World Journal of
Surgical Oncology

RESEARCH

Open Access

H19 may regulate the immune cell infiltration in carcinogenesis of gastric cancer through miR-378a-5p/SERPINH1 signaling

Jianxin Li[†], Ting Han[†], Xin Wang, Yinchun Wang, Xuan Chen, Wangsheng Chen and Qingqiang Yang[*]

Abstract

Background: Increasing studies have indicated that noncoding RNA (ncRNA)-mediated competing endogenous RNA (ceRNA) network serves as a significant role in cancer progression, but the underlying regulatory mechanisms of which in gastric cancer (GC) remain largely unclear.

Methods: Based on Gene Expression Omnibus and The Cancer Genome Atlas datasets, potential biomarkers for GC were screened and validated by machine learning. Then, upstream regulatory ncRNA of potential biomarkers was identified to construct a novel ceRNA network in GC through means of stepwise reverse prediction and validation. Ultimately, tumor immune cell infiltration analysis was performed based on the EPIC algorithm.

Results: A total of 188 differentially expressed genes (DEGs) were screened, and three candidate diagnostic biomarkers (FAP, PSAPL1, and SERPINH1) for GC were identified and validated. Subsequently, H19 and miR-378a-5p were identified as upstream regulatory ncRNAs that could potentially bind SERPINH1 in GC. Moreover, Immune infiltration analysis revealed that each component in the ceRNA network (H19/miR-378a-5p/SERPINH1) was significantly correlated with the infiltration abundances of diverse tumor-infiltrating immune cells.

Conclusions: H19 may regulate the immune cell infiltration in carcinogenesis of GC through miR-378a-5p/SERPINH1 signaling.

Keywords: Gastric cancer, Noncoding RNA, Competing endogenous RNA, Biomarker, Immune cell infiltration

Introduction

Gastric cancer (GC) is one of the most commonly diagnosed malignant tumors in the digestive system, which accounts for the third leading cause of cancer-related deaths despite its worldwide decline in incidence and mortality over the past five decades [1, 2]. Although numerous efforts have been afforded to determine the pathogenesis of GC and substantial improvement in

diagnosis and therapy has been achieved, the prognosis of GC patients is still comparatively poor [3]. The majority of GC patients are diagnosed in the middle to late stage and therefore lose the opportunity to be cured [4]. Therefore, it is imperative to explore the regulatory mechanisms of GC, which not only contributes to improving the understanding of the pathogenesis of GC but also provides novel biomarkers for the diagnosis and therapy of GC.

It is widely accepted that noncoding RNA (ncRNA) could regulate the progression of multiple diseases through modulating gene expression at the transcriptional and posttranscriptional levels [5]. In 2011, Salmena et al. proposed the competing endogenous RNA (ceRNA) hypothesis that long noncoding RNA (lncRNA)

[†]Jianxin Li and Ting Han contributed equally to this work.

*Correspondence: yangqingqiang121@163.com

Department of General Surgery (Gastrointestinal Surgery), The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan 646000, People's Republic of China

Li *et al. World Journal of Surgical Oncology*   (2022) 20:295

Page 2 of 11

can suppress mRNA degradation or silence mRNA translation by sponging microRNA (miRNA), thereby affecting protein coding and modulating the disease process [6]. Recently, increasing studies have demonstrated that the ceRNA network might play critical roles in initiation and progression of multiple diseases, including cancer. For example, Xu et al. reported that lncRNA SNHG1 exerted as a sponge for miR-154-5p, thereby facilitating colorectal cancer cell growth through activating the downstream target of miR-154-5p, CCND2 [7]. Xin et al. demonstrated that lncRNA LINC01133 accelerates proliferation and aggressive of liver cancer cells by sponging miR-199a-5p to activate the ANXA2/STAT3 signaling pathway [8]. Zhao et al. indicated that lncRNA HOTAIR promotes cell growth, metastasis, and apoptosis of breast cancer through the miR-20a-5p/HMGA2 signaling [9]. Moreover, several studies have also reported that ceRNA network could play a vital role in the carcinogenesis of GC [10]. Nevertheless, pivotal lncRNA-miRNA-mRNA ceRNA networks involved in the progression of GC still need to be clarified.

In the present study, we first identified a list of diagnostic biomarkers closely related to GC from Gene Expression Omnibus dataset by machine learning and validated them in The Cancer Genome Atlas dataset. Then, by using multiple bioinformatic methods, the upstream regulatory miRNA and lncRNA were reversely predicted and validated from the perspectives of expression pattern and prognostic value. Ultimately, a novel ceRNA regulatory network was successfully developed, and each component in the network utterly conformed with ceRNA theory and meanwhile associated with the prognosis of GC patients.

Materials and methods
Data collection and processing
Gene expression microarray data sets GSE13911, GSE19826, and GSE79973 were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). Among them are as follows:

- GSE13911 including 31 normal samples and 38 GC samples [11]
- GSE19826 including 15 normal samples and 12 GC samples [12]
- GSE79973 including 10 normal samples and 10 GC samples [13]

All of these three datasets were based on GPL570 platform [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Then, these three datasets were merged into a combined dataset and used as the training cohort, and the surrogate variable analysis (SVA)

algorithm was applied to eliminate the batch effect between any two datasets [14]. Besides, the RNA-sequencing data of GC was downloaded from The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/) and used as the testing cohort.

Differentially expressed gene screening
Differentially expressed genes (DEGs) between normal samples and GC samples were analyzed by using the "limma" package in R in the training cohort [15]. The cutoff criteria for identifying DEGs was as follows: | log2-fold change (FC) | ≥ 2 and adjusted $P < 0.05$.

Biological function enrichment analysis
Gene ontology (GO) analysis is commonly used to annotate the biological functions or localization of genes from the perspective of biological processes (BP), cellular components (CC), and molecular functions (MF) [16]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis database is a knowledge resource for systematic analysis of gene functions, linking genomic information with signaling pathways [17]. Disease ontology (DO) analysis is usually applied to assess how particular genes may be involved in or influenced by specific disease states [18]. In the present study, we performed GO, KEGG, and DO analyses by using the "clusterProfiler" R package to analyze the DEGs at the functional level [19]. Only the enriched terms with $P < 0.05$ were considered statistically significant.

Diagnostic gene identification and verification
The least absolute shrinkage and selection operator (LASSO) is a regression-based algorithm that has the unique feature of penalizing the absolute value of a regression coefficient, thus automatically avoiding the overfitting and removing uninfluential variates [20]. In the present study, we created LASSO logistic regression model using the "glmnet" package to screen the potential diagnostic biomarkers from DEGs for GC in the training cohort. In addition, support vector machine recursive feature elimination (SVM-RFE), an algorithm widely used for cancer classification, biomarker discovery, and cancer driver gene discovery, was performed to further identify these biomarkers with diagnostic value in GC through the "e1071" package [21]. Area under receiver operating characteristic curve (AUC) was calculated to assess the predictive accuracy of the candidate diagnostic biomarkers. Furthermore, we also evaluated the diagnostic and prognostic values of the candidate diagnostic biomarkers in TCGA cohort.

Li *et al. World Journal of Surgical Oncology*   (2022) 20:295

Page 3 of 11

Identification of upstream miRNA

The upstream miRNAs that interacted with candidate diagnostic genes were screened by using miRTar-Base (http://mirtarbase.mbc.nctu.edu.tw/index.html), an online platform whose miRNA-target interactions have been validated using distinct types of experiments, including reporter assay, Western blot, microarray, and next-generation sequencing technologies [22]. Then, we performed correlation analysis, differential analysis, and survival analysis on these upstream miRNAs to filter out candidate miRNAs that could potentially bind to candidate diagnostic genes based on TCGA project. The criteria for candidate miRNAs were defined as follows: (1) negatively correlated with its targeted mRNA, (2) differentially expressed between GC and normal samples, and (3) correlated with the prognosis of patients with GC. A P-value < 0.05 was considered statistically significant.

Identification of upstream lncRNA

The upstream lncRNAs that sponged with candidate miRNA were identified by exploring LncBase v2 (www.microrna.gr/LncBase), a reference repository that contains an extensive collection of miRNA-lncRNA interactions which has been experimentally validated [23]. Similarly, we also performed correlation analysis, differential analysis, and survival analysis to screen out candidate lncRNAs based on TCGA project. The criteria were defined as follows: (1) negatively correlated with its interacted miRNA and meanwhile positively correlated with downstream mRNA, (2) differentially expressed between GC and normal samples, and (3) correlated with the prognosis of patients with GC. P-value < 0.05 was considered statistically significant.

Profile of immune cell infiltration

To determine the correlation between the proportion and composition of tumor-infiltrating immune cells (TIICs) and the expression of candidate diagnostic genes, we applied the EPIC (http://epic.gfellerlab.org) algorithm to assess the abundances of TIICs among the GC samples from the TCGA project [24]. Then, GC samples were divided into low- and high-expression groups according to the median expression level of mRNA, miRNA, and lncRNA, and the differences in TIIC content between the low- and high-expression groups were analyzed. In addition, we further investigated the prognostic value of distinct TIICs in GC patients.

Statistical analysis

All statistical analysis was performed using the R software (v4.1.1 https://www.r-project.org/). Differential expression analysis was assessed by Wilcoxon signed-rank test.

Fisher's test was applied to screen the significant GO, KEGG, and DO enrichment terms. Correlation analysis was estimated by Spearman correlation coefficients. The log-rank test was used in the Kaplan-Meier survival curve analysis. A P-value < 0.05 was considered statistically significant.

Results

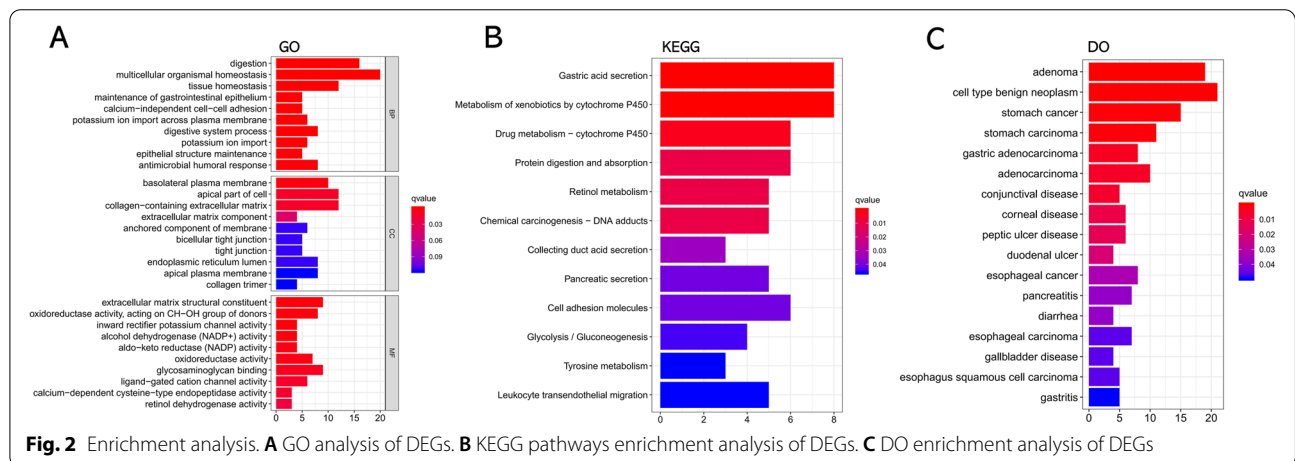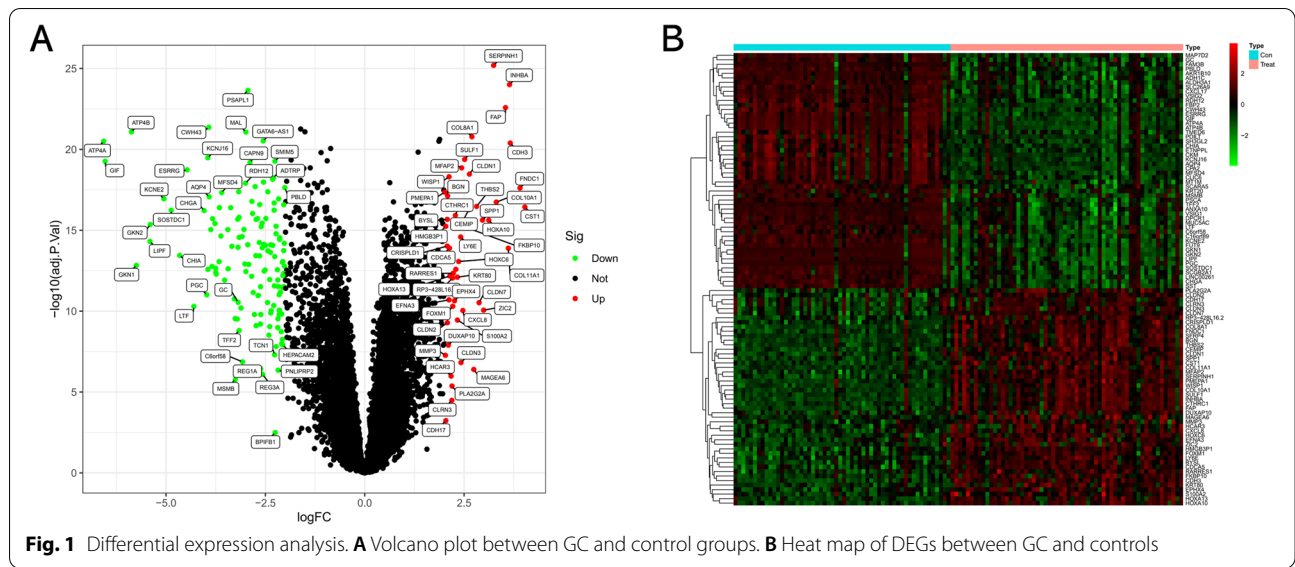### Identification of differentially expressed genes

First, a combined dataset including 56 normal samples and 60 GC samples was generated, and its gene expression matrix was normalized, and the batch effects were removed. Subsequently, a total of 188 DEGs between normal samples and GC samples were identified, including 48 upregulated DEGs and 140 downregulated DEGs (Supplementary Table 1). The volcano plot of these DEGs was displayed in Fig. 1A, and the expression heat map was presented in Fig. 1B.

Function enrichment analysis

The GO annotation analysis found that these DEGs participated in biological process of digestion, tissue homeostasis, and maintenance of gastrointestinal epithelium. Cellular component enriched these DEGs mainly in basolateral plasma membrane, apical part of cell, and collagen-containing extracellular matrix. Besides, molecular function suggested enrichment mainly at extracellular matrix structural constituent, oxidoreductase activity, and glycosaminoglycan binding (Fig. 2A). KEGG enrichment analysis revealed that these DEGs mainly enriched in gastric acid secretion, metabolism of xenobiotics by cytochrome P450, drug metabolism-cytochrome P450, and protein digestion and absorption (Fig. 2B). Moreover, DO enrichment found that these DEGs mainly involved in adenoma, cell type benign neoplasm, and stomach cancer (Fig. 2C).

Identification of candidate biomarkers in GC

We applied LASSO and SVM-RFE algorithms to identify candidate biomarkers among these DEGs. As a result, 13 DEGs were identified as potential biomarkers of GC according to the LASSO algorithm, while 40 DEGs were potential biomarkers of GC based on the SVM-RFE method (Fig. 3 A & B). Finally, six overlapped candidate biomarkers were identified by using the Venn diagram, including ADH7, CDH3, FAP, MT1M, PSAPL1, and SERPINH1 (Fig. 3C). The expression pattern of these potential biomarkers for GC in the training cohort was presented in Fig. 3D. In addition, we further examined the diagnostic efficiency of these potential biomarkers through ROC curves in the training cohort. As shown in Fig. 3E, the results found that the AUC of all these

Li *et al. World Journal of Surgical Oncology*   (2022) 20:295

Page 4 of 11

Fig. 1 Differential expression analysis. **A** Volcano plot between GC and control groups. **B** Heat map of DEGs between GC and controls

Fig. 2 Enrichment analysis. **A** GO analysis of DEGs. **B** KEGG pathways enrichment analysis of DEGs. **C** DO enrichment analysis of DEGs

potential biomarkers was higher than 0.9, suggesting that ADH7, CDH3, FAP, MT1M, PSAPL1, and SERPINH1 are effective indicators of GC.
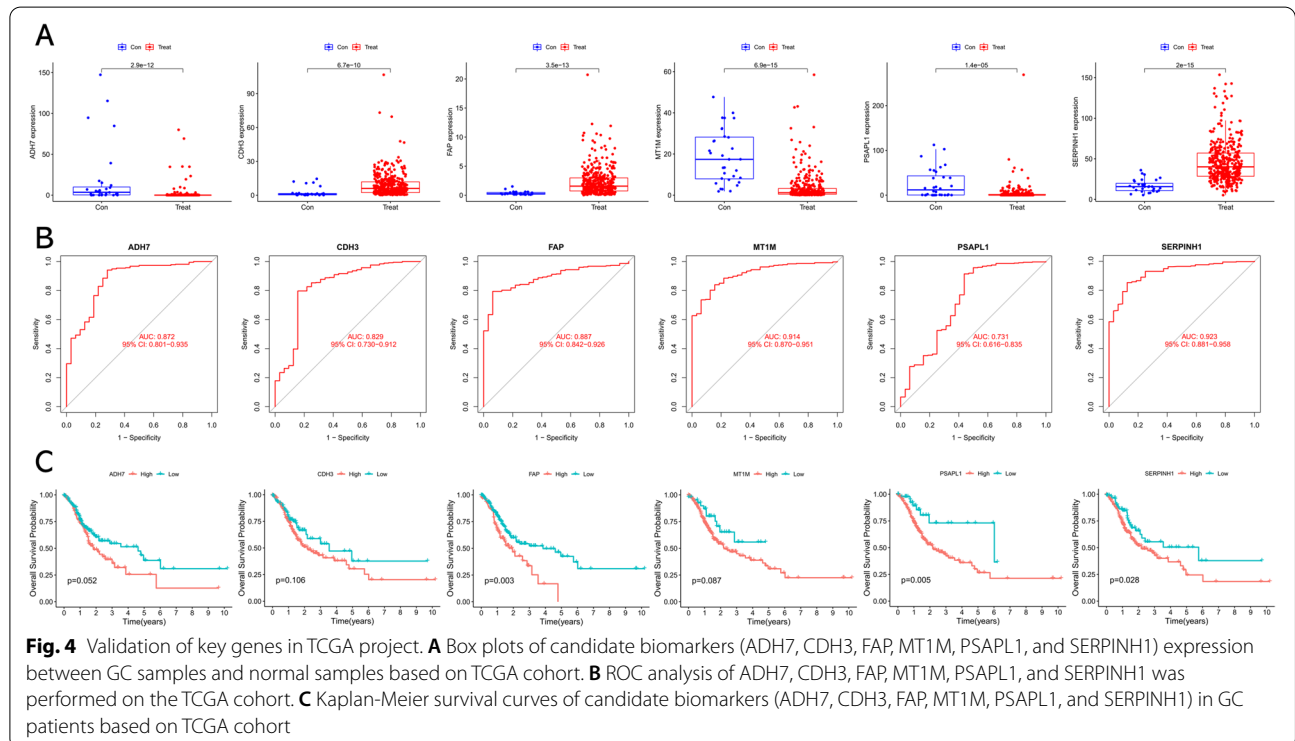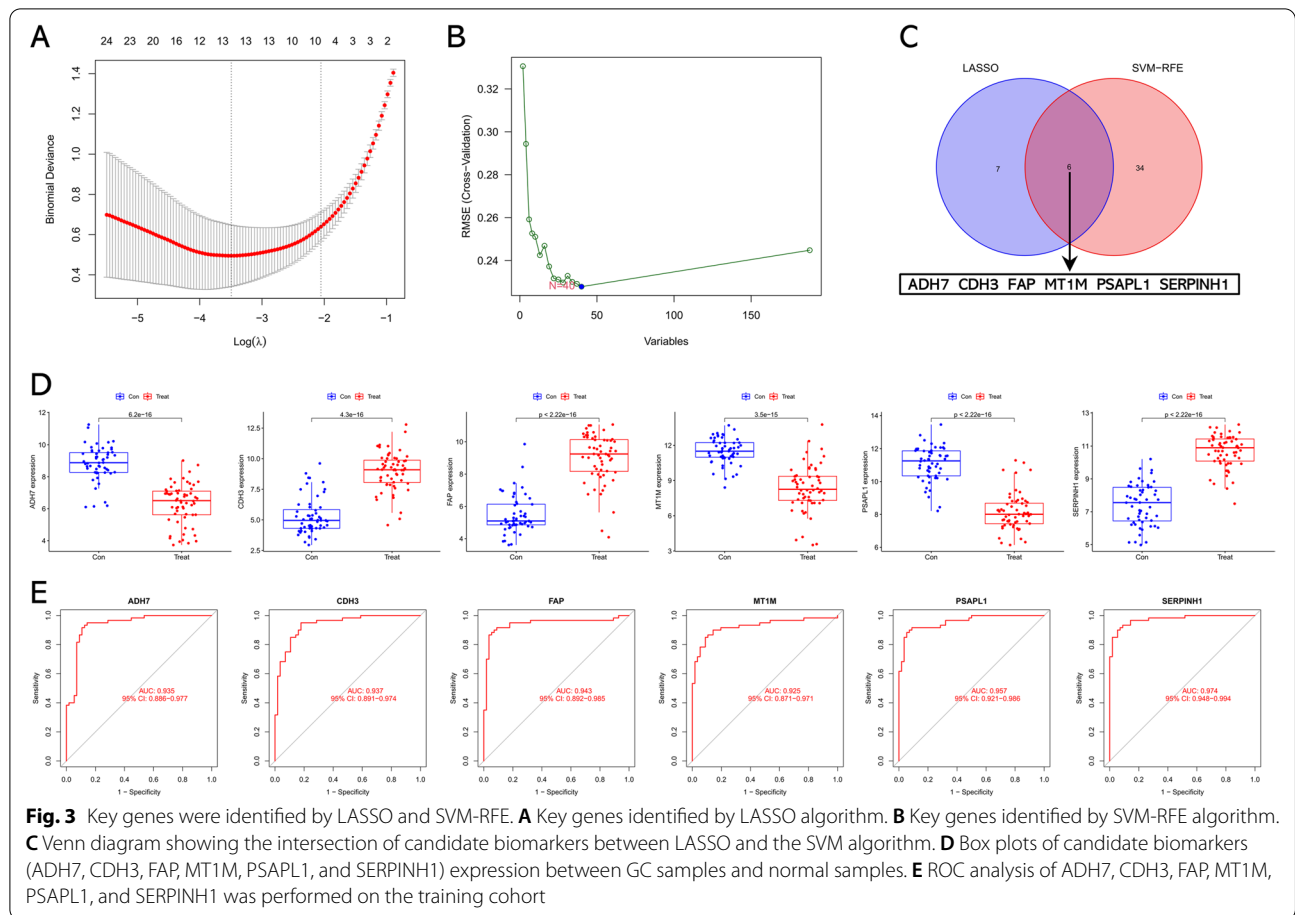
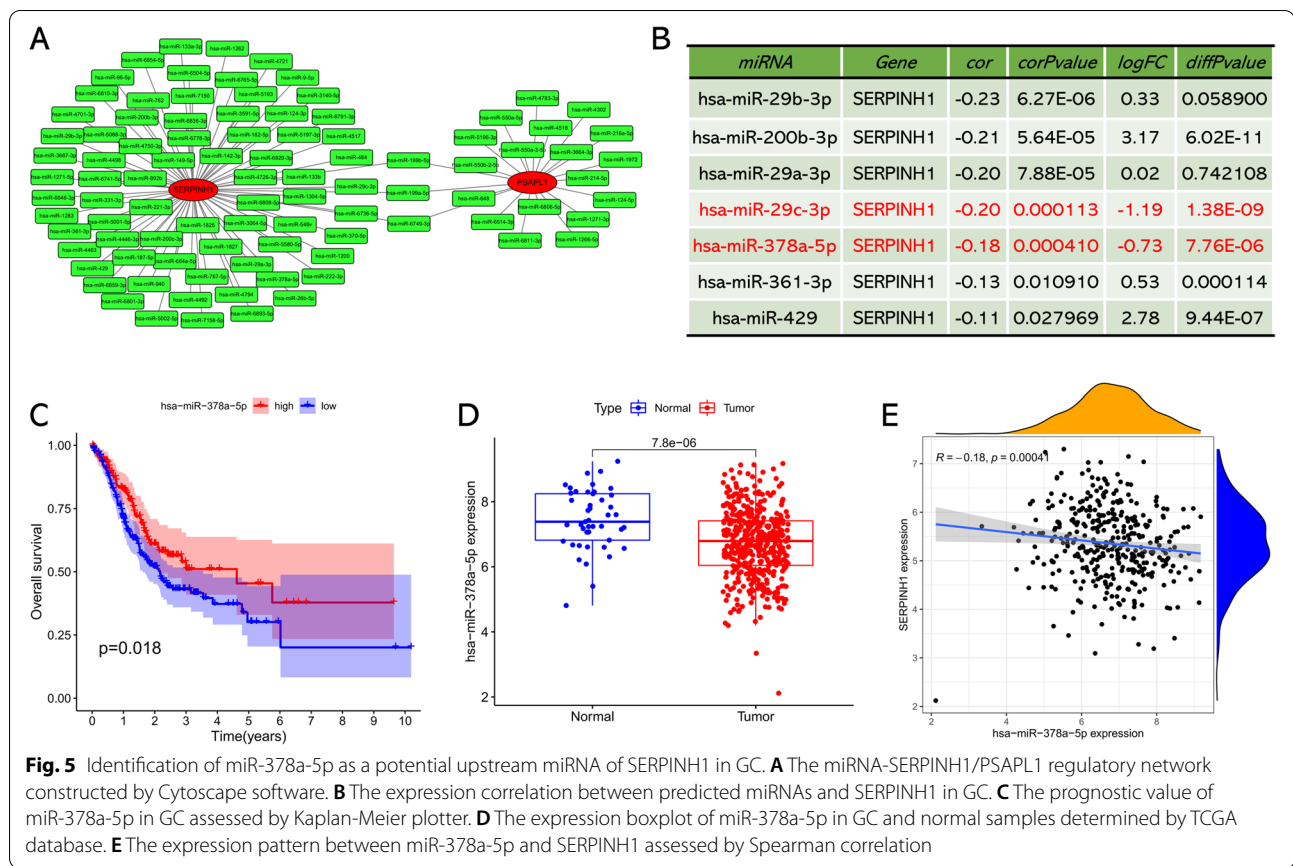### Validation of candidate biomarkers in GC based on TCGA

To enhance the reliability of our findings, the diagnostic efficiency of these potential biomarkers was further validated in the testing cohort based on TCGA project. First, RNA-sequencing data of 32 normal samples and 375 GC samples, and corresponding clinical information, were obtained from TCGA. Then, differential analysis revealed that the expression pattern of these five potential biomarkers was coincident with the results in the training cohort (Fig. 4A). ROC curves showed that the AUC of ADH7, CDH3, FAP, MT1M, PSAPL1, and SERPINH1 were 0.872, 0.829, 0.887, 0.914, 0.731, and 0.923, respectively (Fig. 4B). Finally, we further performed survival

analysis to evaluate the prognostic value of these potential biomarkers and found that dysregulation of FAP, PSAPL1, and SERPINH1 was significantly correlated with the prognosis of GC patients (Fig. 4C). Thus, FAP, PSAPL1, and SERPINH1 were identified as candidate biomarkers in GC and selected for subsequent study.

Identification of candidate upstream miRNA

We conducted miRTarBase database to predict the upstream miRNA of candidate biomarkers and found that 94 miRNAs interacted with PSAPL1 and SER-PINH1, while FAP was not observed to interact with any miRNA in this database. Then, a miRNA-mRNA network comprised of 97 miRNA-mRNA relationship pairs was constructed using Cytoscape (http://cytoscape. org/) (Fig. 5A) [25]. According to the inverse regulatory relationship between mRNA and miRNA, we further

Fig. 3 Key genes were identified by LASSO and SVM-RFE. **A** Key genes identified by LASSO algorithm. **B** Key genes identified by SVM-RFE algorithm. **C** Venn diagram showing the intersection of candidate biomarkers between LASSO and the SVM algorithm. **D** Box plots of candidate biomarkers (ADH7, CDH3, FAP, MT1M, PSAPL1, and SERPINH1) expression between GC samples and normal samples. **E** ROC analysis of ADH7, CDH3, FAP, MT1M, PSAPL1, and SERPINH1 was performed on the training cohort

Fig. 4 Validation of key genes in TCGA project. **A** Box plots of candidate biomarkers (ADH7, CDH3, FAP, MT1M, PSAPL1, and SERPINH1) expression between GC samples and normal samples based on TCGA cohort. **B** ROC analysis of ADH7, CDH3, FAP, MT1M, PSAPL1, and SERPINH1 was performed on the TCGA cohort. **C** Kaplan-Meier survival curves of candidate biomarkers (ADH7, CDH3, FAP, MT1M, PSAPL1, and SERPINH1) in GC patients based on TCGA cohort

Li *et al. World Journal of Surgical Oncology*   (2022) 20:295

Page 6 of 11

Fig. 5 Identification of miR-378a-5p as a potential upstream miRNA of SERPINH1 in GC. **A** The miRNA-SERPINH1/PSAPL1 regulatory network constructed by Cytoscape software. **B** The expression correlation between predicted miRNAs and SERPINH1 in GC. **C** The prognostic value of miR-378a-5p in GC assessed by Kaplan-Meier plotter. **D** The expression boxplot of miR-378a-5p in GC and normal samples determined by TCGA database. **E** The expression pattern between miR-378a-5p and SERPINH1 assessed by Spearman correlation

performed correlation analysis and differential analysis on these upstream miRNAs. As shown in Fig. 5B, seven upstream miRNAs showed a negative correlation with SERPINH1, and only two of which (hsa-miR-29c-3p and hsa-miR-378a-5p) were downregulated in GC. As for PSAPL1, there was no upstream miRNA that showed a negative correlation with it. Subsequent survival analysis revealed that only hsa-miR-378a-5p was significantly associated with a favorable prognosis in GC patients (Fig. 5C). Thus, hsa-miR-378a-5p was identified as a candidate miRNA that could be the most potential regulatory miRNA of SERPINH1 in GC and chosen for further analysis. The expression boxplot of hsa-miR-378a-5p in GC was presented in Fig. 5D, and the correlation between hsa-miR-378a-5p and SERPINH1 was presented in Fig. 5E.
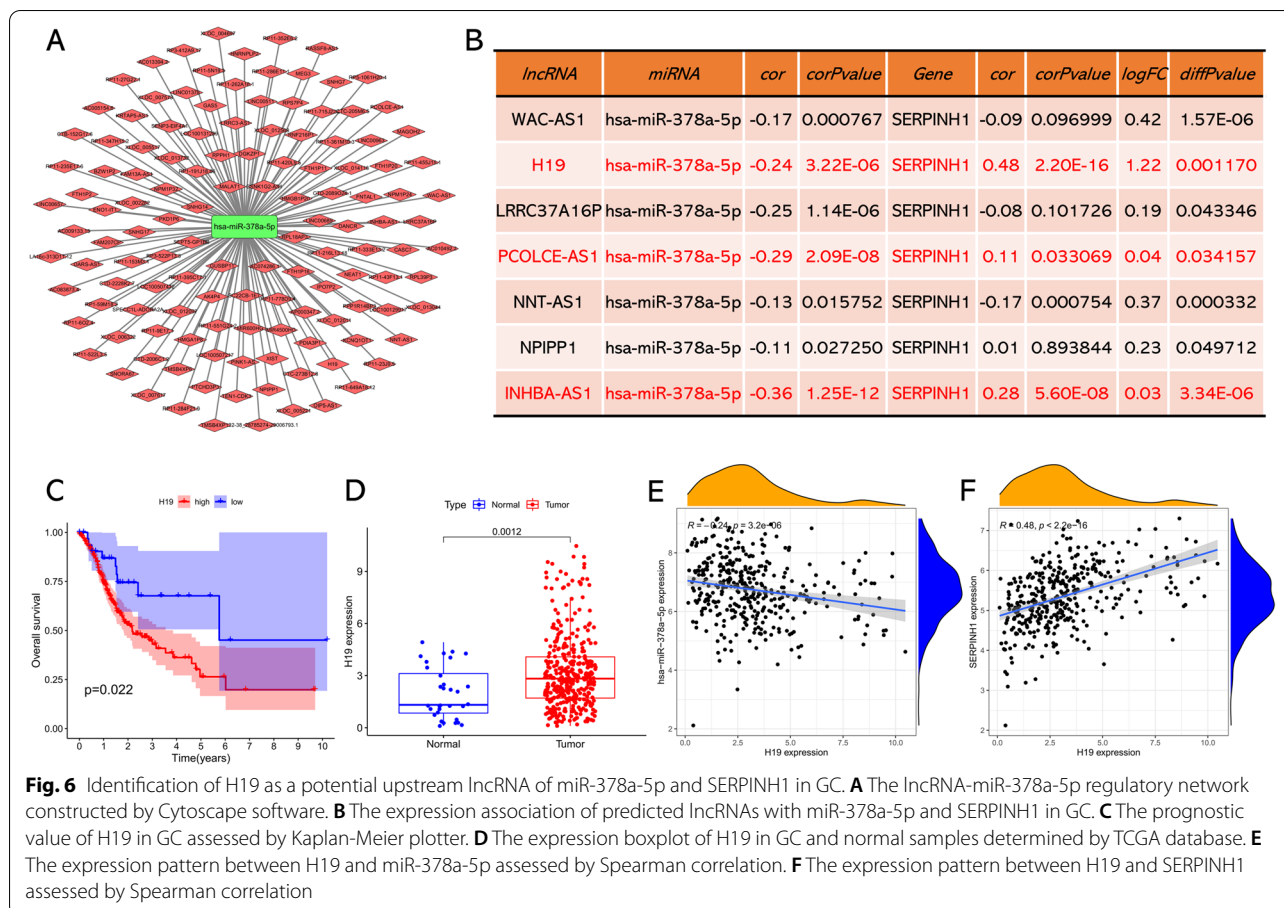
Identification of candidate upstream lncRNA
Previous study has indicated that lncRNA can function as sponge to competitively bind to miRNA [26]. Thus, we used the LncBase v2 database to screen candidate upstream lncRNA that could potentially bind to hsa-miR-378a-5p. As shown in Fig. 6A, a total of 129 possible lncR-NAs were identified. According to the ceRNA theory, the

upstream lncRNA should be negatively associated with miRNA and meanwhile positively associated with mRNA. Therefore, we validate the expression pattern of those predicted lncRNAs based on TCGA project. Among all the 129 lncRNAs, seven lncRNAs were negatively associated with hsa-miR-378a-5p expression in GC, and only three of which (H19, PCOLCE-AS1, and INHBA-AS1) were positively associated with SERPINH1 and meanwhile overexpressed in GC (Fig. 6B). Then, survival analysis of these three lncRNAs showed that only H19 was significantly correlated with poor prognosis in GC patients (Fig. 6C). The expression boxplot of H19 in GC was presented in Fig. 6D, and its correlation with hsa-miR-378a-5p and SERPINH1 was presented in Fig. 6 E and F, respectively. Taken all these results into consideration, H19 serves as a candidate upstream lncRNA that could regulate hsa-miR-378a-5p, and H19/miR-378a-5p/SERPINH1 axis might be a potential regulatory pathway in GC.

Profile of immune infiltration in GC
We further evaluated the correlation of candidate biomarkers' expression with immune cell infiltration level through the EPIC algorithm. As shown in Fig. 7A, infiltrating levels of cancer-associated fibroblasts (CAF)
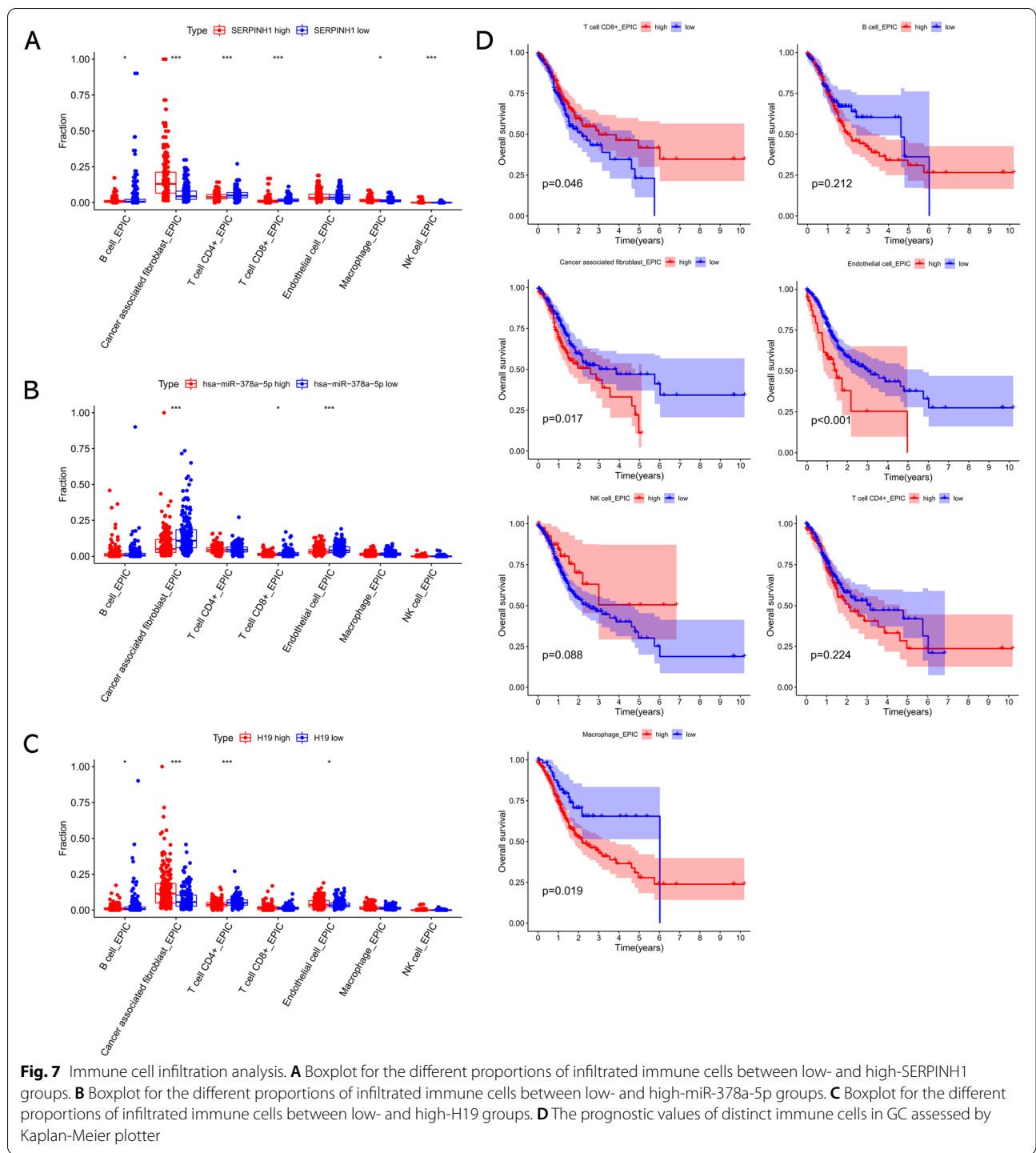
Li *et al. World Journal of Surgical Oncology* (2022) 20:295

Page 7 of 11

Fig. 6 Identification of H19 as a potential upstream lncRNA of miR-378a-5p and SERPINH1 in GC. **A** The lncRNA-miR-378a-5p regulatory network constructed by Cytoscape software. **B** The expression association of predicted lncRNAs with miR-378a-5p and SERPINH1 in GC. **C** The prognostic value of H19 in GC assessed by Kaplan-Meier plotter. **D** The expression boxplot of H19 in GC and normal samples determined by TCGA database. **E** The expression pattern between H19 and miR-378a-5p assessed by Spearman correlation. **F** The expression pattern between H19 and SERPINH1 assessed by Spearman correlation

and macrophages were elevated in high SERPINH1 expression samples, whereas infiltrating levels of B cells, CD4+ T cells, and CD8+ T cells were downregulated in high SERPINH1 expression samples compared to low SERPINH1 expression samples. The infiltrating levels of CAFs, CD8+ T cells, and endothelial cells were decreased in high hsa-miR-378a-5p expression samples compared to low hsa-miR-378a-5p expression samples (Fig. 7B). In addition, GC patients with high H19 expression infiltrated with higher CAFs and endothelial cells and lower B cells and CD4+ T cells compared to those with low H19 expression (Fig. 7C). Survival analysis revealed that infiltrating levels of CD8+ T cells, CAFs, endothelial cells, and macrophages were significantly correlated with the prognosis of GC patients (Fig. 7D). All these findings suggested that H19 might regulate the immune cell infiltration in carcinogenesis of GC through miR-378a-5p/SERPINH1 signaling.

Discussion

To date, the rapidly developed high-throughput sequencing technology and bioinformatics provide us with a more convenient platform to explore the pathogenesis of tumors at the gene level. In the present study, we successfully constructed a ceRNA network comprised of lncRNA (H19), miRNA (hsa-miR-378a-5p), and mRNA (SERPINH1) to provide a more comprehensive view of the RNA regulatory mechanism during GC carcinogenesis by combining multiple bioinformatic platforms.

We first identified three diagnostic biomarkers (FAP, PSAPL1, and SERPINH1) in GC by applying machine learning based on GEO and TCGA projects. Among them, FAP is a type 2 membrane-bound glycoprotein, which belongs to the serine protease family and has been identified as a marker of reactive tumor stromal fibroblasts [27]. FAP was reported to be highly expressed in GC tissues compared to normal controls, and patients in advanced pathological stage showed higher FAP expression levels than those in early pathological stage [28]. In addition, studies also found that FAP was overexpressed in GC cells, and FAP knocking down significantly restrained invasion and migration of GC cells by suppressing the activity of CAF [29]. SERPINH1, also known as HSP47, is an important collagen-specific molecule which is essential for the correct folding and secretion of distinct collagen

Li *et al. World Journal of Surgical Oncology*   (2022) 20:295

Page 8 of 11

Fig. 7 Immune cell infiltration analysis. **A** Boxplot for the different proportions of infiltrated immune cells between low- and high-SERPINH1 groups. **B** Boxplot for the different proportions of infiltrated immune cells between low- and high-miR-378a-5p groups. **C** Boxplot for the different proportions of infiltrated immune cells between low- and high-H19 groups. **D** The prognostic values of distinct immune cells in GC assessed by Kaplan-Meier plotter

types [30]. The mRNA and protein expression levels of SERPINH1 have been reported to be significantly upregulated in GC tissues compared with normal tissues, and inhibition of SERPINH1 significantly suppressed cancer cell migration and invasive abilities [31, 32]. Besides, upregulated SERPINH1 levels have been

reported to be associated with poor prognosis in GC patients [32]. Studies focused on the oncogenic role of PSAPL1 in GC are limited to date, so they are worthy of future research.

MiRNAs are single-stranded noncoding RNAs that regulate gene expression through transcript degradation

Li *et al. World Journal of Surgical Oncology*   (2022) 20:295

Page 9 of 11

or inhibition of protein translation at posttranscriptional level [33]. Accumulating studies have indicated that miRNAs play vital roles in diverse biological processes of multiple diseases, including tumors [34, 35]. In the present study, the upstream miRNAs were predicted and validated based on bioinformatic database for the purpose of exploring candidate ceRNA regulating the diagnostic biomarkers mentioned above. As a result, hsa-miR-378a-5p was identified as a regulatory miRNA that could bind to SERPINH1 in GC. Studies have reported that the expression level of hsa-miR-378a-5p was downregulated in colorectal cancer (CRC) tissues compared to normal controls, and decreased hsa-miR-378a-5p level was significantly associated with advanced histological grade and worse prognosis in CRC patients [36]. Li et al. also indicated that hsa-miR-378a-5p serves as a tumor suppressive role in CRC, and overexpression of hsa-miR-378a-5p inhibited CRC cell proliferation by targeting CDK1. In addition, hsa-miR-378a-5p was reported to promote apoptosis of triple-negative breast cancer cells by targeting SUFU [37]. These results partially enhance the credibility of our finding that hsa-miR-378a-5p serves as a tumor suppressor in GC through targeting SERPINH1.

LncRNAs are a series of transcripts with length greater than 200 nucleotides and no protein-coding ability [38]. Similarly, the aberrant expression of lncRNA has been widely reported to participate in carcinogenesis and progression of diverse tumors [39, 40]. LncRNA H19 is a maternally expressed gene located at human chromosomal 11p15.5, which plays an important role in distinct pathologic processes [41, 42]. Accumulating studies showed that H19 serves as an oncogenic role and was upregulated in many malignancies, including breast cancer [42], ovarian cancer [43], lung cancer [44], and pancreatic cancer [45]. Moreover, existing evidences also demonstrated that H19 participates in GC progression through diverse pathways. For example, the expression level of H19 was found to be significantly upregulated in GC tissues and cell lines compared to that in the normal controls, and elevated H19 expression was remarkably related to advanced pathological stage in GC [46]. Besides, H19 serves as a prognostic biomarker in GC, and patients with high H19 expression showed a worse prognosis than those with low H19 expression [47]. H19 was also found to promote the epithelial-mesenchymal transition (EMT) and metastasis in GC by activating Wnt signaling [48]. Gan et al. indicated that H19 overexpression significantly enhanced, whereas H19 silencing suppressed the proliferation, migration, and invasion of GC cells through regulating miR-22-3p/Snail1 axis in vitro and in vivo [49]. In the present study, we found that H19

was overexpressed in GC and significantly associated with patients' prognosis. Importantly, the expression pattern of H19 was negatively correlated with miR-378a-5p and meanwhile positively correlated with SERPINH1 in GC. Based on the ceRNA hypothesis, H19 was identified as the upstream regulatory lncRNA that could regulate the miR-378a-5p/SERPINH1 axis in GC.

In recent years, anticancer immunotherapy based on the reactivation of the host immunoreaction has revolutionized the treatment of patients with cancer and gained unprecedented progress [50]. However, the clinical use of immunotherapeutic agents is very limited due to the mechanisms of immune dysfunction in GC remain largely unclear. Emerging evidences have demonstrated that the interreaction between cancer cells and immune components in the tumor microenvironment (TME) is the determinant for tumor progression/regression. Thus, we further explored the correlation between the H19/miR-378a-5p/SERPINH1 axis and diverse TIICs in GC. In consequence, dysregulation of H19/miR-378a-5p/SERPINH1 axis was significantly correlated with altered infiltration abundances of CAFs, macrophages, B cells, CD4+ T cells, and CD8+ T cells in GC. CAFs are the prominent component of the tumor stroma, which supports the tumor cells by modifying the TME, boosting angiogenesis, and maintaining inflammatory status [51]. High infiltration of CAFs in TME could promote the malignant progression of GC [52]. Macrophages are the largest fraction of TIICs in the TME and can be divided into two major distinct subtypes according to their phenotype and function [53]. M1 macrophages are involved in the control of tumor growth by secreting pro-inflammatory cytokines, whereas M2 macrophages contribute to tumor progression by the production of immunosuppressive factors and chemokines [54]. B cells are pluripotent lineages which serve as antibody secreting cells but also serve as antigen-presenting cells (APCs) and immunoregulatory cells [55]. Studies have indicated that B-cell infiltration is associated with controlling tumor development in GC [56]. T cells exhibit important antitumor activities, and numerous studies showed that high proportions of infiltrating CD4+ and CD8+ T cells correlated with better prognosis in GC patients [57]. Interestingly, we found that CAFs were highly infiltrated in high H19 and high SERPINH1 groups, whereas its infiltration level in high miR-378a-5p was significantly downregulated. These findings imply that H19 might regulate the infiltration of CAFs to facilitate the carcinogenesis and progression of GC through miR-378a-5p/SERPINH1 pathway.

The present study inevitably exists a limitation that all results and conclusions were achieved based on online public databases; further experimental studies

Li *et al. World Journal of Surgical Oncology*   (2022) 20:295

Page 10 of 11

should be performed to validate our findings. Nevertheless, previous studies focused on constructing survival-related ceRNA network in GC are rare, and the present study successfully constructed a novel ceRNA regulatory network significantly associated with the prognosis of GC patients. Importantly, we proposed a novel hypothesis that H19 regulates the infiltration of CAFs to facilitate the carcinogenesis and progression of GC through miR-378a-5p/SERPINH1 signaling, which provides a promising clue for future research.

Conclusion

In summary, based on machine learning and bioinformatics, the present study identified an immune-related prognostic ceRNA regulatory pathway that H19 might regulate the immune cell infiltration in carcinogenesis of GC through miR-378a-5p/SERPINH1 signaling.

Abbreviations
APC: Antigen-presenting cell; AUC: Area under receiver operating characteristic curve; BP: Biological processes; CAF: Cancer-associated fibroblasts; CC: Cellular components; ceRNA: Competing endogenous RNA; CRC: Colorectal cancer; DEG: Differentially expressed gene; DO: Disease ontology; EMT: Epithelial-mesenchymal transition; GC: Gastric cancer; GEO: Gene Expression Omnibus; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; LASSO: Least absolute shrinkage and selection operator; lncRNA: Long noncoding RNA; MF: Molecular functions; miRNA: Micro RNA; ncRNA: Noncoding RNA; SVA: Surrogate variable analysis; SVM-RFE: Support vector machine recursive feature elimination; TCGA: The Cancer Genome Atlas; TIIC: Tumor-infiltrating immune cell; TME: Tumor microenvironment.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12957-022-02760-6.

> **Additional file 1: Supplementary Table 1**. DEGs between GC and normal samples.

Availability of data and materials
Thedatasets generated and/or analyzed during the current study are available in the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) and The Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga) projects.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Crew KD, Neugut AI. Epidemiology of gastric cancer. World J Gastroenterol. 2006;12(3):354–62. https://doi.org/10.3748/wjg.v12.i3.354.
2. Thrift AP, El-Serag HB. Burden of gastric cancer. Clin Gastroenterol Hepatol. 2020;18(3):534–42. https://doi.org/10.1016/j.cgh.2019.07.045.
3. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. Cancer Epidemiol Biomark Prev. 2014;23(5):700–13. https://doi.org/10.1158/1055-9965.EPI-13-1057.
4. Wagner AD, Grothe W, Haerting J, Kleber G, Grothey A, Fleig WE. Chemotherapy in advanced gastric cancer: a systematic review and meta-analysis based on aggregate data. J Clin Oncol. 2006;24(18):2903–9. https://doi.org/10.1200/JCO.2005.05.0245.
5. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861–74. https://doi.org/10.1038/nrg3074.
6. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146(3):353–8. https://doi.org/10.1016/j.cell.2011.07.014.
7. Xu M, Chen X, Lin K, et al. The long noncoding RNA SNHG1 regulates colorectal cancer cell growth through interactions with EZH2 and miR-154-5p. Mol Cancer. 2018;17(1):141. https://doi.org/10.1186/s12943-018-0894-x.
8. Yin D, Hu ZQ, Luo CB, et al. LINC01133 promotes hepatocellular carcinoma progression by sponging miR-199a-5p and activating annexin A2. Clin Transl Med. 2021;11(5):e409. https://doi.org/10.1002/ctm2.409.
9. Zhao W, Geng D, Li S, Chen Z, Sun M. LncRNA HOTAIR influences cell growth, migration, invasion, and apoptosis via the miR-20a-5p/HMGA2 axis in breast cancer. Cancer Med. 2018;7(3):842–55. https://doi.org/10.1002/cam4.1353.
10. Wu H, Liu B, Chen Z, Li G, Zhang Z. MSC-induced lncRNA HCP5 drove fatty acid oxidation through miR-3619-5p/AMPK/PGC1α/CEBPB axis to promote stemness and chemo-resistance of gastric cancer. Cell Death Dis. 2020;11(4):233. https://doi.org/10.1038/s41419-020-2426-z.
11. D'Errico M, de Rinaldis E, Blasi MF, et al. Genome-wide expression profile of sporadic gastric cancers with microsatellite instability. Eur J Cancer. 2009;45(3):461–9. https://doi.org/10.1016/j.ejca.2008.10.032.
12. Wang Q, Wen YG, Li DP, et al. Upregulated INHBA expression is associated with poor survival in gastric cancer. Med Oncol. 2012;29(1):77–83. https://doi.org/10.1007/s12032-010-9766-y.
13. He J, Jin Y, Chen Y, et al. Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer. Onco Targets Ther. 2016;9:6099–109. https://doi.org/10.2147/OTT.S110203.
14. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007;3(9):1724–35. https://doi.org/10.1371/journal.pgen.0030161.
15. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.
16. Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. Nucleic Acids Res. 2006;34(Database issue):D322–6. https://doi.org/10.1093/nar/gkj021.
17. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27–30. https://doi.org/10.1093/nar/28.1.27.
18. Bello SM, Shimoyama M, Mitraka E, et al. Disease ontology: improving and unifying disease annotations across species. Dis Model Mech. 2018;11(3):dmm032839. https://doi.org/10.1242/dmm.032839.
19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7. https://doi.org/10.1089/omi.2011.0118.

Li *et al. World Journal of Surgical Oncology*   (2022) 20:295

Page 11 of 11

20. McEligot AJ, Poynor V, Sharma R, Panangadan A. Logistic LASSO regression for dietary intakes and breast cancer. Nutrients. 2020;12(9):2652. https://doi.org/10.3390/nu12092652.

21. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (svm) learning in cancer genomics. Cancer Genomics Proteomics. 2018;15(1):41–51. https://doi.org/10.21873/cgp.20063.

22. Chou CH, Shrestha S, Yang CD, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res. 2018;46(D1):D296–302. https://doi.org/10.1093/nar/gkx1067.

23. Paraskevopoulou MD, Vlachos IS, Karagkouni D, et al. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. Nucleic Acids Res. 2016;44(D1):D231–8. https://doi.org/10.1093/nar/gkv1270.

24. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. Elife. 2017;6:e26476. https://doi.org/10.7554/eLife.26476.

25. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. https://doi.org/10.1101/gr.1239303.

26. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. Nat Rev Genet. 2016;17(5):272–83. https://doi.org/10.1038/nrg.2016.20.

27. Abbas O, Richards JE, Mahalingam M. Fibroblast-activation protein: a single marker that confidently differentiates morpheaform/infiltrative basal cell carcinoma from desmoplastic trichoepithelioma. Mod Pathol. 2010;23(11):1535–43. https://doi.org/10.1038/modpathol.2010.142.

28. Gao LM, Wang F, Zheng Y, Fu ZZ, Zheng L, Chen LL. Roles of fibroblast activation protein and hepatocyte growth factor expressions in angiogenesis and metastasis of gastric cancer. Pathol Oncol Res. 2019;25(1):369–76. https://doi.org/10.1007/s12253-017-0359-3.

29. Wang RF, Zhang LH, Shan LH, et al. Effects of the fibroblast activation protein on the invasion and migration of gastric cancer. Exp Mol Pathol. 2013;95(3):350–6.

30. Duarte BDP, Bonatto D. The heat shock protein 47 as a potential biomarker and a therapeutic agent in cancer research. J Cancer Res Clin Oncol. 2018;144(12):2319–28. https://doi.org/10.1007/s00432-018-2739-9.

31. Kawagoe K, Wada M, Idichi T, et al. Regulation of aberrantly expressed SERPINH1 by antitumor miR-148a-5p inhibits cancer cell aggressiveness in gastric cancer. J Hum Genet. 2020;65(8):647–56. https://doi.org/10.1038/s10038-020-0746-6.

32. Tian S, Peng P, Li J, et al. SERPINH1 regulates EMT and gastric cancer metastasis via the Wnt/β-catenin signaling pathway. Aging (Albany NY). 2020;12(4):3574–93. https://doi.org/10.18632/aging.102831.

33. Wang Y, Wang L, Chen C, Chu X. New insights into the regulatory role of microRNA in tumor angiogenesis and clinical implications. Mol Cancer. 2018;17(1):22. Published 2018 Feb 7. https://doi.org/10.1186/s12943-018-0766-4.

34. Annese T, Tamma R, De Giorgis M, Ribatti D. microRNAs biogenesis, functions and role in tumor angiogenesis. Front Oncol. 2020;10:581007. https://doi.org/10.3389/fonc.2020.581007.

35. Xie T, Huang M, Wang Y, Wang L, Chen C, Chu X. MicroRNAs as regulators, biomarkers and therapeutic targets in the drug resistance of colorectal cancer. Cell Physiol Biochem. 2016;40(1-2):62–76. https://doi.org/10.1159/000452525.

36. Li H, Dai S, Zhen T, et al. Clinical and biological significance of miR-378a-3p and miR-378a-5p in colorectal cancer. Eur J Cancer. 2014;50(6):1207–21. https://doi.org/10.1016/j.ejca.2013.12.010.

37. Zheng S, Li M, Miao K, Xu H. lncRNA GAS5-promoted apoptosis in triple-negative breast cancer by targeting miR-378a-5p/SUFU signaling. J Cell Biochem. 2020;121(3):2225–35. https://doi.org/10.1002/jcb.29445.

38. Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. Cell. 2018;172(3):393–407. https://doi.org/10.1016/j.cell.2018.01.011.

39. Chen R, Li WX, Sun Y, et al. Comprehensive analysis of lncRNA and mRNA expression profiles in lung cancer. Clin Lab. 2017;63(2):313–20. https://doi.org/10.7754/Clin.Lab.2016.160812.

40. Peng WX, Koirala P, Mo YY. LncRNA-mediated regulation of cell signaling in cancer. Oncogene. 2017;36(41):5661–7. https://doi.org/10.1038/onc.2017.184.

41. Ding D, Li C, Zhao T, Li D, Yang L, Zhang B. LncRNA H19/miR-29b-3p/PGRN axis promoted epithelial-mesenchymal transition of colorectal

42. Wang J, Xie S, Yang J, et al. The long noncoding RNA H19 promotes tamoxifen resistance in breast cancer via autophagy. J Hematol Oncol. 2019;12(1):81. https://doi.org/10.1186/s13045-019-0747-0.

43. Ma H, Gao J, Yu H, Song X. Long non-coding RNA H19 correlates with unfavorable prognosis and promotes cell migration and invasion in ovarian cancer. Ginekol Pol. 2022:10.5603/GP.a2021.0079. https://doi.org/10.5603/GP.a2021.0079.

44. Zhao Y, Feng C, Li Y, Ma Y, Cai R. LncRNA H19 promotes lung cancer proliferation and metastasis by inhibiting miR-200a function. Mol Cell Biochem. 2019;460(1-2):1–8. https://doi.org/10.1007/s11010-019-03564-1.

45. Wang J, Zhao L, Shang K, et al. Long non-coding RNA H19, a novel therapeutic target for pancreatic cancer. Mol Med. 2020;26(1):30. https://doi.org/10.1186/s10020-020-00156-4.

46. Chen JS, Wang YF, Zhang XQ, et al. H19 serves as a diagnostic biomarker and up-regulation of H19 expression contributes to poor prognosis in patients with gastric cancer. Neoplasma. 2016;63(2):223–30. https://doi.org/10.4149/207_150821N454.

47. Zhou H, Shen W, Zou H, Lv Q, Shao P. Circulating exosomal long non-coding RNA H19 as a potential novel diagnostic and prognostic biomarker for gastric cancer. J Int Med Res. 2020;48(7):300060520934297. https://doi.org/10.1177/0300060520934297.

48. Liu J, Wang G, Zhao J, et al. LncRNA H19 promoted the epithelial to mesenchymal transition and metastasis in gastric cancer via activating Wnt/β-catenin signaling. Dig Dis. 2021:10.1159/000518627. https://doi.org/10.1159/000518627.

49. Gan L, Lv L, Liao S. Long non-coding RNA H19 regulates cell growth and metastasis via the miR-22-3p/Snail1 axis in gastric cancer. Int J Oncol. 2019;54(6):2157–68. https://doi.org/10.3892/ijo.2019.4773.

50. Picard E, Verschoor CP, Ma GW, Pawelec G. Relationships between immune landscapes, genetic subtypes and responses to immunotherapy in colorectal cancer. Front Immunol. 2020;11:369. https://doi.org/10.3389/fimmu.2020.00369.

51. Grunberg N, Pevsner-Fischer M, Goshen-Lago T, et al. Cancer-associated fibroblasts promote aggressive gastric cancer phenotypes via heat shock factor 1-mediated secretion of extracellular vesicles. Cancer Res. 2021;81(7):1639–53. https://doi.org/10.1158/0008-5472.CAN-20-2756.

52. Qin Y, Wang F, Ni H, et al. Cancer-associated fibroblasts in gastric cancer affect malignant progression via the CXCL12-CXCR4 axis. J Cancer. 2021;12(10):3011–23. https://doi.org/10.7150/jca.49707.

53. Yang Y, He W, Wang ZR, et al. Immune cell landscape in gastric cancer. Biomed Res Int. 2021;2021:1930706. Published 2021 Jan 9. https://doi.org/10.1155/2021/1930706.

54. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. Nat Rev Clin Oncol. 2017;14(7):399–416. https://doi.org/10.1038/nrclinonc.2016.217.

55. Wei Y, Huang CX, Xiao X, et al. B cell heterogeneity, plasticity, and functional diversity in cancer microenvironments. Oncogene. 2021;40(29):4737–45. https://doi.org/10.1038/s41388-021-01918-y.

56. Ni Z, Xing D, Zhang T, et al. Tumor-infiltrating B cell is associated with the control of progression of gastric cancer. Immunol Res. 2021;69(1):43–52. https://doi.org/10.1007/s12026-020-09167-z.

57. Zhang N, Cao M, Duan Y, Bai H, Li X, Wang Y. Prognostic role of tumor-infiltrating lymphocytes in gastric cancer: a meta-analysis and experimental validation. Arch Med Sci. 2019;16(5):1092–103. https://doi.org/10.5114/aoms.2019.86101.

Publisher's Note

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