts (11). In the present study, we identified and validated three distinct CIMP subgroups (termed CIMP-L, CIMP-M, and CIMP-H subgroup) in 448 PC patients with the data downloaded from TCGA and ICGC. In agreement with previous studies, we observed that CIMP status was significantly associated with clinical characteristics, OS and RFS in PC patients. There were more patients with advanced T stage and TNM stage in the CIMP-M subgroup and CIMP-H subgroup. Patients of CIMP-H subgroup had the worst OS and RFS. Moreover, the CIMP status was also an independent prognostic factor for OS and DFS in PC patients.

Accumulation of somatic mutations in oncogenes and TSGs were common in the development and progression of cancer (27). Based on the mutation analysis, significantly higher somatic mutation burdens in KRAS, TP53, SMAD4, CDKN2A were observed in patients of CIMP-H subgroup, which had been shown to be major driver genes in PC (23). Encoding a small GTPase that involved in cellular proliferation, motility, and cytoskeletal remodeling, KRAS was the most frequently mutated oncogene in PC. More than 90% of PC patients showed somatic mutations in KRAS (28). CDKN2A encoded an essential cell-cycle regulator and was reported to be the most frequently mutated TSGs in PC. Similarly, more than 90% of PC patients exhibited function defect of CDKN2A because of gene mutation (28). Besides, studies have showed that somatic KRAS and CDKN2A mutations were early events of PC development as they were the earlier alteration genes in most low-grade pancreatic intraepithelial neoplasia (29). TP53 played a vital role in the cellular stress response. Somatic mutations in TP53 were also observed frequently in a wide range of tumor types, including PC. SMAD4 was found to mainly mediate signaling downstream of TGF $\beta$  receptor and was inactivated in about 50% of PC patients. Alterations in p53 and SMAD4 were late events in PC, as they often occurred in pancreatic patients with histologic grade 3 and high invasiveness (29). Consistent with the results in gene mutation, KEGG pathways, such as "P53 signaling pathway," was found to be significantly enriched in patients of CIMP-H subgroup. Moreover, CIMP may also act as a tumor promoter in PC carcinogenesis by influencing mutation of major driver genes, and PC patients of different CIMP subtypes originated from precursor cells might have a distinct epigenetic background of the cell of origin.

Consists of cancer cells, stromal cells (mainly composed of fibroblasts and immune cells) and extracellular components, TME has been found to promote tumor progression, metastasis niche formation and therapeutic resistance in PC (24). In the present study, we also found that CIMP status was significantly associated with TME. On one hand, significantly higher tumor purity and cancer stemness indices were found in patients of CIMP-H subgroup. Researches have proven that loss of differentiated phenotype, acquisition of progenitor and stem cell-like features were the main hallmarks of cancer progression. It was speculated that cancer cells might arise from a cell population with self-renewal ability, which was thought to be CSCs. Pancreatic cancer stem cells (PCSCs) have been first identified in 2007 and were reported to take part in the resistance to standard chemotherapy and radiation treatment as they could express multidrug resistant membrane transporters, aberrantly activate proliferation signaling pathways and increase the capability of repairing DNA (30). On the other hand, lower stromal score and immune score were found at PC patients of CIMP-H subgroup. Consistently, lower proportion of total T cells, total B cells, naive B cells, CD8 T cells, CD4 T cells, memory CD4 T cells and activated memory CD4 T cells were also found at patients of CIMP-H subgroup, which indicated that patients of CIMP-H subgroup had a distinct immune phenotype, characterized by less immune cells infiltration, lower cytotoxic potential and immune activation. Moreover, many kinds of chemokines, such as CCL2, XCL2, CCR2, CCL5 and CCR5, were also found to be decreased at patients of CIMP-H subgroup, which may in part account for the lower infiltration of immune cells as many different kinds of immune cells were mainly recruited into the TME via interactions between chemokines and their chemokine receptors (25). Taken together, the higher tumor purity and cancer stemness indices, and lower infiltration of immune cells in patients of CIMP-H subgroup may contribute to their worse OS and DFS.

The advent of immunotherapy, especially checkpoint inhibitor-based immunotherapy, have revolutionized cancer treatments, especially for advanced cancer patients (26). Currently, the monoclonal antibodies, targeted for PD-1 and its ligands, have been successfully applied in clinical practice and have been approved for several cancers (such as melanoma, non-small cell lung carcinoma, renal cancer, and bladder cancer) (31). Besides, researches have showed that combination of anti-PD-1/PD-L1 antibody and CTLA-4 inhibitor can improve treatment effect of patients with advanced melanoma, which was approved by FDA in treating BRAF V600E wild-type patients with unresectable or metastatic melanoma (32, 33). For example, a recent phase-II clinical trial have proved that combination of ipilimumab (a CTLA-4 inhibitor) and nivolumab (a PD-1 inhibitor) significantly improved treatment efficacy in advanced melanoma patients compared with monotherapy with ipilimumab (34). In our study, significant association of CIMP status with expression of immune checkpoint genes was found. Lower expression of PD-1, CTLA4, CD86, VTCN1 and LAG-3 were found at PC patients of CIMP-H subgroup. Consider the expression of immune checkpoint genes were necessary for checkpoint inhibitor immunotherapy. Effective

immunotherapy of immune checkpoint inhibitors depended on the generation of neoantigen-specific T-cells and its infiltration in the TME. Immunotherapies were less likely to be efficacious in patients of CIMP-H phenotype as lower expression of immune checkpoint genes and T cells infiltration were found.

As changeable and possibly heritable genetic alteration, epigenetic regulation may prove promising clues for the treatment of various kinds of diseases, including cancers (35). In-vitro studies have shown that 5-Aza-CdR, a DNA methyltransferase 1 (DNMT1) inhibitor, could induce cell death and apoptosis of pancreatic cancer cells by reactivation of RASSF1A and up-regulation of Bax genes (36). Han et al. also observed the synergistic effects of the combination of 5-Aza-CdR and suberoylanilide hydroxamic acid on the anticancer property of PC (37). More inspiring, phase I/II clinical trials of DNMT1 inhibitors (azacitidine, decitabine and guadecitabine) in PC patients were currently underway, the inhibitors of which exhibited potential treatment outcomes (38). It is a promising future in drugs designing for epigenetic targets, while CIMP may help to make an assertion to provide specific and efficient treatment options for patients with different CIMP status.

However, there were some limitations to be addressed in our study. First, our analysis was performed on the basis of single-omics (DNA methylation). Patients with the same CIMP status might have heterogeneity due to the different characteristics in terms of other omics data. Second, our analysis was performed on the basis of retrospective cohort. Prospective studies with larger sample sizes should be performed to validate our findings. Finally, the biological functions and molecular mechanisms of CIMP status to influence the prognosis of PC patients should be further validated in-vitro experiment.

In conclusion, we identified and validated three distinct CIMP subgroups in PC patients. Patients of different CIMP subgroups showed distinct clinical characteristic, prognosis, gene mutation landscape and tumor microenvironment (TME). It may help to make an assertion to provide specific and efficient treatment options for patients with different CIMP status.

## Abbreviations

PC: pancreatic cancer, TSGs: tumor suppressor genes, CIMP: CpG island methylator phenotype, TME: tumor microenvironment, TCGA: The Cancer Genome Atlas, ICGC: International Cancer Genome Consortium, TMB: tumor mutational burden, OS: overall survival, PFS: progression free status, PCA: principal component analysis, ESTIMATE: estimation of stromal and immune cells in malignant tumor tissues using expression data, ssGSEA: single sample Gene Set Enrichment Analysis, OCLR: one-class logistic regression machine learning algorithm, GSEA: Gene set enrichment analysis, CSCs: cancer stem cells, PCSCs: pancreatic cancer stem cells, DNMT1: DNA methyltransferase 1.

## Declarations

### Ethics approval

All the data analyzed in the present study were got from TCGA and ICGC. Informed consents had already been obtained from the patients before the present study.

### Consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The data of the study are available from the corresponding web page link, including GDC Data portal (<https://cancergenome.nih.gov/>) and ICGC portal ICGC (<https://dcc.icgc.org/>).

### Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Funding

The study was supported by National Natural Science Foundation of China (81970507), Natural Science Foundation of Guangdong Province (2020A1515010100), Technology Planning Project of Guangzhou City (201904010132), Guangzhou High Technology Project (2019GX05), Guangdong Medical Science and Research Foundation (A2020411), Guangzhou General Science and Technology Project of Health and Family Planning (20191A01100).

## Code availability

Not applicable.

## Authors' contributions

NG and L-YQ performed most of the data analysis and wrote the manuscript. C-WJ assisted in the data collection and analysis of the study. T-WJ assisted in data collection and analysis. S-DW was involved in writing the manuscript. L-QL assisted in the revision of the manuscript. Z-YJ and C-HT designed the study and obtained funding to support this research.

## Acknowledgements

Not applicable.

## References

1. Bray F, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J Clin*. 2018;68:394–424.
2. Thompson MJ, Rubbi L, Dawson DW, Donahue TR, Pellegrini M. Pancreatic cancer patient survival correlates with DNA methylation of pancreas development genes. *PLoS one*. 2015;10:e0128814.
3. Neureiter D, Jager T, Ocker M, Kiesslich T. Epigenetics and pancreatic cancer: pathophysiology and novel treatment aspects. *World journal of gastroenterology*. 2014;20:7830–48.
4. Chakraborty S, Baine MJ, Sasson AR, Batra SK. Current status of molecular markers for early detection of sporadic pancreatic cancer. *Biochim Biophys Acta*. 2011;1815:44–64.
5. Tan AC, et al. Characterizing DNA methylation patterns in pancreatic cancer genome. *Molecular oncology*. 2009;3:425–38.
6. Yang N, Sen P. The senescent cell epigenome. *Aging*. 2018;10:3590–609.
7. Hansen KD, et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*. 2011;43:768–75.
8. Bailey P, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*. 2016;531:47–52.
9. Kamekar S, et al. Exosomes facilitate therapeutic targeting of oncogenic KRAS in pancreatic cancer. *Nature*. 2017;546:498–503.
10. Noshmehr H, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer cell*. 2010;17:510–22.
11. Mishra NK, Guda C. Genome-wide DNA methylation analysis reveals molecular subtypes of pancreatic cancer. *Oncotarget*. 2017;8:28990–9012.
12. Zhang C, et al. CpG island methylator phenotype association with elevated serum alpha-fetoprotein level in hepatocellular carcinoma. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2007;13:944–52.
13. Malta TM, et al. Glioma CpG island methylator phenotype (G-CIMP): biological and clinical implications. *Neurooncology*. 2018;20:608–20.
14. Sato N, Fukushima N, Hruban RH, Goggins M. CpG island methylation profile of pancreatic intraepithelial neoplasia. *Modern pathology: an official journal of the United States Canadian Academy of Pathology Inc*. 2008;21:238–44.
15. (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513: 202–209.
16. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research*. 2018;28:1747–56.
17. Chalmers ZR, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome medicine*. 2017;9:34.

18. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015;43:e47.
19. Yoshihara K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications*. 2013;4:2612.
20. Malta TM, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell*. 2018;173:338–54 e315.
21. Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*. 2015;12:453–7.
22. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–50.
23. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. *Lancet*. 2016;388:73–85.
24. Neesse A, et al. Stromal biology and therapy in pancreatic cancer: ready for clinical translation? *Gut*. 2019;68:159–71.
25. Nagarsheth N, Wicha MS, Zou W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat Rev Immunol*. 2017;17:559–72.
26. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol*. 2016;17:e542–51.
27. Vogelstein B, et al. Cancer genome landscapes. *Science*. 2013;339:1546–58.
28. Collisson EA, Bailey P, Chang DK, Biankin AV. Molecular subtypes of pancreatic cancer. *Nat Rev Gastroenterol Hepatol*. 2019;16:207–20.
29. Kanda M, et al. Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. *Gastroenterology*. 2012;142:730–3 e739.
30. Zhou P, et al. The epithelial to mesenchymal transition (EMT) and cancer stem cells: implication for treatment resistance in pancreatic cancer. *Mol Cancer*. 2017;16:52.
31. Raval RR, Sharabi AB, Walker AJ, Drake CG, Sharma P. Tumor immunology and cancer immunotherapy: summary of the 2013 SITC primer. *J immunother Cancer*. 2014;2:14.
32. Boutros C, et al. Safety profiles of anti-CTLA-4 and anti-PD-1 antibodies alone and in combination. *Nat Rev Clin Oncol*. 2016;13:473–86.
33. Baumeister SH, Freeman GJ, Dranoff G, Sharpe AH. Coinhibitory Pathways in Immunotherapy for Cancer. *Annu Rev Immunol*. 2016;34:539–73.
34. Hodi FS, et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol*. 2016;17:1558–68.
35. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. *Nat Rev Genet*. 2016;17:630–41.
36. Nikbakht Dastjerdi MP, et al. An Effective Concentration of 5-Aza-CdR to Induce Cell Death and Apoptosis in Human Pancreatic Cancer Cell Line through Reactivating RASSF1A and Up-Regulation of Bax Genes. *Iranian journal of medical sciences*. 2018;43:533–40.
37. Han T, Zhuo M, Hu H, Jiao F, Wang LW. Synergistic effects of the combination of 5-AzaCdR and suberoylanilide hydroxamic acid on the anticancer property of pancreatic cancer. *Oncol Rep*. 2018;39:264–70.
38. Wong KK. (2020) DNMT1 as a therapeutic target in pancreatic cancer: mechanisms and clinical implications. *Cell Oncol (Dordr)*.

## Tables

Table 1 Demographic and clinical characteristics of PAAD patients with different CIMP status from TCGA

Variables	CIMP-L	CIMP-M	CIMP-H	<i>p</i> -value
Number of patients	46	82	56	
Gender ( Male/female)	21/25	48/34	32/24	0.343
Age (years, ≤65/>65)	27/19	43/39	26/29	0.466
T stage (T1/T2/ T3/T4/NA)	5/10/30/0/1	0/8/70/3/1	2/5/48/1/0	0.008*
N stage (N0/N1/NX+NA)	14/29/3	17/64/1	19/36/1	0.13
M stage (M0/M1/MX+NA)	22/1/23	34/2/46	28/2/26	0.829
TNM stage (I/II/III /IV/NA)	12/32/0/1/1	4/71/4/2/1	5/48/1/2/0	0.014*

**Table 2 Univariate and multivariate analysis of overall survival and progression free survival in PAAD patients of TCGA cohort.**

Variables	OS						PFS					
	Univariate analysis			Multivariate analysis			Univariate analysis			Multivariate analysis		
	Hazard ratio	95% CI	P value	Hazard ratio	95% CI	P value	Hazard ratio	95% CI	P value	Hazard ratio	95% CI	P value
Gender (Male vs female)	0.84	0.565-1.25	0.39				0.929	0.637-1.354	0.701			
Age (>65 vs ≤65)	1.302	0.873-1.943	0.195				1.237	0.843-1.815	0.277			
T stage (T3+T4 vs T1+T2)	2.28	1.179-4.407	0.014*	1.584	0.796-3.15	0.19	2.651	1.406-4.999	0.003*	2.037	1.053-3.942	0.035*
N stage (N1+Nx vs N0)	2.117	1.278-3.508	0.004*	1.828	1.096-3.049	0.021*	1.603	1.035-2.482	0.034*	1.48	0.946-2.316	0.086
M stage (M1+Mx vs M0)	0.844	0.567-1.257	0.404				0.58	0.396-0.851	0.005*	0.725	0.491-1.0	0.105
TNM stage (stage III+IV vs stage I+II)	1.0	0.31-3.225	1.0				0.994	0.461-2.154	0.994			
CIMP status												
CIMP-M vs CIMP-L	2.039	1.142-3.642	0.016*	1.72	0.948-3.122	0.074	1.585	0.934-2.688	0.088			
CIMP-H vs CIMP-L	2.289	1.245-4.206	0.008*	2.043	1.096-3.806	0.024*	2.393	1.405-4.075	0.001*	1.956	1.129-3.388	0.017*

Table 3 Expression of 59 kinds of chemokine among of PC patients with different CIMP status.

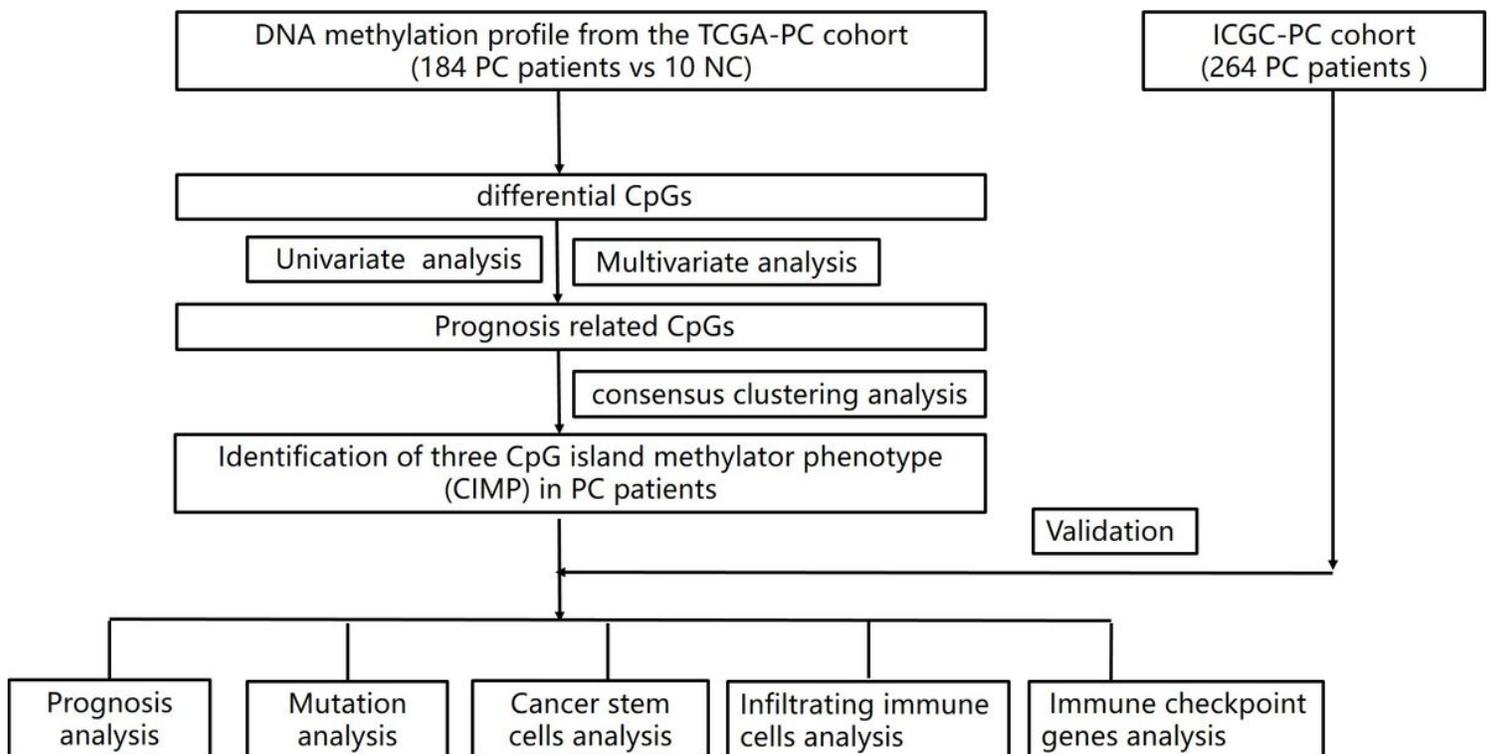
Gene	CIMP-L(N=46)		CIMP-M(N=76)		CIMP-H(N=56)		p value
	Mean	Sd	Mean	Sd	Mean	Sd	
CCL1	0.019	0.034	0.021	0.037	0.020	0.045	0.9417
CXCR6	2.238	1.739	1.344	1.406	0.669	0.483	0.000***
CCL2	35.294	33.754	19.973	17.815	11.123	9.482	0.000***
CCL19	89.272	118.604	52.134	78.701	16.316	27.712	0.001***
XCL2	1.092	0.904	0.696	0.675	0.446	0.439	0.000***
CX3CL1	11.107	6.235	14.391	10.037	15.302	17.829	0.2071
CCR10	0.616	1.214	0.360	0.212	0.298	0.183	0.0331*
CCL25	0.224	0.518	2.914	14.716	0.225	1.103	0.1878
CXCR4	86.490	125.355	57.819	71.923	31.478	18.115	0.0030**
CCL28	7.608	8.420	7.050	6.102	6.124	4.299	0.4828
CXCL2	13.775	24.638	9.186	7.417	7.984	24.495	0.2841
CCR2	2.095	1.697	1.318	1.272	0.603	0.563	0.000***
CCL5	32.342	26.573	23.419	21.406	13.148	10.918	0.000***
CCR9	0.185	0.186	0.248	0.333	0.248	0.225	0.4016
XCR1	0.636	0.602	0.341	0.432	0.148	0.224	0.000***
CCL8	1.405	1.377	1.051	1.784	0.751	1.491	0.1232
CXCL11	2.560	3.744	3.279	6.261	2.675	4.345	0.6973
CXCL1	18.895	26.105	16.609	12.493	15.503	28.623	0.7422
CXCL17	26.545	34.664	62.098	85.526	64.257	86.579	0.0215*
CXCR5	0.229	0.490	0.120	0.292	0.028	0.051	0.0069**
CCL16	0.141	0.109	0.084	0.065	0.038	0.041	0.000***
CCR5	3.709	2.898	2.485	2.357	1.251	0.993	0.000***
CXCL8	18.044	18.167	25.857	27.721	16.058	16.663	0.0311*
CXCL14	62.254	72.601	97.927	80.887	132.875	168.641	0.0091**
CCR8	0.536	0.583	0.446	0.435	0.307	0.366	0.0402*
CCL15	1.298	1.478	1.786	2.572	1.556	2.122	0.4906
CCL21	108.128	193.077	72.864	195.196	24.227	43.603	0.0340*
CCL23	1.964	1.921	1.821	3.553	0.828	1.388	0.0494*
CCR7	8.640	15.561	4.779	12.140	1.401	1.492	0.0061**
CCL3	3.676	3.275	2.509	1.904	1.737	1.501	0.001***
CCL7	0.290	0.598	0.386	0.491	0.529	0.928	0.2030
CXCL10	10.127	15.566	11.441	24.734	9.581	13.448	0.8535
CCL24	3.600	3.986	23.088	85.332	20.610	94.433	0.3719
CX3CR1	1.521	1.261	1.015	0.941	0.576	0.733	0.000***
CCL4	4.747	3.162	2.978	2.468	1.757	1.277	0.000***

CCL22	4.591	5.368	3.777	3.472	2.011	1.733	0.0014**
CXCL13	21.519	45.303	14.589	37.830	4.694	11.043	0.0462*
CXCR2	0.684	0.870	0.648	0.801	0.217	0.310	0.0007**
CXCL6	11.406	16.524	10.446	13.190	5.032	4.956	0.0152*
CCL14	1.066	0.998	0.577	0.508	0.162	0.145	0.000***
CCL26	3.888	9.430	3.928	16.757	1.956	3.101	0.6057
CCR4	2.337	2.796	1.304	1.546	0.520	0.547	0.000***
CCL17	10.386	19.405	8.800	11.544	3.906	5.125	0.0245*
CXCR1	0.662	0.839	0.646	0.819	0.288	0.357	0.0076**
CCR1	5.030	3.568	3.859	3.112	2.588	1.995	0.000***
CXCL5	38.527	64.259	81.274	124.396	36.474	49.043	0.008**
CXCL9	18.918	42.376	14.706	50.569	9.541	19.449	0.5112
XCL1	0.452	0.345	0.409	0.472	0.377	0.674	0.7662
CXCR3	2.298	2.170	2.379	2.067	1.941	1.557	0.4239
CCL18	21.990	42.000	29.714	29.962	15.548	23.234	0.0408*
CCL13	4.006	5.276	4.984	4.601	4.791	7.459	0.6551
CCR6	0.141	0.194	0.098	0.117	0.049	0.057	0.0019**
CCR3	0.086	0.101	0.110	0.159	0.147	0.449	0.5221
CXCL16	38.590	19.162	48.610	16.384	49.334	14.780	0.0018**
CCL20	8.951	12.180	25.344	41.318	19.521	29.584	0.0270*
CCL11	3.908	3.675	3.800	3.876	1.647	1.317	0.000***
CCL27	0.000	0.003	0.000	0.000	0.001	0.005	0.2795
CXCL3	4.555	9.367	5.720	6.384	5.472	13.359	0.8110
CXCL12	25.525	16.445	12.292	10.336	4.826	3.670	0.000***

Table 4 Demographic and clinical characteristics of PAAD patients with different CIMP status from ICGC

Variables	CIMP-L	CIMP-M	CIMP-H	<i>p</i> -value
Number of patients	58	171	35	
Gender ( Male/female/NA)	35/23/0	95/75/1	20/15	0.904
Age (years, ≤65/>65)	28/30/0	77/93/1	11/20/4	0.009**
T stage (T1/T2/ T3/T4/NA)	5/3/46/02/2	1/817/145/2/6	0/3/20/1/11	0.000***
N stage (N0/N1/NX+NA)	12/44/2	46/119/6	7/16/12	0.000***
M stage (M0/M1/MX+NA)	14/2/42	24/6/141	1/4/30	0.016*
TNM stage (I/II/III /IV/NA)	6/48/1/2/1	8/150/2/6/5	1/20/0/4/10	0.000***

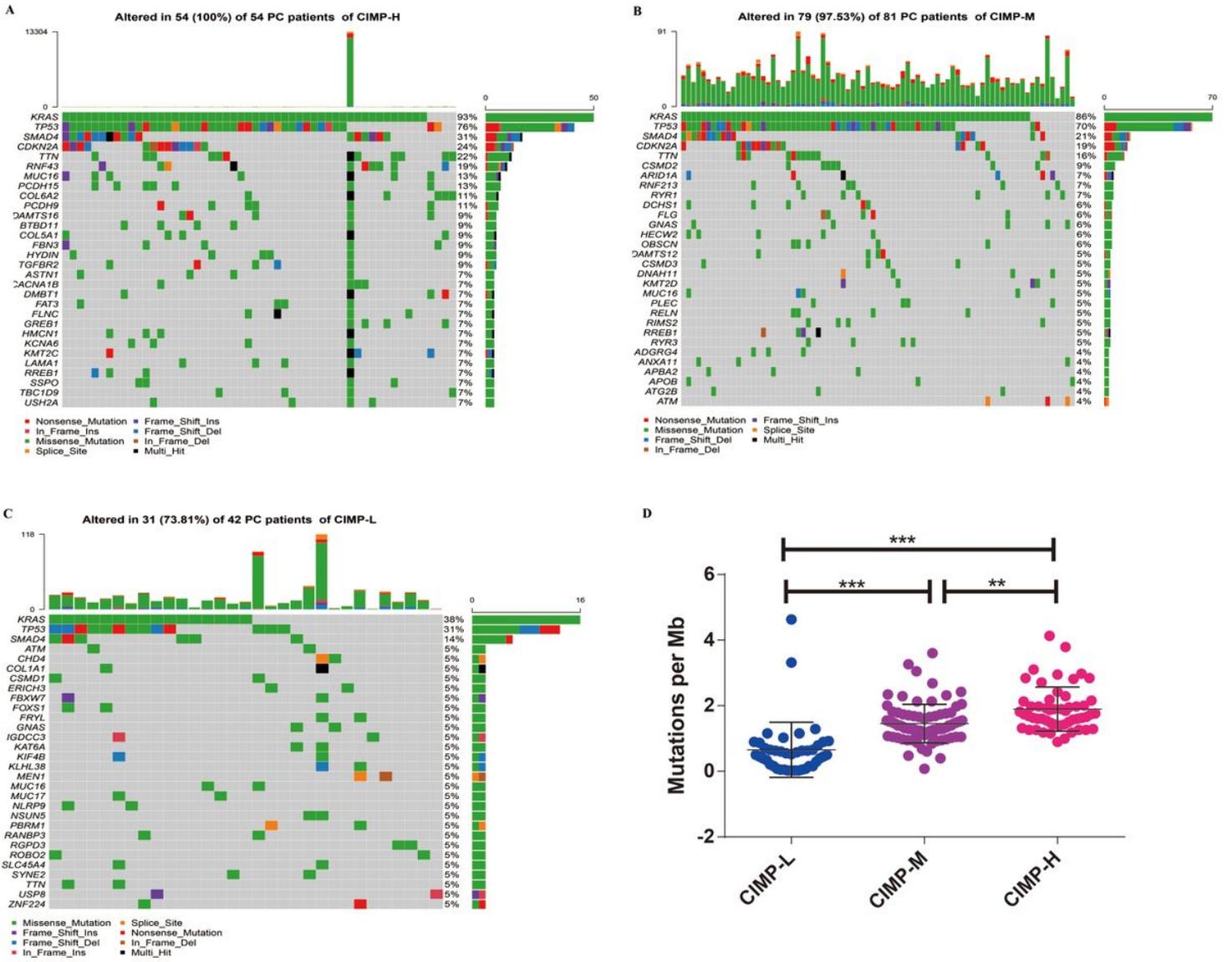
## Figures



**Figure 1**

The workflow chart of the present study.





**Figure 3**

Mutational landscapes of PC patients with different CIMP status. 30 most significantly mutated genes in the PC patients from CIMP-H subgroup (A), CIMP-M subgroup (B) and CIMP-L subgroup (C); expression of TMB among PC patients with different CIMP status (D).

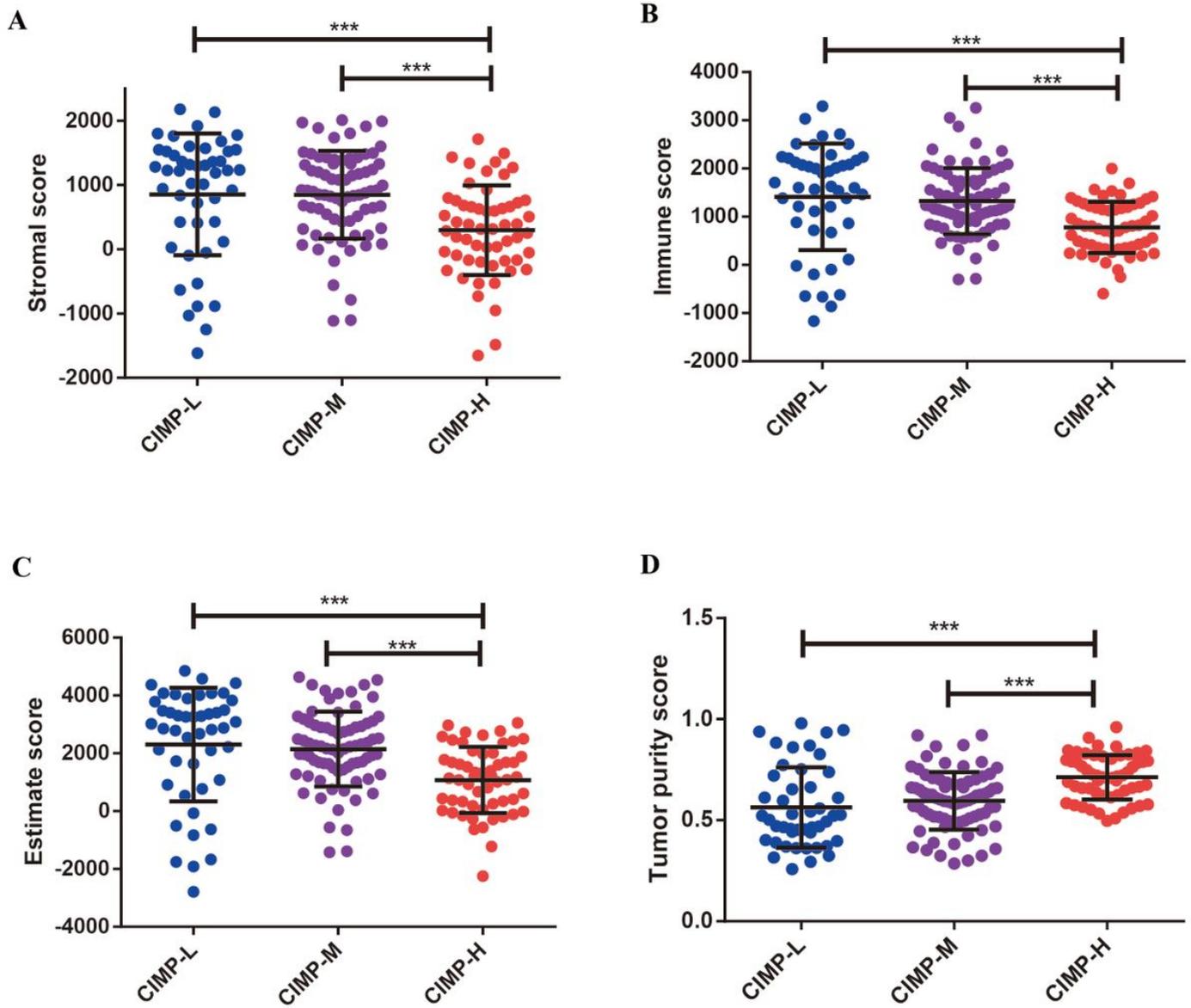


Figure 4

Association of CIMP status with tumor microenvironment in PC patients with distinct CIMP status. Expression of stromal score (A), immune score (B), estimate score (C) and tumor purity (D) among PC patients with distinct CIMP status.

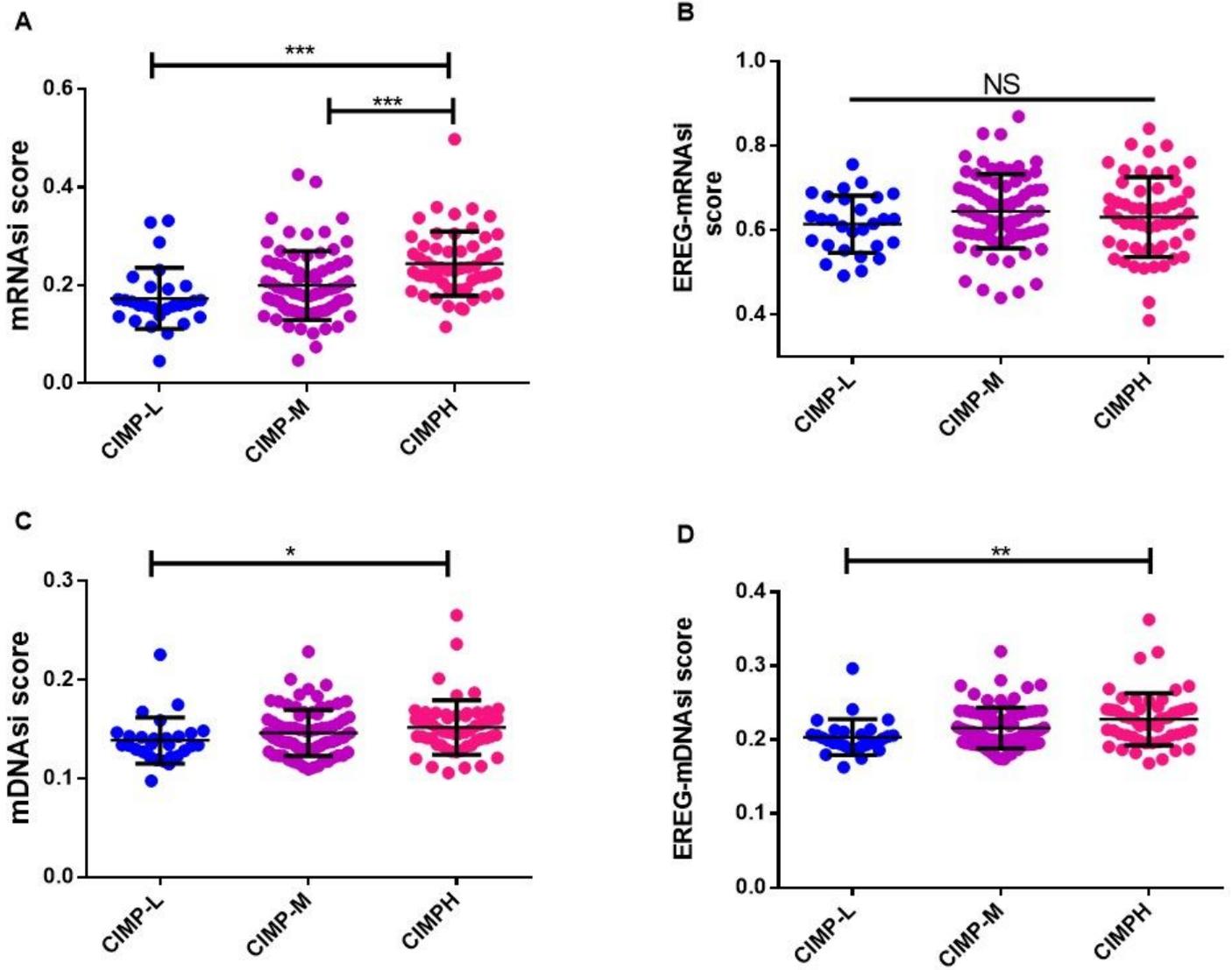
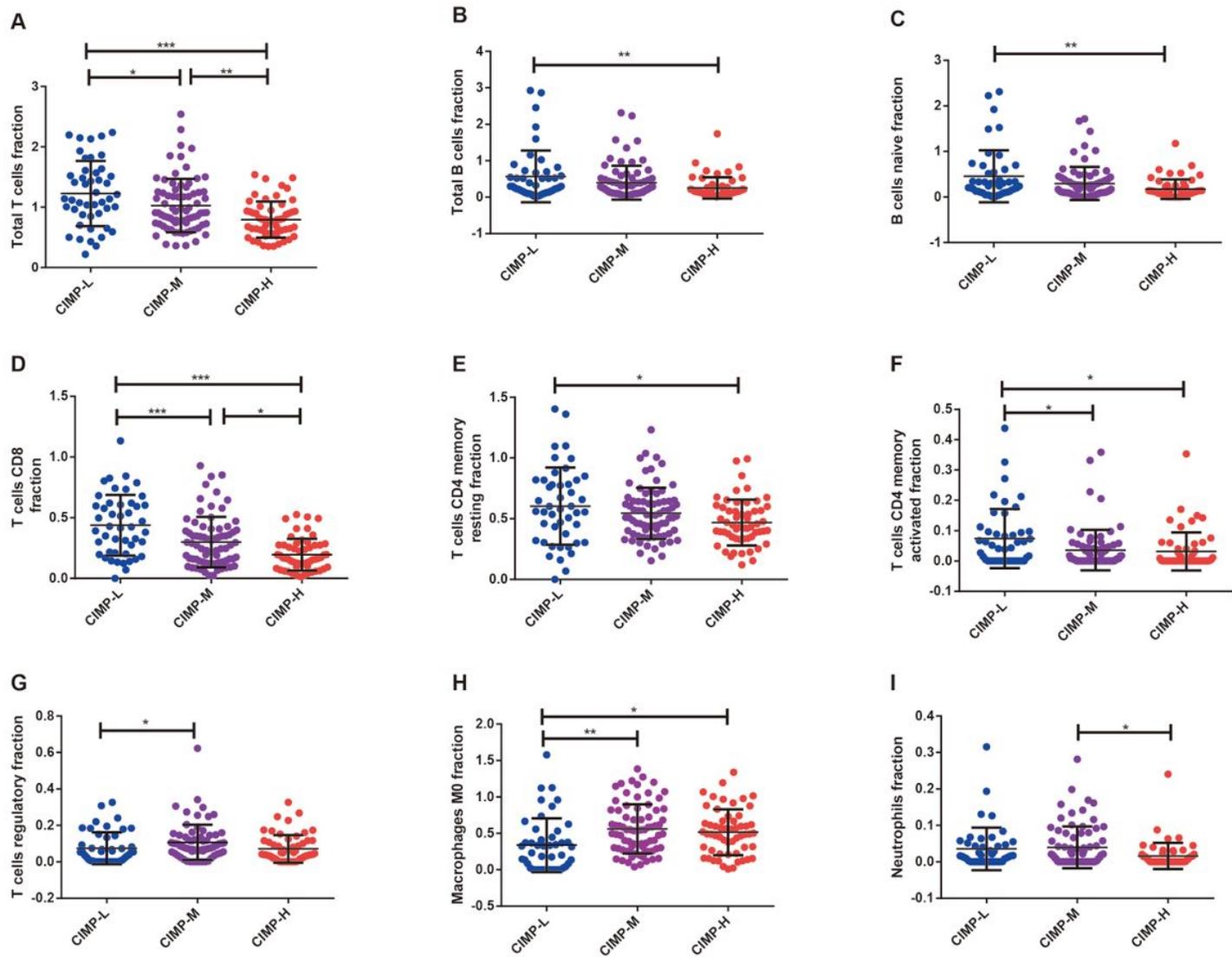


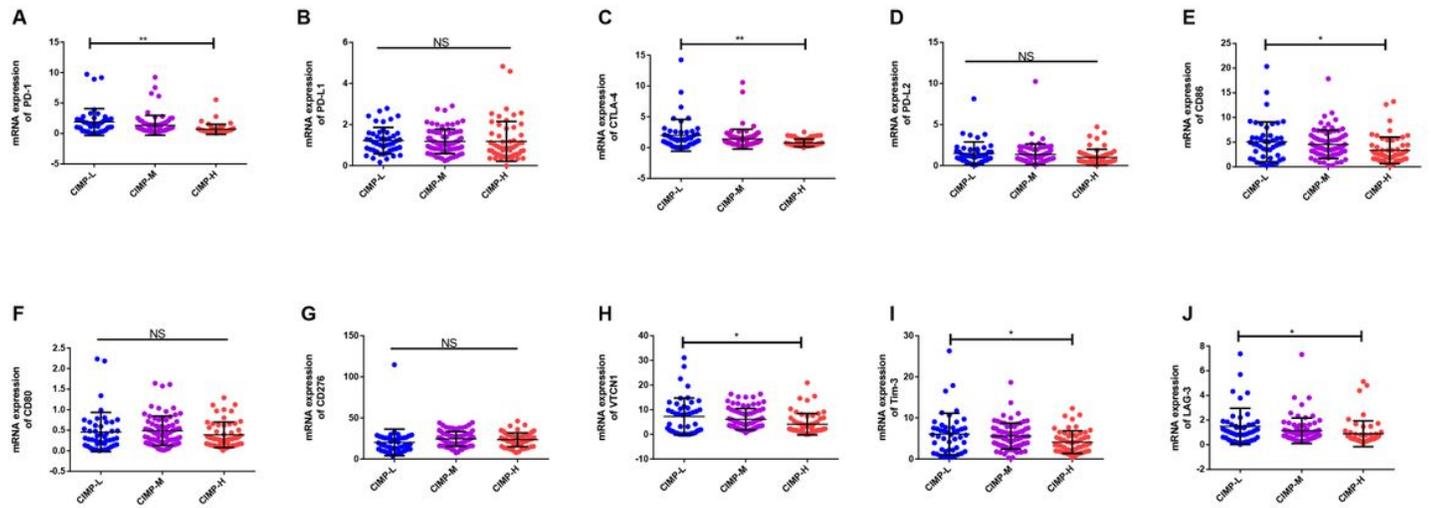
Figure 5

Association of CIMP status with cancer stemness indices of PC patients. Expression of mRNA<sub>si</sub> (A), EREG-mRNA<sub>si</sub> (B), mDNA<sub>si</sub> (C) and EREG-mDNA<sub>si</sub> (D) among PC patients with different CIMP status.



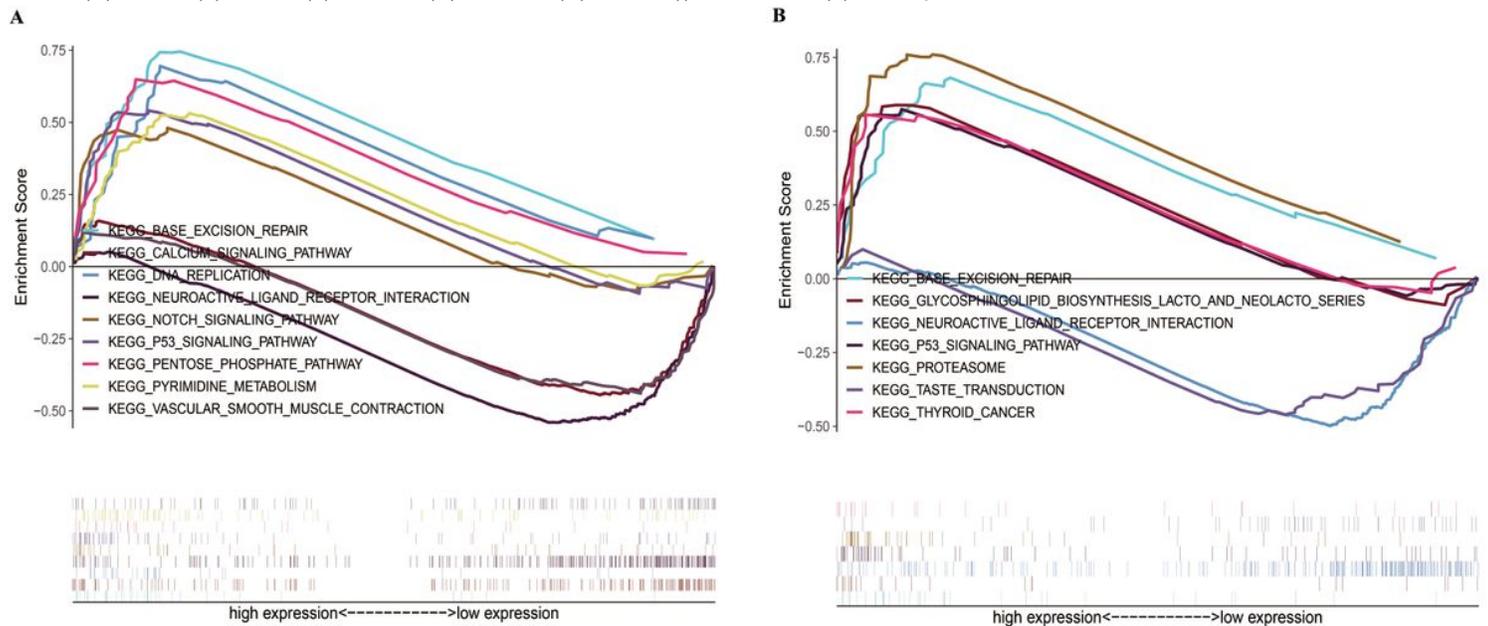
**Figure 6**

Landscapes of tumor-infiltrating immune cells in PC patients with different CIMP status. Expression of total T cells (A), total B cells (B), naive B cells (C), CD8 T cells (D), resting memory CD4 T cells (E), activated memory CD4 T cells (F), regulatory T cells (G) and M0 macrophages (H) among PC patients with different CIMP status.



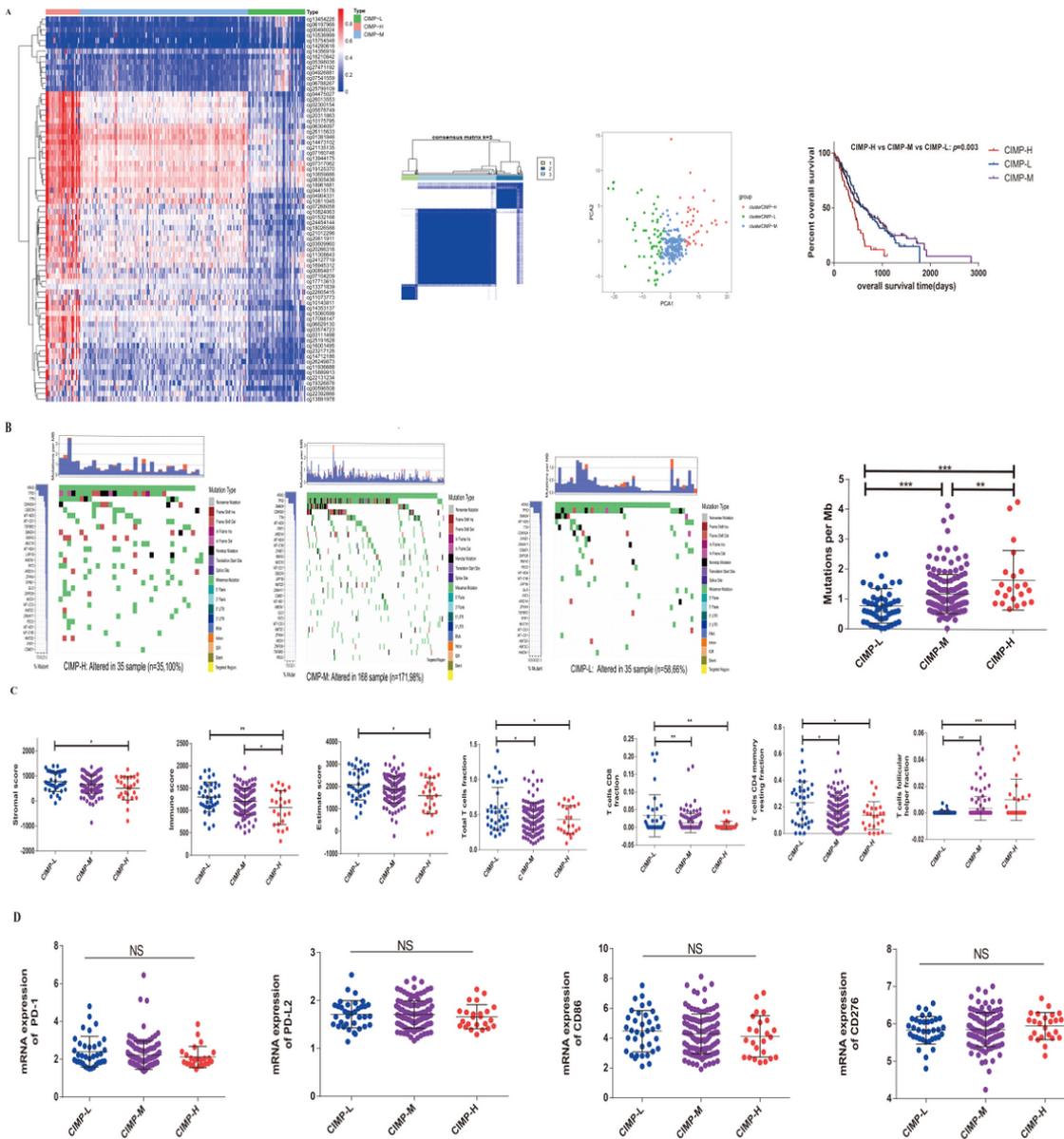
**Figure 7**

Expression of 10 immune checkpoint genes in PC patients with different CIMP status. Expression of PD-1 (A), PD-L1 (B), CTLA4 (C), PD-L2 (D), CD86 (E), CD80 (F), CD276 (G), VTCN1 (H), Tim-3 (I) and LAG-3 (J) in PC patients with distinct CIMP status.



**Figure 8**

Potential mechanism by which CIMP status to influence the prognosis of PC patients. Significant KEGG pathways enriched in PC patients of CIMP-H subgroup (A); significant KEGG pathways enriched in PC patients of CIMP-M subgroup (B).



**Figure 9**

Validation of CIMP status in 264 PC patients of ICGC dataset. Validation of three CIMP subgroup in PC patients of ICGC dataset and the association of CIMP status with clinical-pathologic parameters and prognosis (A); mutational landscapes of three CIMP subgroup and expression of TMB among PC patients with different CIMP status (B); expression of stromal score, immune score, estimate score and tumor-infiltrating immune cells among PC patients with distinct CIMP status (C); expression of 4 immune checkpoint genes in PC patients with different CIMP status (D).