Computational and Structural Biotechnology Journal 21 (2023) 677–687

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Dynamic gene screening enabled identification of a 10-gene panel for early detection and progression assessment of gastric cancer

Fei Long [a,b,1], Shuo Li [a,b,1], Yaqi Xu [a], Min Liu [a], Xuan Zhang [a], Junting Zhou [a], Yiyi Chen [a], Yuan Rong [b,d,*], Xiangyu Meng [b,c,*], Fubing Wang [a,b,c,**]

[a] Department of Laboratory Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China
[b] Center for Single-Cell Omics and Tumor Liquid Biopsy, Zhongnan Hospital of Wuhan University, Wuhan, China
[c] Wuhan Research Center for Infectious Diseases and Cancer, Chinese Academy of Medical Sciences, Wuhan, China
[d] Forensic Center of Justice, Zhongnan Hospital of Wuhan University, Wuhan China

ARTICLE INFO

Article history:
Received 28 September 2022
Received in revised form 10 December 2022
Accepted 19 December 2022
Available online 22 December 2022

Keywords:
Biomarker
Gastric cancer
Precancerous lesions
Diagnosis
Progression assessment

ABSTRACT

Early diagnosis and progression assessment are critical for the timely detection and treatment of gastric cancer (GC) patients. Identification of diagnostic biomarkers for early detection of GC represents an unmet clinical need, and how these markers further influence GC progression is explored rarely. We performed dynamic gene screening based on high-throughput data analysis from patients with precancerous lesions and early gastric cancer (EGC) and identified a 10-gene panel by the lasso regression model. This panel demonstrated good diagnostic performance in TCGA (AUC = 0.95, sensitivity = 86.67 %, specificity = 90.63 %) and GEO (AUC = 0.84, sensitivity = 91.67 %, specificity = 78.13 %) cohorts. Moreover, three GC subtypes were clustered based on this panel, in which cluster 2 (C2) demonstrated the highest tumor progression level with a high expression of 10 genes, showing a decreased tumor mutation burden, significantly enriched epithelial-mesenchymal transition hallmark and increased immune exclusion/exhausted features. Finally, the cell localization of these panel genes was explored in scRNA-seq data based on more than 40,000 cells. The 10-gene panel is expected to be a new clinical early detection signature for GC and may aid in progression assessment and personalized treatment of patients.

1. Introduction

Gastric cancer (GC) is the fifth most common human malignancy (5.6 % of total cases) and the fourth most common cause of cancer-related deaths (7.7 % of total cancer deaths) worldwide [1]. Its 5-year survival rate is less than 20 % in advanced cases, but can go up to 90 % in early stage gastric cancer (EGC) patients [2,3]. Therefore, early diagnosis is important to enable timely and effective treatment decisions to improve patients' survival. According to Lauren's classification, GC is classified into intestinal, diffuse, and rarely mixed type; with the intestinal type having a higher incidence and longer course, often driven by multiple events such as chronic superficial gastritis, chronic atrophic gastritis (CAG), intestinal metaplasia (IM), dysplasia, and ultimately carcinoma [4,5]. Gastric carcinogenesis is thus a slow, progressive and multistep process involving various contributing molecules. To this effect, dynamic changes in gene expressions are important indicators of transformation from precancerous lesions to EGC.

Dynamic gene screening based on precancerous lesions is helpful to identify pivotal biomarkers for early diagnosis and progression assessment of GC. For instance, Zhang et al. constructed single-cell dynamic transcriptome profiles of precancerous and EGC lesions through single-cell sequencing analysis of 13 gastric mucosal samples of 9 patients with non-atrophic gastritis (NAG), CAG, IM, or EGC, which resolved their cellular changes at different disease stages, and identified a set of early tumor cell-specific markers [6]. In the early stages of gastric carcinogenesis, precancerous tissues are infiltrated by various immune cells such as T cells, B cells, macrophages, mast cells, and dendritic cells that secrete various inflammatory cytokines

* Corresponding authors at: Center for Single-Cell Omics and Tumor Liquid Biopsy, Zhongnan Hospital of Wuhan University, Wuhan, China.
** Corresponding author at: Department of Laboratory Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China.
E-mail addresses: rongyuan323@whu.edu.cn (Y. Rong), mengxy_whu@163.com (X. Meng), wfb20042002@sina.com (F. Wang).
[1] Contribute equally.

https://doi.org/10.1016/j.csbj.2022.12.036

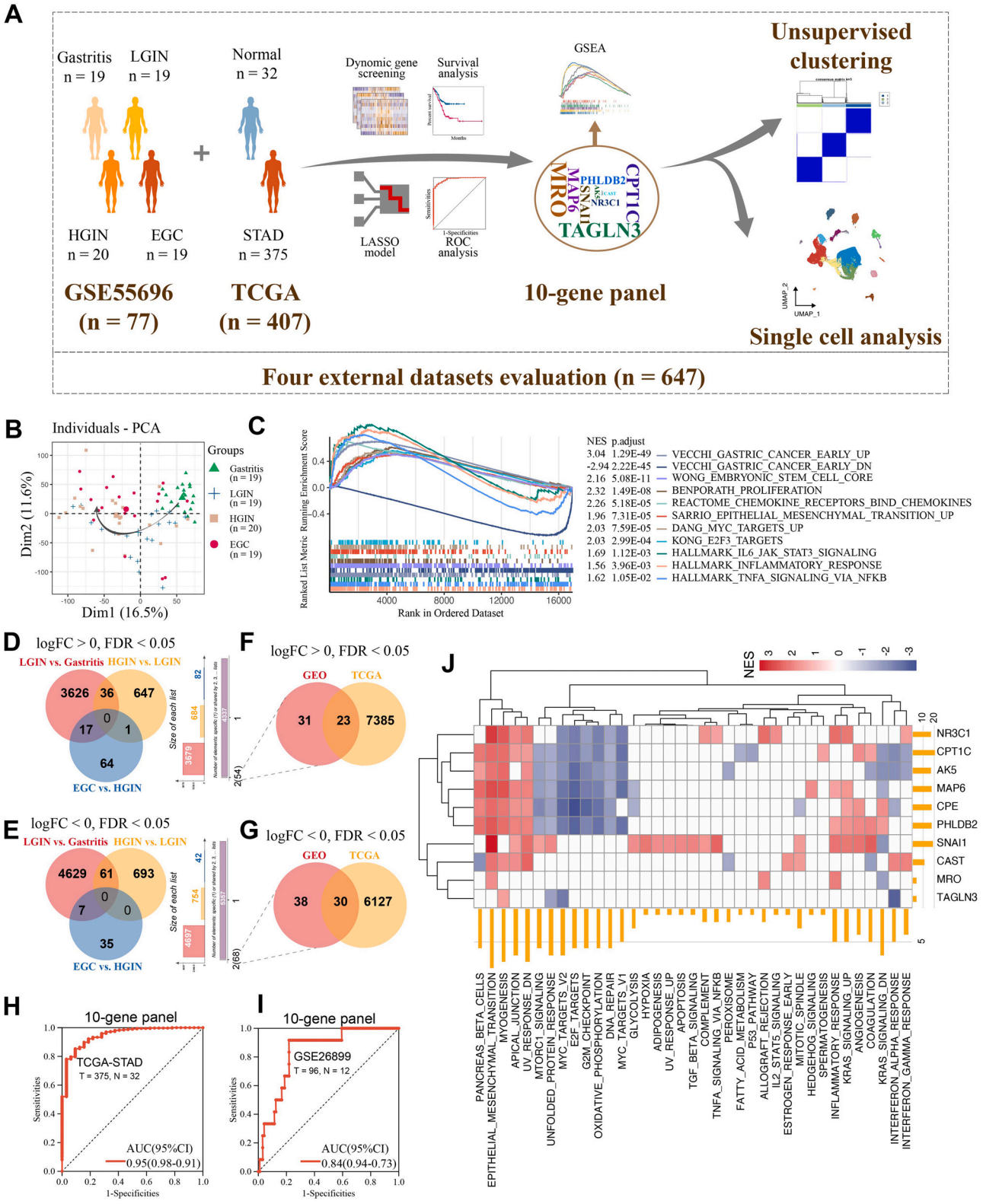

Fig. 1. The construction of 10-gene diagnostic panel. (A) Schematic diagram highlighting the experimental workflow for the whole study. (B) The principal component analysis (PCA) chart shows the progression of the patient from gastritis to low-grade intraepithelial neoplasia (LGIN), high-grade intraepithelial neoplasia (HGIN) and early gastric cancer (EGC). (C) The genes sorted by principal component rotation were significantly enriched in the processes related to early gastric cancer by gene set enrichment analysis (GSEA). (D &E) The Venn diagram shows the intersection of genes obtained from the three groups of difference analyses. (F&G) The Venn diagram shows the genes successfully validated in TCGA. (H&I) Receiver operating characteristic curve (ROC) shows the area under the curve (AUC) of the 10-gene panel in TCGA and GSE26899, with larger AUC indicating better diagnostic effect. (J) The heatmap shows the enrichment level of cancer hallmarks estimated by GSEA for 10 genes. The bottom and right bar plots show the number of genes enriched for each hallmark and the number of hallmarks enriched for each gene, respectively.

resulting into chronic inflammation, immune dysregulation and transformation to gastric cancer [7,8]. In addition to immune cells, epithelial and stromal cells of normal gastric mucosa, including mucus cells, secretory cells, endocrine cells, fibroblasts, and endothelial cells, are often disrupted in the precancerous lesions, further facilitating cancer development and progression [9]. However, the exploration of biomarkers that can indicate transformation of precancerous lesions to EGC is still limited, and how these biomarkers influence tumor microenvironment and GC progression is unclear.

Here, we report a dynamic gene analysis based on the conversion of precancerous tissues to EGC (Fig. 1A). We first performed differential gene expression analyses on samples of precancerous and EGC lesions, combined with expression validation, survival analysis and key gene screening based on the lasso regression model in the TCGA-STAD cohort. A 10-gene diagnostic panel was obtained, and their associated cancer pathways identified through gene set enrichment analysis (GSEA). Three gene clusters with specific cancer mutations, transcriptomic processes and differential tumor microenvironment status were identified based on the 10-gene panel. The gene expression, panel diagnostic performance, and cluster classification stability were estimated in four additional independent datasets. Finally, the expression levels of the genes were characterized by analyzing single-cell RNA sequencing (scRNA-seq) data of precancerous and EGC lesions involving over 40,000 cells. These 10 precancerous lesion-associated genes could be potential new early diagnostic biomarkers for GC, and may aid GC prognostic predictions and personalized therapy.

2. Materials and methods

2.1. Data retrieval

We downloaded six datasets from Gene Expression Omnibus (GEO), including five bulk RNA microarray datasets (GSE55696, GSE60662, GSE26899, GSE63089, and GSE84437), and one scRNA-seq dataset (GSE134520) [10–14]. GSE55696 contained 77 samples [19 cases of gastritis, 19 cases of low-grade intraepithelial neoplasia (LGIN), 20 cases of high-grade intraepithelial neoplasia (HGIN), and 19 cases of EGC]; GSE60662 contained 16 samples (4 normal samples, 4 cases of mild gastritis, 4 cases of severe gastritis and 4 cases of IM); GSE26899 contained 108 samples (96 GC samples and 12 normal samples); GSE63089 contained 90 samples (45 paired GC and normal para-cancer samples); GSE84437 contained 433 GC samples; and GSE134520 contained 13 samples from 9 patients [3 NAG, 3 CAG, 2 wild intestinal metaplasia (IMW), 4 severe intestinal metaplasia (IMS), and 1 EGC]. In addition, we downloaded a bulk RNA-seq count matrix and corresponding clinical data from 407 TCGA-STAD samples (375 tumor samples and 32 normal samples) from the UCSC Xena database (https://xenabrowser.net/datapages/).

2.2. Differential gene analysis, survival analysis and cancer stage analysis

We then performed differential expression analysis of all the genes in the microarray datasets using the limma package, and normalized gene expressions using the normalizeBetweenArrays function [15]. RNA-seq data of TCGA-STAD were analyzed by the edgeR package and normalized by the trimmed mean of m-values (TMM) [16]. The dynamically altered genes of precancerous lesions from GSE55696 and differentially expressed genes of TCGA-STAD were screened for intersecting genes according to their expression patterns and visualized by jvenn [17]. The intersecting genes obtained were further subjected to cancer stage and survival analysis using GEPIA online tools [18].

2.3. Least absolute shrinkage and selection operator (LASSO) and diagnostic analysis

Key precancerous genes were obtained by LASSO regression analysis (default 10-fold cross-validation) of the glmnet package, and the receiver operating characteristic (ROC) curves generated using the pROC package [19] to assess the discriminatory performance of each gene for cancer and normal samples. To evaluate the combined diagnostic performance of multiple genes, binary logistic regression model was constructed for multiple genes by SPSS (Version 25), and their diagnostic performance evaluated using ROC curves.

2.4. Gene set enrichment analysis (GSEA)

We downloaded H: hallmark gene sets from the MSigDB database (https://www.gsea-msigdb.org/) [20] as background gene sets and performed GSEA [21] for each key gene based on TCGA-STAD RNA-seq data by ClusterProfiler package [22]. In brief, the cancer samples were first sorted by the expression levels of each gene, median expression level was taken for sample grouping, and differential gene expression analysis performed by edgeR package. Subsequently, all genes were sorted by logFC from the highest to the lowest as the input gene set, and GSEA performed with H: hallmark gene sets as background gene sets. Significantly enriched entries ($FDR < 0.05$) were obtained and visualized by the pheatmap package.

2.5. Subtype classification of GC based on key genes

Unsupervised clustering [paritioning around medoids (PAM)] was performed on 375 TCGA-STAD samples by ConsensusClusterPlus package [23] based on key genes obtained from LASSO regression screening, generating 1000 resampling to ensure the stability of the classification. Survival analysis and significance assessment were performed based on Kaplan-Meier method and Log-rank test using survival package, and the reliability of clustering verified by tSNE algorithm. The above analysis was repeated on GSE84437 datasets (n = 437).

2.6. Mutation analysis and gene set variation analysis (GSVA)

Somatic mutation data processed and integrated by MuTect2 in TCGA-STAD were downloaded from UCSC Xena for overall mutation burden calculation and gene mutation frequency analysis using maftools package [24]. Overall and progression-free survival analysis based on mutant and wild-type gene was performed in cBioPortal [25].

To further assess the cancer hallmark changes in different patients, GSVA [26] was performed on 375 TCGA-STAD samples based on the H: hallmark background gene sets to obtain the enrichment score for each sample. Subsequently, differential hallmark analysis was performed on the different subtypes by limma package to obtain the specific hallmark in each subtype ($FDR < 0.05$).

2.7. Immune cell infiltration abundance analysis

We first downloaded gene markers of 28 immune cell types as background genes from the study of Charoentong P et al. [27] and estimated immune cell abundance in 375 samples of TCGA-STAD by single sample gene set enrichment analysis (ssGSEA). The differences in immune cell infiltration among subtypes were assessed using Kruskal-Wallis test, and the correlation between immune cell abundance and key precancerous gene expression was assessed by Spearman correlation analysis. Subsequently, we verified the infiltration of immune cells and evaluated the infiltration of immune cell subtypes as well as non-immune cells by the CIBERSORT [28]

and EPIC [29] algorithms. Finally, immune score, stromal score and ESTIMATE score were calculated for each sample using the estimate package [30].

2.8. Single cell data processing and definition of major cell types

Single-cell data from 13 precancerous lesions and EGC samples were preprocessed and subsequently analyzed by the Seurat package [31]. To filter low quality cells, only cells with transcripts > 400 and < 7000 per cell; expressed in more than 3 cells per transcript; percentage of mitochondrial reads < 25 %; percentage of ribosomal reads < 60 % were included in the analysis. Resolution = 0.2 was chosen for initial clustering of epithelial and non-epithelial cells in all cells, followed by further clustering of epithelial or non-epithelial cells by choosing resolution = 1.5 or 0.2, respectively. Major cell types were defined using known cell markers derived from literature or CellMarker database [32].

2.9. Statistical analysis

All statistical analyses for this study were performed using R software (v.4.1.1), and SPSS (Version 25). Comparisons between two groups were made using the Wilcoxon nonparametric rank sum test, multiple group comparisons were made using the Kruskal-Wallis test, and sample composition ratio comparisons between groups were made using the Chi-square test. $P < 0.05$ was considered statistically significant and shown as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. All *P*-value corrections were made by the Benjamin Hochberg method.

3. Results

3.1. Screening for dynamically changed genes in the progression of precancerous lesions

First, we performed a principal component analysis (PCA) of gene expression profiles based on patients with gastritis (n = 19), LGIN (n = 19), HGIN (n = 20) and EGC (n = 19) from GSE55696. This allowed us to observe the dynamic processes of patients' progression from gastritis to EGC (Fig. 1B). GSEA performed based on the PCA rotation score sorted genes (with C2 and H of MSigDB as background gene sets), showed that GASTRIC_CANCER_EARLY_UP was the most significantly enriched entry (*NES* = 3.04, *p.adjust* = 1.29E-49), demonstrating the validity and feasibility of our strategy. Cell proliferation, cell cycle, stemness and immune regulation related entries were also enriched (Fig. 1C). Next, differentially expressed genes (DEGs) analysis was performed for gastritis vs. LGIN (Fig. S1A), LGIN vs. HGIN (Fig. S1B), and HGIN vs. EGC (Fig. S1C), respectively. The gastritis vs. LGIN group (DEGs = 8376, *FDR* < 0.05) had the highest number of DEGs compared to the other two groups (LGIN vs. HGIN, DEGs = 1438; HGIN vs. EGC, DEGs = 124), suggesting that the gastritis to neoplasia transition involves the greatest alterations of gene expression. To further identify dynamically changing genes with similar expression patterns, we divided the DEGs from the three groups into up- and down-regulated groups, and extracted the genes present in each group (Fig. 1D&1E). The extracted sets of genes were then verified by TCGA-STAD data for differential expressions between tumor (n = 373) and normal (n = 32) samples (*FDR* < 0.05), and a total of 23 up-regulated and 30 down-regulated genes were obtained (Fig. 1F & G).

3.2. The screening of survival-related genes

To identify GC prognosis associated genes, the above 50 up-and down-regulated genes were further analyzed in GEPIA. A total of 13 genes significantly associated with survival of GC patients (Log-rank $P < 0.05$) were obtained; 11 were down-regulated (*TAGLN3*, *PHLDB2*, *NR3C1*, *MRO*, *MREG*, *MAP6*, *KIF5C*, *GLDN*, *CPE*, *CAST* and *AK5*; Fig. S2A) and two up-regulated (*SNAI1* and *CPT1C*; Fig. S2B). Interestingly, high expression of all the above genes except for *MREG* were associated with poor prognosis, suggesting that gene expression levels is not an absolute indicator of the effect of genes on cancer progression. Furthermore, we concurrently investigated changes in the expression levels of the above genes from one GC stage to another (Fig. S2C) using GEPIA. *PHLDB2*, *NR3C1*, *MAP6* and *CPE* had the most significant positive correlation with cancer stage [$Pr\ (>F) < 0.05$]; and *MRO*, *GLDN*, *SNAI1* and *CPT1C* had trends of elevated expression with increasing GC staging. Finally, we used box plots to visualize the expression of these genes in the GSE55696 and TCGA-STAD cohorts, most genes had consistent expression patterns in both cohorts (Fig. S2D&S2E). We noted that two genes (*SNAI1* and *CPT1C*) showed stable upregulation in both precancerous lesion and cancer samples in four additional cohorts (GSE130823, GSE60662, GSE26899 and GSE63089) (Fig. S3), suggesting that they could have potentials for clinical application.

3.3. Construction of the 10-gene diagnostic panel

To screen the gene panel with the best diagnostic value, a linear model was constructed based on the 13 earlier identified GC survival-related genes using the LASSO regression analysis with variable screening and regularization (Fig. S4A). The mean-squared error of the model was minimized when $log(\lambda) = -5.22$ (Fig. S4B). 10 genes (*AK5*, *CAST*, *CPE*, *MAP6*, *MRO*, *NR3C1*, *PHLDB2*, *TAGLN3*, *CPT1C*, and *SNAI1*) having the best diagnostic capacity to distinguish GC from normal samples were finally obtained. AUC of the gene combination reached 0.95 (95 % *CI*: 0.91–0.98; accuracy = 86.98 %, sensitivity = 86.98 %, specificity = 90.63 %) in the TCGA-STAD training cohort and 0.84 (95 % *CI*: 0.73–0.94; accuracy = 79.63 %, sensitivity = 91.67 %, specificity = 78.13 %) in the GSE26899 validation cohort. (Fig. 1H&1I, Fig. S4C). Moreover, our gene panel was also compared with reported biomarkers[33,34] and panels[35,36] (Fig. S5). Overall, the diagnostic efficiency of panel was better than individual biomarker in training and validation cohorts. In the validation cohort, compared with the other two panels from Yap et al. and Cui et al., our panel demonstrated higher sensitivity than both of them (our panel vs. Yap et al./Cui et al. = 91.7 % vs. 75.0 %/75.0 %) but lower in specificity than Cui et al. (our panel vs. Yap et al./Cui et al. = 78.1 % vs. 67.7 %/89.6 %). To further determine the diagnostic effectiveness of the 10-gene panel for precancerous lesions and EGC, ROC curve analysis was performed on the GSE55696 datasets (Fig. S6). The results showed that the 10-gene panel best differentiated EGC from gastritis (AUC > 0.90). Its ability to differentiate EGC from other neoplastic stages was rather low (EGC vs. LGIN, AUC = 0.78; EGC vs. HGIN, AUC = 0.83). The Human protein Atlas (HPA) database was concurrently used to determine the protein expression levels of these genes in GC tumor and normal tissues. (https://www.proteinatlas.org/). The expression trends of most of the proteins were consistent with mRNA levels, except for CPT1C and TAGLN3, where no significant protein expression was found (Fig. S7A), which may be related to their protein localization and tissue specificity (Fig. S7B&S7C). In addition to this, the cancer processes that are likely influenced by these genes were explored through GSEA (Fig. 1J). Epithelial mesenchymal transition (EMT) was the most affected entry, suggesting that these genes may play a crucial role in cancer progression and metastasis. Moreover, some cell cycle-related entries such as E2F TARGETS, G2M CHECKPOINT, DNA REPAIR, and immune-related entries such as INFLAMMATORY RESPONSE and INTERFERON ALPHA RESPONSE were also enriched.

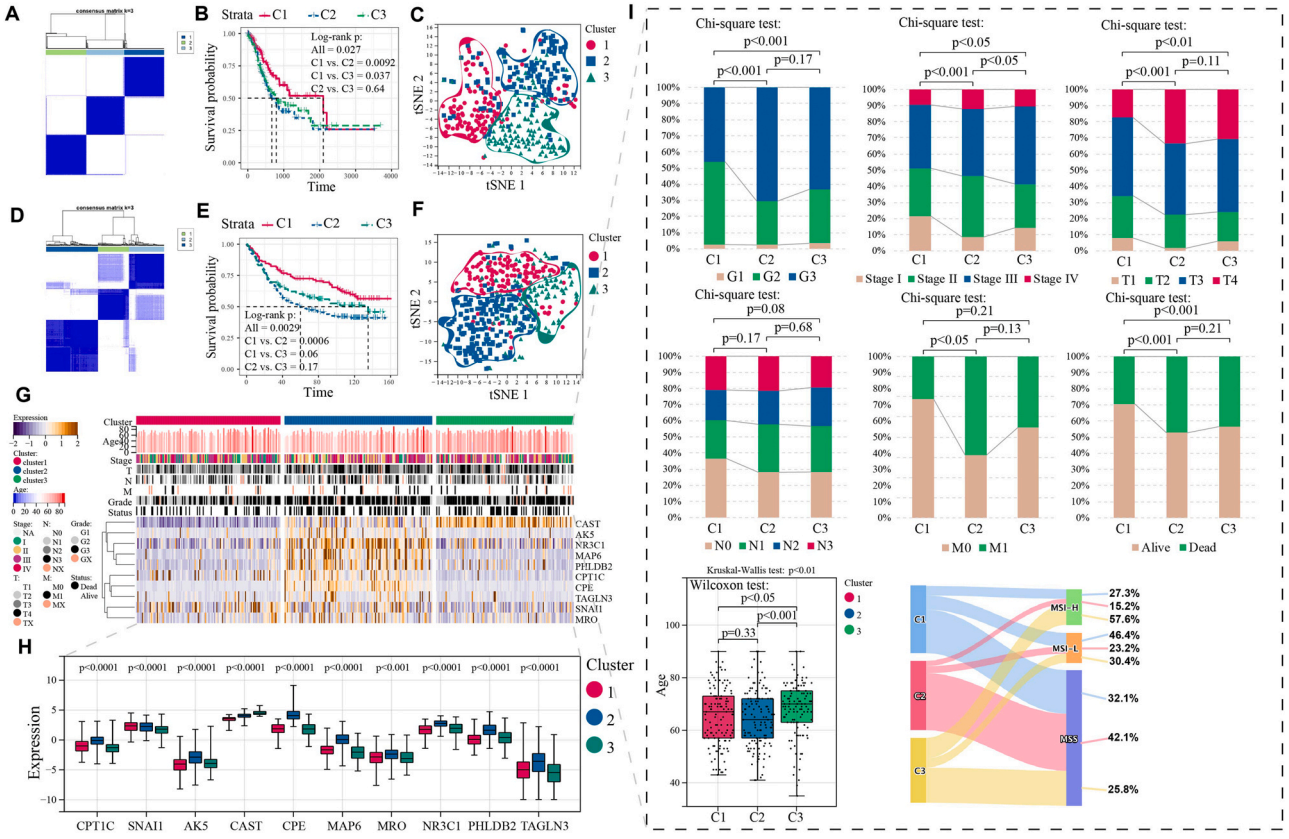

Fig. 2. Unsupervised clustering based on 10-gene panel. (A) The heatmap shows that gastric cancer patients can be distinguished into three subtypes (or clusters) by consensus clustering method in TCGA-STAD (n = 375). (B) Kaplan-Meier curve demonstrates different overall survival for three clusters. (C) The tSNE plots shows that the three clusters can also be distinguished based on the 10-gene panel through tSNE algorithm. (D-F) The results of consensus clustering, survival analysis and tSNE can be replicated in the GSE84437 cohort (n = 433). (G&H) The heatmap and box plot show the expression of 10 genes in different clusters. (I) Proportional bars, box plot, and Sankey diagram show the distribution of patients corresponding to typical clinical features (include grade, stage, T, N, M and status), age, and microsatellite instability (MSI), respectively.

3.4. Potential of 10-gene panel for subtype classification

We then performed unsupervised clustering on 375 TCGA-STAD tumor samples based on the 10-gene panel. The results indicated that all samples could be clearly classified into three clusters (subtypes) (Fig. 2A), with cluster 1 (C1) patients having the best overall survival (OS) while C2 had the worst OS (Fig. 2B). Moreover, the tSNE algorithm was also able to differentiate the three clusters, hence confirming the classification reliability (Fig. 2C). The above results were validated in the GSE84437 cohort (n = 433) (Fig. 2D–F). Next, we investigated the expression levels of panel genes in the three TCGA-STAD clusters. All panel genes exhibited significantly increased expression in C2, while generally decreased expression in C1 (Fig. 2G&H). The independent cohort, GSE84437 demonstrated similar expression pattern (Fig. S8). Finally, we explored the distribution of patients with different clinical characteristics across the three clusters. Compared to C1 and C3, C2 had; 1) a decrease number of stage I patients and an increase in stage II and stage IV; 2) a decrease in T1 and an increase in T4; 3) no significant change in lymph node; 4) an increase in metastatic patients; 5) an increase in G3 patients; 6) an increase in death; 7) and an overall decrease in age (Fig. 2I). These results suggested that C2 had a more malignant GC progression and was not affected by age. We also observed that microsatellite stable (MSS) tumors had the highest percentage of C2 patients (42.1 %), while C3 (57.6 %) and C1 (46.4 %) patients dominated in tumors with high and low microsatellite instability (MSI), respectively (Fig. 2I). MSI-type tumors usually predict better patient survival, and detection of MSI status can provide patients with personalized options for targeted therapies and immunotherapy [37]. Concurrently, in comparison with the classical gastrointestinal tumor subtypes [38], the genome stable (GS) type contained more C2 patients, while the MSI type contained more C1 and C3 patients, which is consistent with the previous results (Fig. S9).

3.5. Mutation in the subtypes

We further examined the overall tumor mutation burden (TMB) in the three clusters. C3 had the highest TMB, while C2 had a significantly lower TMB compared to C1 and C3 ($P < 0.0001$) (Fig. 3A). Mutation frequency analysis of the genes in each cluster revealed that *TP53* was the most frequently mutated gene in C2, while *TTN* was the most frequently mutated gene in C1 and C3 (Fig. 3B–D). Meanwhile, the mutation frequencies of the top3 mutated genes *TTN*, *TP53*, and *MUC16* were all decreased in C2, and patients with *TP53* and *MUC16* mutations had better OS (Log-rank $P < 0.05$), suggesting that the reduced TMB and mutation frequencies of *TP53* and *MUC16* may lead worse prognosis in C2 patients (Fig. 3E). Chi-square test was then used to obtain the mutated genes with the most significant differences among the three clusters (Fig. 3F). Among the top10 differentially mutated genes, *TTN*, *CSMD3*, *RYR1*, *PLEC*, and *ANK3* mutant samples were all associated with better progression-free survival (PFS). These genes may thus serve as potential drug targets for suppressing GC progression (Fig. 3G–K).

3.6. Altered cancer hallmark and immune microenvironment in different subtypes

To explore the unique activation of cancer processes in the above three clusters, we calculated the enrichment score for TCGA-STAD samples by GSVA based on 50 cancer hallmarks collected from

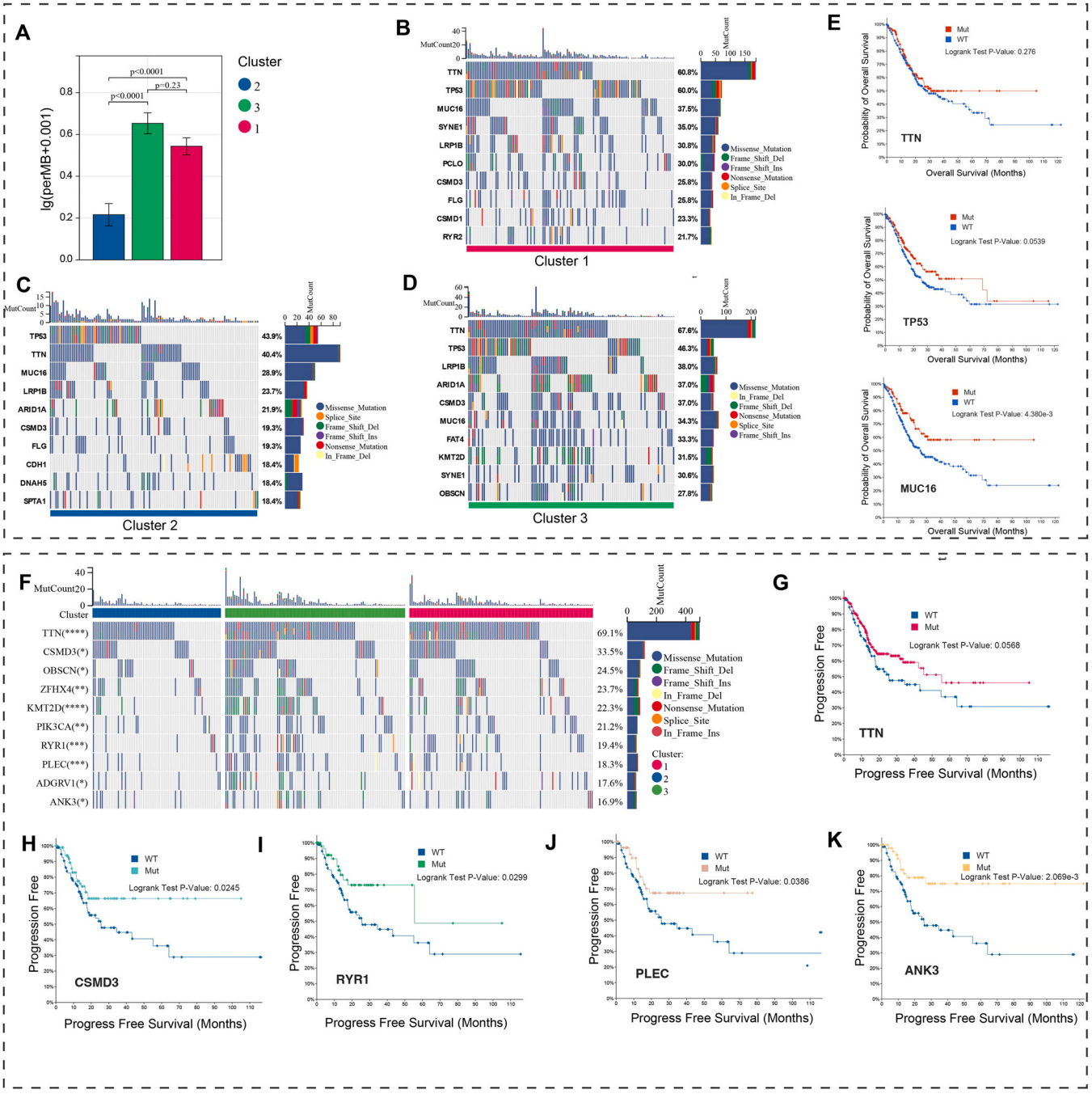

Fig. 3. Mutation characteristics of patients with different subtypes. (A) Bar plot demonstrates the tumor mutation burden (TMB) among three clusters in TCGA-STAD. (B–D) Oncoplots show the mutation landscape of mutant genes with high mutation frequency in each cluster. (E) Kaplan-Meier curves show the overall survival of mutant and wild-type patients with the three most frequently mutated genes (*TTN*, *TP53* and *MUC16*). (F) The oncoplot shows the mutated genes with significantly different mutation frequencies among the three clusters by chi-square test. (G–K) Kaplan-Meier curves show that 5 of the top 10 differentially mutated genes were associated with progression-free survival (PFS) of patients.

MSigDB (Fig. S10A). C1 and C3 shared some cell cycle-related hallmarks such as MYC TARGETS, E2F TARGETS, and G2M CHECKPOINT. C2 showed a high enrichment of tumor progression, metastasis, and immune-related hallmarks such as EPITHELIAL MESENCHYMAL TRANSITION, MYOGENESIS, ANGIOGENESIS, KRAS SIGNALING, and INFLAMMATORY RESPONSE. Immune-related entries such as INTERFERON ALPHA & GAMMA RESPONSE and PROTEIN SECRETION were significantly enriched in C3. To gain insight into the differences in the tumor immune microenvironment among the clusters, immune cell infiltration of each patient was assessed by ssGSEA based on the reported marker of 28 immune cells. Overall, all samples could be clustered into two groups, high immune level and low immune level group, with more C2 and C3 samples distributed in the high immune group and predominantly C1 in the low immune group (Fig. S10B). Moreover, compared to C1 and C3, C2 had the highest ssGSEA scores of most immune cells, such as activated B cells, activated CD8 + T cells, macrophages, mast cells, and natural killer cells (Fig. 4A, Fig. S10C); while decreased infiltration of activated CD4 + T cells, neutrophil, type 17 T helper cells, and type 2 T helper cells. In addition, we also investigated the correlation between the genes in the earlier identified 10-gene panel and these immune cells. The results showed significant positive correlations of 10 genes with most of the

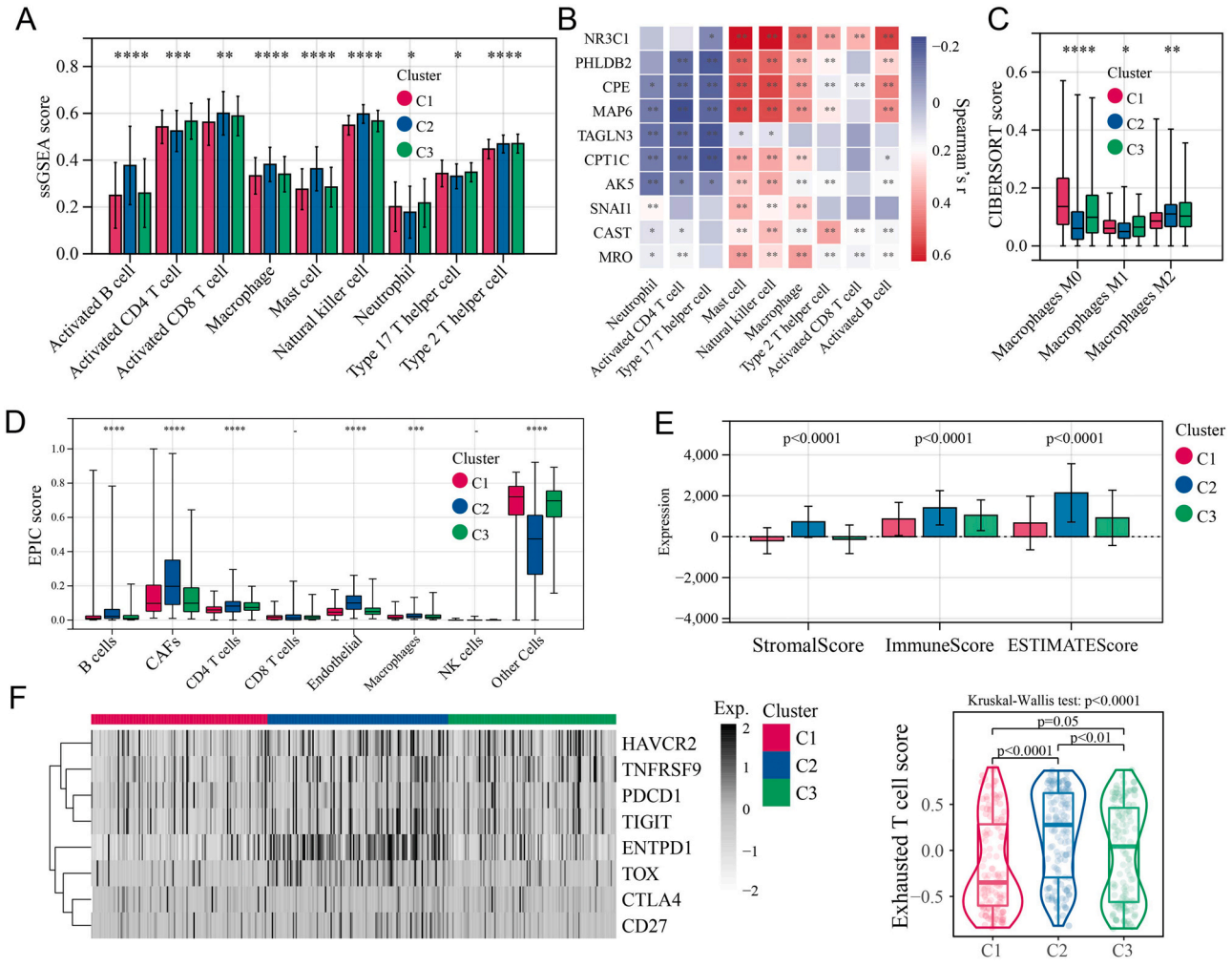

Fig. 4. Increased immune cell infiltration in C2. (A) The standard error bar graph shows that the infiltration score (calculated by single-sample gene set enrichment analysis) of some immune cells was highest in C2 or C3. (B) Heatmap shows the Spearman correlation of 10 key genes with immune cell infiltration score. (C) Box plot shows the infiltration of macrophage subtype immune estimated by CIBERSORT algorithm in three clusters. (D) Box plot shows the infiltration of immune cells and non-immune cells estimated by EPIC algorithm in three clusters. (E) The standard error bar graph shows the stromal score, immune score and ESTIMATE score (usually reflecting tumor purity) for the three cluster samples calculated by the ESTIMATE algorithm. (F) The heatmap shows the scaled expression (FPKM value) of exhausted T cell markers in the three clusters. Violin combination plot shows the distribution of exhausted T cell score in three clusters as calculated by GSVA algorithm.

immune cells (Spearman analysis, $P < 0.05$) and negative correlations with activated CD4 + T cells, neutrophil and Th17 (Fig. 4B, Fig. S10D). These findings provide a molecular basis for classifying subtypes based on the nature of the infiltrated immune cells. To validate the results of the ssGSEA, we also performed cell abundance estimation using the CIBERSORT and EPIC algorithms based on the built-in immune cell maker, where EPIC also provided a partial maker of non-immune cells. Results showed that C2 still exhibited a high infiltration by major immune cell types, but there were differences in some immune subpopulations, for example, macrophages M0 and M1 were reduced in C2, but increased in M2 (Fig. 4C, Fig. S10E). The pro-tumor immune infiltration in C2 may thus favor tumor growth. C2 also had a significant increase in cancer-associated fibroblasts (CAFs), showing a strong immune exclusion phenotype [39] (Fig. 4D). Also, we evaluated the overall immune score, stromal score, and ESTIMATE score for the three clusters, and C2 still had the highest score (Fig. 4E). Moreover, we observed increased expression of exhausted T cell markers (*HAVCR2*, *TNFRSF9*, *PDCD1*, *TIGIT*, *ENTPD1*, *TOX*, *CTLA4*, and *CD27*) [40], and increased exhausted T cell score in C2 (Fig. 4F&G). Therefore, we hypothesize that these precancerous genes may be involved in the formation of a pro-tumor microenvironment in GC patients and are associated with immune exclusion/exhausted subtype shaping.

3.7. Single-cell localization of panel genes

We further investigated the expression of the 10 panel genes in scRNA-seq data containing 13 samples of precancerous lesions and EGC. First, the downloaded scRNA-seq data were cleaned to remove low-quality cells (see Materials and Methods, Fig. S11A). A total of 41,471 cells were then obtained, of which 5418 were derived from 3 NAG samples, 18,765 from 3 CAG samples, 3597 from 2 IMW samples, 10,847 from 4 IMS samples, and 2844 from 1 EGC sample (Fig. S11B). Subsequently, we examined the distribution of cells grouped by disease, sample and patient origin to ensure that there was no significant batch effect (Fig. S11C-S11E). Although patient-grouped cells showed some specific aggregation, this aggregation was correlated with disease type, where the specific aggregation likely represented biological differences, rather than batch effect. Next, we evaluated the expression changes of the 10 genes at the single-cell transcriptome level in precancerous lesions. *SNAI1* gradually increased during the transition from NAG to EGC and reached the highest expression in EGC; *CAST* showed significant upregulation mainly at the IMS stage and was also highly expressed in EGC; *CPE* and *NR3C1* were upregulated at the initial CAG stage and then decreased, they both demonstrated the lowest expression in EGC (Fig. 5A). The expression of the remaining six genes were relatively

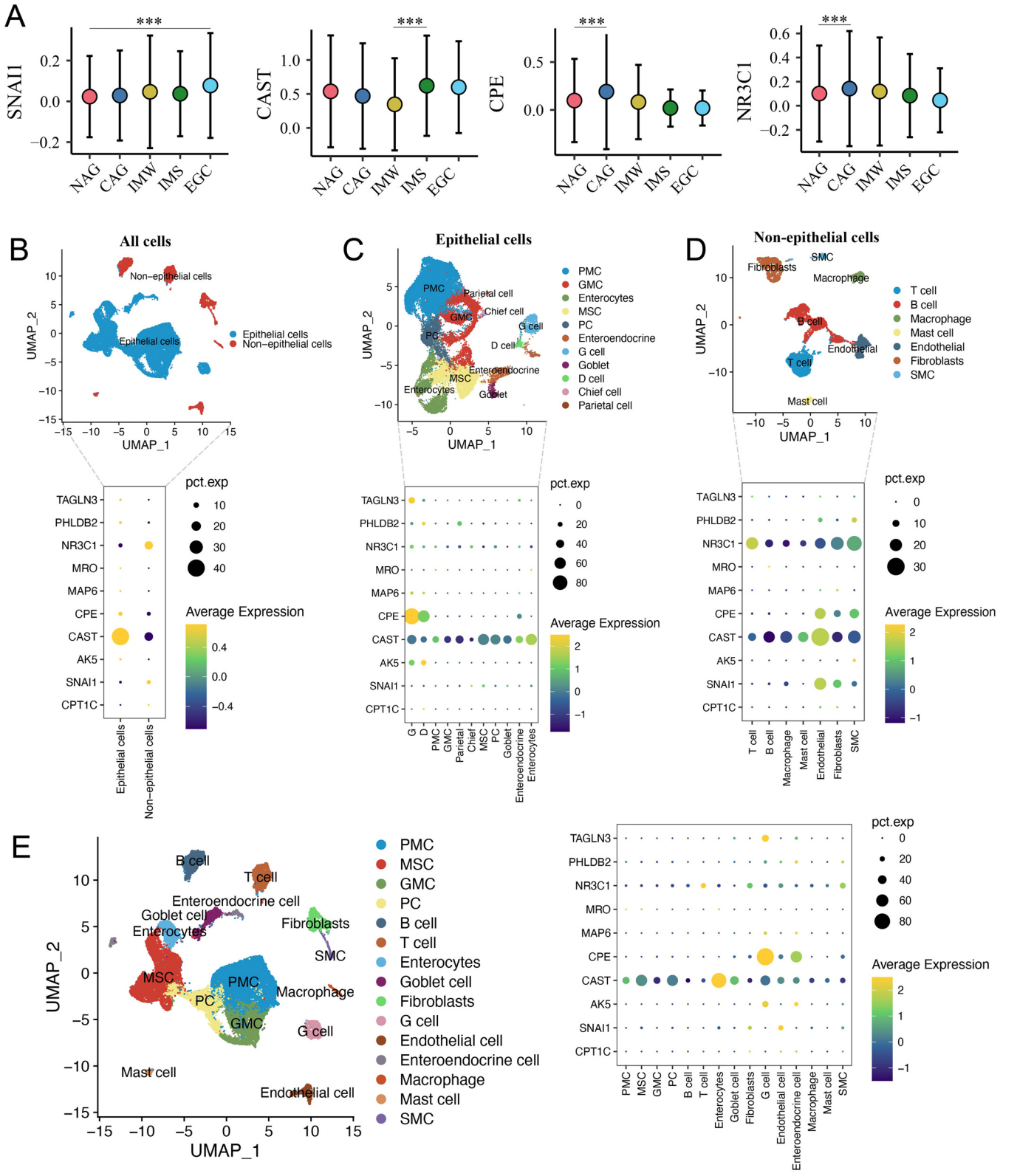

Fig. 5. Single-cell expression localization of key genes. (A) The mean dot plot shows four genes with typical expression changes during gastric precancerous lesion progression based on single-cell RNA sequencing data (GSE134520). (B–D) The UMAP plot (upper) and dot plot (lower) show the cell clusters and key genes expression in all cells, epithelial cells and non-epithelial cells. (E) The UMAP plot (left) and dot plot (right) show the cell clusters and key genes expression in all annotated cells.

low (*AK5*, *MAP6*) in EGC or stable (*CPT1C*, *MRO*, *PHLDB2*, *TAGLN3*) from NAG to EGC (Fig. S12). This indicated that gene expression in scRNA-seq was different from that in bulk RNA-seq in some degree.

We then annotated the cell types and classified all cells into epithelial cells (*EPCAM*, *KRT18*) and non-epithelial cells (*VIM*, *PTPRC*), Here, *CAST* was significantly expressed in the epithelial cells, while *NR3C1* was expressed in non-epithelial cells (Fig. 5B, Fig. S13A).

Epithelial cells were further classified as G cells (*GAST*), D cells (*SST*), pit mucous cells (PMC) (*GKN1*, *GKN2*), gland mucous cells (GMC) (*MUC6*), parietal cells (*ATP4A*, *ATP4B*, *GIF*), chief cells (*PGA3*, *PGA4*, *LIPF*), metaplastic stem-like cells (MSC) (*OLFM4*, *EPHB2*), proliferative cells (PC) (*MKI67*, *TOP2A*, *BIRC5*), goblet cells (*TFF3*, *SPINK4*, *MUC2*), enteroendocrine cells (*CHGA*, *TPH1*), and enterocytes (*FABP1*, *CA1*, *VIL1*) (Fig. 5C, Fig. S13B). Among the panel genes, *CPE* was significantly expressed in gastric secretory G and D cells, which may be involved in the process of gastritis; *CAST* showed high expression in various cells, but was highest in enterocytes, and associated with *H. pylori* infection, suggesting that it may affect *H. pylori*-induced intestinal epithelial metaplasia (Fig. 5C, Fig. S14). Non-epithelial cells were further classified as T cells (*CD3D*), B cells (*CD79A*), macrophages (*CD14*, *CD68*, *CSF1R*), mast cells (*TPSAB1*), endothelial cells (*VWF*, *ENG*), fibroblasts (*PDPN*, *COL1A2*, *DCN*, *COL3A1*, *COL6A1*), and smooth muscle cells (SMC) (*ACTA2*) (Fig. 5D, Fig. S13C). Gene *NR3C1* was significantly expressed in T cells, and showed the strongest correlation with immune cell infiltration among all the panel genes based on bulk RNA-seq data analysis (Fig. 5D, Fig. 4D), hinting that it may have an important role in immune system regulation. Moreover, *SNAI1* was significantly expressed in endothelial cells and fibroblasts, suggesting a potential *SNAI1*-dependent pro-tumor mechanism in fibroblasts and endothelial cells [41]. *CAST* was the most widely expressed gene in various cell comparisons, suggesting that it may have an important contribution to the shaping of the gastric precancerous microenvironment (Fig. 5E). Finally, the dynamic changes in the different cell types during progression from NAG to EGC was investigated. MSC and T cells were significantly increased at the EGC stage, suggesting elevated cell stemness and enhanced regulation of the T cell system in the early stages of gastric carcinogenesis (Fig. S15). In conclusion, 4 of the 10 precancerous lesion-related genes had some changes in expression, and these changes were specific to the different precancerous stages and cell types, and therefore could be associated with the development of specific cell-mediated precancerous lesions.

4. Discussion

The use of a panel of differentially expressed genes to diagnose early stage cancer is a research hot spot. Although a number of gene panel biomarkers have been developed for GC, few studies have focused on the dynamic changes that occur in genes as precancerous lesions transform to GC, and how they influence GC progression. In this study, we based on the dynamic gene expression profiles of precancerous lesions and EGC, to identify a 10-gene diagnostic panel, which in addition to exhibiting stepwise changes in expression from precancerous lesions to EGC, also influence cancer progression and patient survival. Through unsupervised clustering, this panel was able to divide GC patients into different subtypes and demonstrated significant relationship with OS. Furthermore, these subtypes had distinct mutational profiles, and showed transcriptome-based pathway alterations during cancer progression (especially dysregulation of the immune microenvironment). Analysis of scRNA-seq data revealed that these genes had significantly different expression profiles at the single-cell level compared to bulk.

Most studies consider genes that are upregulated in tumor tissues to be oncogenic, however, this assertion is not always true as mRNA in bulk samples come from various sources and cancer development is a complex process. Eight of the 10 precancerous lesion-associated genes in this study were down-regulated in GC, but their upregulation was associated with worse patient survival, moreover, most of them showed increase in expression as the tumor stage increased, indicating that they probably possess tumor promoting properties in GC. Additionally, we found that all the genes, except *TAGLN3*, were significantly enriched in the EMT process, further confirming their association to GC progression and metastasis. Subsequently, through unsupervised clustering, the 10 genes distinguished the GC patients into three clusters, where C2 had a generally high expression of the genes, showed more malignant clinical features and was associated with the worst overall patient survival. In the analysis of subtype mutation profiles, the top mutated genes *TTN*, *TP53* and *MUC16* all showed decreased mutation frequency in C2, a result consistent with reports from other studies stating that mutations in these genes is predictors of good survival in GC patients [42–44]. Moreover, we found a significant association between some of the differentially mutated genes and PFS of patients, a further proof that these genes are closely associated with GC progression. In the transcriptomic profiling of subtypes, C2 was enriched in more malignant tumor processes and showed increased immune cell infiltration as well as typical features of a pro-tumor microenvironment, which may be associated with a reduced frequency of *TP53* mutations. Mutations in *TP53* have been found to be associated with lower immune activity in GC patients and may result in a worse response to immunotherapy [45]. The 10-gene-based subtype classification system may thus be indicative of the patient's response to immunotherapy.

Finally, in an attempt to explore the association between the panel genes and cell types, we re-evaluated their expression levels in scRNA-seq data, and found the expression patterns significantly different from those of the bulk RNA-seq data. Snail family transcriptional repressor 1 (*SNAI1*), a widely reported zinc-finger transcription factor, and a major regulator of EMT in various cancers [46], was highly expressed in endothelial cells and fibroblasts. Just a few studies have reported the association between *SNAI1* and precancerous lesions, therefore, our findings suggest that *SNAI1* could be a key promoter of the transition from gastritis to EGC, hence is a potential target gene for early diagnosis of GC. In addition, *SNAI1*'s high expression in endothelial cells and fibroblasts could promote the exclusion of immune cells, shape the immune exclusion microenvironment and cause immune escape of tumors [47,48]. *CAST* was the most widely expressed gene across all cell types. *CAST* encodes a calpain inhibitor calpastatin, which is normally found at the plasma membrane and around the nucleus, where it can be translocated into the nucleus to regulate the WNT/β-catenin pathway [49,50]. Yang et al. revealed that *CAST* is a potential oncogene in GC, where its high expression level is associated with poor OS in GC patients, and linked to macrophage infiltration [51]. In this study, *CAST* is expressed in multiple cell types especially enterocytes and endothelial cells and may regulate cellular behavior towards promoting the formation of a pro-tumor microenvironment. We also found that *CAST* was particularly expressed in IMS patients, which may have an important contribution to intestinal metaplasia of the stomach. *CPE* was mainly expressed in the gastric secretory G and D cells. Naturally, *CPE* is involved in the synthesis of gastrin, secreted by G cells. An association between gastrin and cancer has been reported and gastrin-related genes have been found to be frequently upregulated in cancer [52]. Studies in several in vitro models have shown that gastrin can activate cancer-related signaling pathways such as MAPK, PI3K/Akt, and EGFR signaling pathways via CCK2R [53–55]. The *NR3C1* gene meanwhile was highly expressed in T cells. *NR3C1* encodes the glucocorticoid receptor (GR), which is involved in the inflammatory response, cell proliferation and differentiation of target tissues. Studies show that GR may have dual (pro- or anti-tumor) effects on some tumors, such as breast and prostate cancers, but its role in GC is unclear [56]. We found that both mRNA and protein expression of *NR3C1* were lower in GC samples than in normal samples, but its high mRNA expression tended to predict poorer OS and the expression increased with cancer staging. In single-cell expression assessment, significant elevations of NR3C1 were present at the CAG stage. These results suggest that the role of *NR3C1* in GC may be diverse and correlate with GC stages, hence

deeper understanding of its specific roles in each GC stage could facilitate its application in clinical practice.

Our study had the following limitations: First, there is heterogeneity in the expression and function of these precancerous lesion-associated genes both at the bulk and single-cell levels, as well as in pre-cancer development and cancer progression. To fully explain these seemingly contradictory results requires larger samples and further experimental validation. Second, gastric carcinogenesis does not always evolve according to a stereotypical process and varies among the intestinal, diffuse and mixed types of Lauren's classification. Considering that we only screened genes in the three stages of inflammation, neoplasia and cancer, skipping some precancerous stages (e.g. CAG and IM), certain genes with important impact on gastric carcinogenesis, that are found in these skipped precancerous stages could have been missed [57]. In future studies, a more rational screening strategy needs to be designed for different evolutionary types of GC. Furthermore, the specific expression of these markers in the blood of patients with precancerous lesions and specific stages of GC is an important criterion for their use in clinical practice.

5. Conclusions

In summary, our study provides a reference for early diagnostic marker discovery based on dynamic gene screening from precancerous lesions to EGC, and the identified 10-gene panel is expected to aid in the early diagnosis, progression assessment and personalized treatment of GC. Moreover, these genes are expected to be used for further cellular heterogeneity and mechanistic exploration.

CRediT authorship contribution statement

Fei Long: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Shuo Li:** Conceptualization, Investigation, Writing – review & editing. **Yaqi Xu:** Data curation, Writing – review & editing. **Min Liu:** Writing – review & editing. **Xuan Zhang:** Writing – review & editing. **Junting Zhou:** Writing – review & editing. **Yiyi Chen:** Writing – review & editing. **Yuan Rong:** Writing – review & editing, Project administration. **Xiangyu Meng:** Methodology, Writing – review & editing. **Fubing Wang:** Writing – review & editing, Supervision, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the research fund from Medical Sci-Tech Innovation Platform of Zhongnan Hospital (No. PTXM2021001) and the Fundamental Research Funds for the Central Universities (No. 2042021kf0227). This work was also funded by Medical Top-talented Youth Development Project of Hubei Province and the Health Commission of Hubei Province Scientific Research Project (WJ2021M172).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2022.12.036.

References

[1] Rao H-L, Chen J-W, Li M, Xiao Y-B, Fu J, Zeng Y-X, et al. Increased intratumoral neutrophil in colorectal carcinomas correlates closely with malignant phenotype and predicts patients' adverse prognosis. PLoS One 2012;7(1):e30806.
[2] Yuasa N, Nimura Y. Survival after surgical treatment of early gastric cancer, surgical techniques, and long-term survival. Langenbeck's Arch Surg 2005;390(4):286–93.
[3] Park J-M, Ryu W-S, Kim J-H, Park S-S, Kim S-J, Kim C-S, et al. Prognostic factors for advanced gastric cancer: stage-stratified analysis of patients who underwent curative resection. Cancer Res Treatment: Off J Korean Cancer Assoc 2006;38(1):13–8.
[4] Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma: an attempt at a histo-clinical classification. Acta Pathol Microbiol Scand 1965;64(1):31–49.
[5] Correa P, Haenszel W, Cuello C, Tannenbaum S, Archer M. A model for gastric cancer epidemiology. The Lancet 1975;306(7924):58–60.
[6] Zhang P, Yang M, Zhang Y, Xiao S, Lai X, Tan A, et al. Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer. Cell Rep 2019;27(6):1934–47. e5.
[7] Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? The Lancet 2001;357(9255):539–45.
[8] Coussens LM, Werb Z. Inflammation and cancer. Nature 2002;420(6917):860–7.
[9] Kim J, Park C, Kim KH, Kim EH, Kim H, Woo JK, et al. Single-cell analysis of gastric pre-cancerous and cancer lesions reveals cell lineage diversity and intratumoral heterogeneity. NPJ Prec Oncol 2022;6(1):1–11.
[10] Xu X, Feng L, Liu Y, Zhou W-X, Ma Y-C, Fei G-J, et al. Differential gene expression profiling of gastric intraepithelial neoplasia and early-stage adenocarcinoma. World J Gastroenterology: WJG 2014;20(47):17883.
[11] Hanada K, Uchida T, Tsukamoto Y, Watada M, Yamaguchi N, Yamamoto K, et al. Helicobacter pylori infection introduces DNA double-strand breaks in host cells. Infect Immun 2014;82(10):4182–9.
[12] Oh SC, Sohn BH, Cheong J-H, Kim S-B, Lee JE, Park KC, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. Nat Commun 2018;9(1):1–14.
[13] Zhang X, Ni Z, Duan Z, Xin Z, Wang H, Tan J, et al. Overexpression of E2F mRNAs associated with gastric cancer progression identified by the transcription factor and miRNA co-regulatory network analysis. PLoS One 2015;10(2):e0116979.
[14] Yoon S-J, Park J, Shin Y, Choi Y, Park SW, Kang S-G, et al. Deconvolution of diffuse gastric cancer and the suppression of CD34 on the BALB/c nude mice model. BMC Cancer 2020;20(1):1–10.
[15] Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor. Springer; 2005. p. 397–420.
[16] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinform 2010;26(1):139–40.
[17] Bardou P, Mariette J, Escudié F, Djemiel C, Klopp CJBb. jvenn: an interactive Venn diagram viewer, 15(1); 2014, p. 1–7.
[18] Tang Z., Li C., Kang B., Gao G., Li C., Zhang Z.J.Nar GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. 45(W1); 2017. p. W98–W102.
[19] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform 2011;12(1):1–8.
[20] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinform 2011;27(12):1739–40.
[21] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102(43):15545–50.
[22] Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innov 2021;2(3):100141.
[23] Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinform 2010;26(12):1572–3.
[24] Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28(11):1747–56.
[25] Gao J, Aksoy B, Dogrusoz U, Dresdner G, Gross B, Sumer S, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6(269):pl1.
[26] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinform 2013;14(1):1–15.
[27] Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep 2017;18(1):248–62.
[28] Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Cancer Syst Biol: Springer 2018:243–59.
[29] Racle J, Gfeller D. EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data. Bioinform Cancer Immunotherapy: Springer 2020:233–48.

[30] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4(1):1–11.
[31] Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36(5). 411-20.
[32] Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res 2019;47(D1):D721–8.
[33] Necula L, Matei L, Dragu D, Neagu AI, Mambet C, Nedeianu S, et al. Recent advances in gastric cancer early diagnosis. World J Gastroenterol 2019;25(17):2029.
[34] Yasui W, Oue N, Sentani K, Sakamoto N, Motoshita J. Transcriptome dissection of gastric cancer: identification of novel diagnostic and therapeutic targets from pathology specimens. Pathol Int 2009;59(3):121–36.
[35] Yap YL, Zhang XW, Smith D, Soong R, Hill J. Molecular gene expression signature patterns for gastric cancer diagnosis. Comput Biol Chem 2007;31(4):275–87.
[36] Cui J, Chen Y, Chou W-C, Sun L, Chen L, Suo J, et al. An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer. Nucleic Acids Res 2011;39(4):1197–207.
[37] Puliga E, Corso S, Pietrantonio F, Giordano S. Microsatellite instability in Gastric Cancer: Between lights and shadows. Cancer Treat Rev 2021;95:102175.
[38] Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. Cancer Cell 2018;33(4):721–35. e8.
[39] Joyce JA, Fearon DT. T cell exclusion, immune privilege, and the tumor microenvironment. Science 2015;348(6230):74–80.
[40] Chen D, Zhang Y, Wang W, Chen H, Ling T, Yang R, et al. Identification and characterization of robust hepatocellular carcinoma prognostic subtypes based on an integrative metabolite-protein interaction network. Adv Sci 2021;8(17):2100311.
[41] Herrera A, Herrera M, Guerra-Perez N, Galindo-Pumariño C, Larriba MJ, García-Barberán V, et al. Endothelial cell activation on 3D-matrices derived from PDGF-BB-stimulated fibroblasts is mediated by Snail1. Oncogen 2018;7(9):1–15.
[42] Li X, Pasche B, Zhang W, Chen K. Association of MUC16 mutation with tumor mutation load and outcomes in patients with gastric cancer. JAMA Oncol 2018;4(12):1691–8.
[43] Yang Y, Zhang J, Chen Y, Xu R, Zhao Q, Guo W. MUC4, MUC16, and TTN genes mutation correlated with prognosis, and predicted tumor mutation burden and immunotherapy efficacy in gastric cancer and pan-cancer. Clin Transl Med 2020;10:4.
[44] Tahara T, Shibata T, Okamoto Y, Yamazaki J, Kawamura T, Horiguchi N, et al. Mutation spectrum of TP53 gene predicts clinicopathological features and survival of gastric cancer. Oncotarget 2016;7(27):42252.
[45] Jiang Z, Liu Z, Li M, Chen C, Wang X. Immunogenomics analysis reveals that TP53 mutations inhibit tumor immunity in gastric cancer. Transl Oncol 2018;11(5):1171–87.
[46] Dong B, Wu Y. Epigenetic regulation and post-translational modifications of SNAI1 in cancer metastasis. Int J Mol Sci 2021;22(20):11062.
[47] Klein D. The tumor vascular endothelium as decision maker in cancer therapy. Front Oncol 2018;8:367.
[48] Mhaidly R, Mechta-Grigoriou F. Fibroblast heterogeneity in tumor micro-environment: role in immunosuppression and new therapies. Seminars in immunology. Elsevier; 2020.
[49] Burchett R, Brun M, Monckton EA, Poon H-Y, Godbout R. Effects of nuclear factor I phosphorylation on calpastatin (CAST) gene variant expression and subcellular distribution in malignant glioma cells. J Biol Chem 2019;294(4):1173–88.
[50] Benetti R, Copetti T, Dell'Orso S, Melloni E, Brancolini C, Monte M, et al. The calpain system is involved in the constitutive regulation of β-catenin signaling functions. J Biol Chem 2005;280(23):22070–80.
[51] Yang K-T, Yen C-C, Chang R, Wang J-T, Chen J-S. CAST as a potential oncogene, identified by machine search, in gastric cancer infiltrated with macrophages and associated with Lgr5. Biomol 2022;12(5):670.
[52] Ferrand A, Wang TC. Gastrin and cancer: a review. Cancer Lett 2006;238(1):15–29.
[53] Dehez S, Daulhac L, Kowalski-Chauvel A, Fourmy D, Pradayrol L, Seva C. Gastrin-induced DNA synthesis requires p38-MAPK activation via PKC/Ca2+ and Src-dependent mechanisms. FEBS Lett 2001;496(1):25–30.
[54] Ferrand A, Kowalski-Chauvel A, Bertrand C, Pradayrol L, Fourmy D, Dufresne M, et al. Involvement of JAK2 upstream of the PI 3-kinase in cell–cell adhesion regulation by gastrin. Exp Cell Res 2004;301(2):128–38.
[55] Miyazaki Y, Shinomura Y, Tsutsui S, Zushi S, Higashimoto Y, Kanayama S, et al. Gastrin induces heparin-binding epidermal growth factor–like growth factor in rat gastric epithelial cells transfected with gastrin receptor. Gastroenterol 1999;116(1):78–89.
[56] Mayayo-Peralta I, Zwart W, Prekovic S. Duality of glucocorticoid action in cancer: tumor-suppressor or oncogene? Endocrine-related Cancer 2021;28(6):R157–71.
[57] de Vries AC, van Grieken NC, Looman CW, Casparie MK, de Vries E, Meijer GA, et al. Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. Gastroenterol 2008;134(4):945–52.