Pathology & Oncology Research (2020) 26:1903–1919
https://doi.org/10.1007/s12253-019-00781-3

ORIGINAL ARTICLE

Identification of Key Genes and Signaling Pathways Associated with the Progression of Gastric Cancer

Chaoran Yu[1] · Jie Chen[1,2] · Junjun Ma[1] · Lu Zang[1] · Feng Dong[1] · Jing Sun[1] · Minhua Zheng[1]

Received: 3 August 2019 / Accepted: 19 November 2019 / Published online: 17 December 2019

Abstract

Genomic features have been gradually regarded as part of the fundamentals to the clinical diagnosis and treatment for gastric cancer. However, the molecular alterations taking place during the progression of gastric cancer remain unclear. Therefore, identification of potential key genes and pathways in the gastric cancer progression is crucial to clinical practices. The gene expression profile, GSE103236, was retrieved for the identification of the differentially expressed genes (DEGs), followed by gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichments, gene set enrichment analysis (GSEA) and the protein-protein interaction (PPI) networks. Multiple bioinformatics platforms were employed for expression and prognostic analysis. Fresh frozen gastric cancer tissues were used for external validation. A total of 161 DEGs were identified from GSE103236. The PPI network-derived hub genes included collagen type I alpha 1 chain (*COL1A1*), tissue inhibitor of the metalloproteinases (*TIMP1)*, Secreted Phosphoprotein 1 (*SPP1*), somatostatin (*SST*), neuropeptide Y (*NPY*), biglycan (*BGN*), matrix metallopeptidase 3 (*MMP3*), apolipoprotein E (*APOE*), ATPase H+/K+ transporting alpha subunit (*ATP4A*), lysyl oxidase (*LOX*). *SPP1* (log rank $p = 0.0048$, HR = 1.39 [1.1–1.75]) and *MMP3* (log rank $p < 0.0001$, HR = 1.77 [1.44–2.19]) were significantly associated with poor overall survival. Stage-specifically, both *COL1A1* and *BGN* were correlated with significant in stage III and IV gastric cancer cases. *LOX* showed significant correlation with prognosis in stage I and stage II gastric cancer cases. Furthermore, cg00583003 of *SPP1* and cg16466334 of *MMP3* exhibited highly methylation level and significant prognostic values (*SPP1*: HR = 1.625, $p = 0.013$; *MMP3*: HR = 0.647, $p = 0.011$). Hub genes signature displayed a favorable prognostic value (p value = 5.227e-05). *APOE* demonstrated the highest correlation with $CD8^+$ T cells, neutrophils, and dendritic cells whereas *BGN* had the highest correlation with macrophages. This study systematically explored the key genes and pathways involved in PGC and AGC, providing insights into therapeutic individualized management.

Keywords Differentially expressed genes · Gene ontology · KEGG pathway · Protein-protein interaction network · Gastric cancer · Gene set enrichment analysis

Chaoran Yu, Jie Chen and Junjun Ma contributed as co-first authors.

Minhua Zheng and Jing Sun are listed as corresponding authors (equally contributed).

✉ Chaoran Yu
chaoran_yu@sjtu.edu.cn

Jie Chen
misc_jc@yeah.net

Junjun Ma
marsnew1997@163.com

Lu Zang
zanglu@yeah.net

Feng Dong
dfsnake@vip.163.com

Jing Sun
sj11788@rjh.com.cn

Minhua Zheng
zmhtiger@yeah.net

[1] Department of Gastrointestinal Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, People's Republic of China

[2] Department of Nursing, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200025, People's Republic of China

Background

Gastric cancer (GC) is one of the most common, death-caused diseases with marked heterogeneity and phenotypic diversity, which continues to be a major global health challenge [1]. In China, despite both the mortality and morbidity is descending, GC remains a leading malignant disease across the nation [2]. The prognosis of GC is largely related to the tumor progression at presentation and surgical intervention is listed as the primarily therapeutic option [3, 4]. Of note, the global standardization of both the diagnosis and treatment in GC remains elusive.

The biological alterations occurred during the progression has been gradually regarded as one of the fundamentals to the clinical diagnosis, treatment and prognostic values of GC. The genomic features and biomarkers of GC had been previously investigated [5–7]. Noteworthy, the intrinsic features among the progression of GC are yet to be fully disclosed. The GSE103236 profile in this study had been previously studied by Chivu et al in 2010 for the identification of differentially expressed genes (DEGs) in primary gastric cancer (PGC) (American Joint Committee on Cancer, AJCC staging I, II) and advanced gastric cancer (AGC) (AJCC staging III, IV) [7]. Noteworthy, given the updated bioinformatics resources and well-established in silico analytic strategies, the GSE103236 was therefore processed for re-analysis and re-annotation of key genes, signaling pathways and prognostic values.

Hereby, with the multi-dimensional bioinformatics analysis strategy and public available resource, including Gene Expression Omnibus (GEO) profiles and The Cancer Genome Atlas (TCGA) data, this study provided multilevel bioinformatics analysis strategy, including gene set enrichment analysis (GSEA) and the protein-protein interaction (PPI) networks analysis between the GC versus normal tissue, the PGC versus normal tissue, as well as the PGC versus AGC comparisons subgroups. This study offered insights for potential candidate biomarkers, signaling pathways and mechanisms for molecular clinical detection and prognosis values of GC.

Materials and Methods

Gene Expression Profile from GEO

The gene expression profile, GSE103236, deposited by Chivu et al [7], was retrieved from the GEO (http://www.ncbi.nlm.nih.gov/geo), which offers massive public available genomic data and integrative web-based analysis tools, enabling comprehensive and dynamic genes analysis [8, 9]. The corresponding platform, GPL4133 (Agilent-014850 Whole Human Genome Microarray 4*44 K G4112F (Feature Number Version)) was also downloaded. The GSE103236 profile consisted of ten gastric cancer samples (including seven PGC and three AGC) and nine corresponding paired normal tissues.

Identification of the DEGs in Profile

The identification process of the DEGs was performed by a web-based analysis tool, GEO2R, which provided a reliable and repeatable comparison analysis [10]. The cut-off values for DEGs included adjusted-p value<0.05 and log fold-change ($|\log_2 FC|$) values ≥2. The DEG expressions between tumor tissues and the paired normal tissues were processed for volcano plot using R program and hierarchical clustering using FunRich program version 3 (Bundoora,Victoria,Australia) (http://www.funrich.org) [11].

GO and KEGG for the DEGs

Gene ontology (GO) was primarily established as an interactive encyclopedic vocabulary covering all information in eukaryotes while Kyoto Encyclopedia of Genes and Genomes (KEGG) was among the major resource providers for pathways and genomic annotations [12, 13]. Both GO and KEGG were processed for systematic and comprehensive information, further marking the potential biological processes, cellular components, molecular functions and pathways among the DEGs. Database for Annotation, Visualization and Integrated Discovery (DAVID) was employed for the functional enrichment analysis. It offered powerful bioinformatics analysis for all species [14]. The cut-off value for significant GO and KEGG results was false discovery rate (FDR) <0.25.

GSEA Analysis

GSEA software was established to determine the statistically significant gene sets in the comparison between different subgroups. GSEA was obtained from the Broad Institute (http://software.broadinstitute.org/gsea/index.jsp) [15]. The GSE103236 dataset was input with the annotation file "hallmark gene sets". The cut-off values were predefined as $p < 0.05$ and FDR < 0.25.

PPI Network and Module Analysis of the DEGs

The selected DEGs were then inputted to the Search Tool for the Retrieval of Interacting Genes (STRING) database for an integrative assessment for PPI annotation [16]. The data was processed by the Cytoscape program for further illustration with Molecular Complex Detection (MCODE). The preferred cut-off values were determined as the value of degree cut-off = 1, the value of max.depth = 100, the value of node score = 0.2 and the k-score = 2.

Analysis of the mRNA Expressions of the Hub Genes

The Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/index.html) is established for producing customized functionalities and integrative information for wide range analysis based on TCGA and the genotype-tissue expression (GTEx) projects [17]. The expressions of hub genes between tumor tissues and normal tissues, as well as those between different stages were analyzed in the stomach adenocarcinoma (STAD) of TCGA. The genomic alterations of hub genes were generated by the cBioPortal for Cancer Genomics [18, 19]. The gene expression profiles, GSE13861 and GSE27342, were used for external validation of mRNAs expression of the hub genes. GSE13861 consists of 65 primary gastric adenocarcinoma tissues and 19 surrounding normal tissues for microarray (Illumina human V3) [20]. GSE27342 consists of 80 paired gastric cancer and normal tissues (Affymetrix Human Exon 1.0 ST Array, transcript (gene) version) [21].

Protein Expression of Hub Genes in Human Protein Atlas

The proteins expression of the hub genes, both in gastric cancer and normal tissues, were determined from the public-available human protein atlas (HPA, www.proteinatlas.org) [22].

Clinical Tissues Samples for Validation by Quantitative Real-Time PCR (qRT-PCR)

Fresh gastric cancer and paired normal tissues were retrieved from the biobank of Shanghai Minimally Invasive Surgery Center at Ruijin Hospital (Shanghai, China) between April 2018 and July 2018 with written informed consents. A total of 10 paired fresh gastric cancer tissues were frozen in liquid nitrogen and stored at −80°Cfor further qRT-PCR. No chemo-radiotherapy was received for included patients prior to surgery. The total RNA was extracted with the Trizol reagent (Invitrogen, Carlsbad, CA, USA). The cDNA was generated by the RNA with PrimeScriptTM RT Master Mix (TaKara Bio, Otsu, Japan), and further amplified by the SYBR Green Real-time PCR Master Mix (TaKara Bio, Otsu, Japan) in a 20 μL system. The primers were designed by PrimerBank (https://pga.mgh.harvard.edu/primerbank) [23] and synthesized by Sangon (Shanghai, China). The fold changes (FC) ($FC = 2^{-\Delta\Delta CT}$) of the hub genes in tumor compared to paired normal tissues were calculated with 18S RNA as the internal reference. The primers (5′ to 3′) were as follow: COL1A1 (F:GAGGGCCAAGACGAAGACATC; R:CAGATCACGTCATCGCACAAC);

TIMP1 (F:CTTCTGCAATTCCGACCTCGT; R:ACGCTGGTATAAGGTGGTCTG);

SPP1 (F:CTCCATTGACTCGAACGACTC; R:CAGGTCTGCGAAACTTCTTAGAT); SST (F:ACCCAACCAGACGGAGAATGA; R:GCCGGGTTTGAGTTAGCAGA); NPY (F:CGCTGCGACACTACATCAAC; R:CTCTGGGCTGGATCGTTTTCC);

BGN (F:CAGTGGCTTTGAACCTGGAG; R:GGGAGGTCTTTGGGGATGC);

MMP3 (F:TCTATGGACCTCCCCCTGAC; R:GATTTGCGCCAAAAGTGCCT);

APOE (F:GTTGCTGGTCACATTCCTGG; R:GCAGGTAATCCCAAAAGCGAC);

ATP4A (F:GATGGAGATTAACGACCACCAG; R:GCAACCCACATGAGGCACT);

LOX (F:CGGCGGAGGAAAACTGTCT; R:TCGGCTGGGTAAGAAATCTGA);

18 s (F:GAGAGTGAGCGGCAGAGC; R:GCTCCCAAGATCCAACTACGAG).

Survival Analysis of the Hub Genes

Kaplan-Meier (KM) Plotter is one of the publicly available datasets for cancer microarray with clinical annotations, which enables reliable assessment of prognostic values among input genes [24]. It incorporates public available genomic resources with clinical data from GEO, including GSE14210, GSE15459, GSE22377, GSE29272, GSE51105 and GSE62254. The top selected genes were inputted as a query and the data were collected for overall survival (OS) analysis. The output results were displayed with log rank p values and the hazard ratio (HR) with 95% confidence intervals (95% CI).

Methylation Data

The DNA methylation of SPP1 and MMP3 in TCGA was retrieved from MethSurv, a website-based tool for DNA methylation analysis (https://biit.cs.ut.ee/methsurv/) [25]. The CpG methylation levels were depicted by heatmap with average linkage method and correlation distance. The prognostic value of each single CpG was further calculated.

Prognostic Values of the Hub Genes-Signature by the SurvExpress

The TCGA database (STAD) was used for the prognostic value assessment of the hub gene signatures ($n = 352$) in the SurvExpress (http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp) [26]. The low risk and high risk groups were categorized according to the maximized risk group algorithm [26].

The Correlation between the Hub Genes and the Tumor Immune Infiltrates

The correlation of hub genes and immune infiltrates (B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages and dendritic cells) in STAD of TCGA was analyzed via the Tumor IMmune Estimation Resource (TIMER) platform (https://cistrome.shinyapps.io/timer/), a comprehensive resource for systematical visualization of immune infiltrating cells across different cancers [27]. All the correlation values were calculated via purity-corrected partial Spearman's correlation. Generally, genes of negative association with tumor purity are highly expressed in the microenvironment whereas those of positive association indicate high expression in the tumor cells [27].

Statistics

The statistical significance between the tumor and normal groups was calculated by Student's *t* test. The statistical analysis and illustration were performed by GraphPad Prism 6 software program (San Diego, CA) and R software (Version 3.5.1, www.r-project.org). Adjusted *p* value <0.05 and $|\log_2 FC| >= 2$ were set up as the significant cut-off for DEGs only.

Results

The Identification of DEGs in Different Comparison Subgroups

Based on the predefined cut-off-adjusted *p* value <0.05 and $|\log2FC| \geq 2$, 161 DEGs were identified from tumor versus normal tissues in GSE103236, including 89 DEGs with down-regulation and 72 DEGs with upregulation (Fig. 1a, b). Next, the same cut-off values were implemented to screen the DEGs between the normal tissues versus PGC, as well as the PGC versus AGC. However, only 34 DEGs were identified between normal tissues and PGC. No significant DEGs were identified between PGC and AGC (Table 1). Therefore, only the 161 DEGs from the tumor versus normal tissues were used for subsequent GO and KEGG annotations.

GO and KEGG Analysis for the DEGs in Tumor Versus Normal Tissues

For the all DEGs, the extracellular matrix organization/extracellular structure organization/cell-cell signaling were the most enriched in the biological process (BP), the extracellular space/proteinaceous extracellular matrix/extracellular matrix were the most enriched gene terms in the cellular component (CC), the receptor binding/glycosaminoglycan binding were the most enriched gene terms in molecular function (MF) and yet no term was significant enriched in KEGG analysis (Fig. 2a). Furthermore, the enriched analysis of BP/CC/MF was also demonstrated in up/down-regulated gene clusters. For up-regulated genes, extracellular matrix organization/extracellular structure organization/skeletal system development were the most enriched gene items in BP, and proteinaceous extracellular matrix/extracellular matrix/extracellular space were the most enriched gene items in CC, while no term was found enriched in MF and KEGG. For down-regulated genes, extracellular space was significantly enriched in CC, while the rest terms remain insignificant in BP/MF and KEGG.

GSEA Analysis of the Gene Expression Files in Tumor Versus Normal, Normal Versus PGC and PGC Versus AGC Subgroups

The GSEA analysis was performed with the GSE103236 profile based on the predefined "hallmark signature". Noteworthy, only one gene set was found significantly

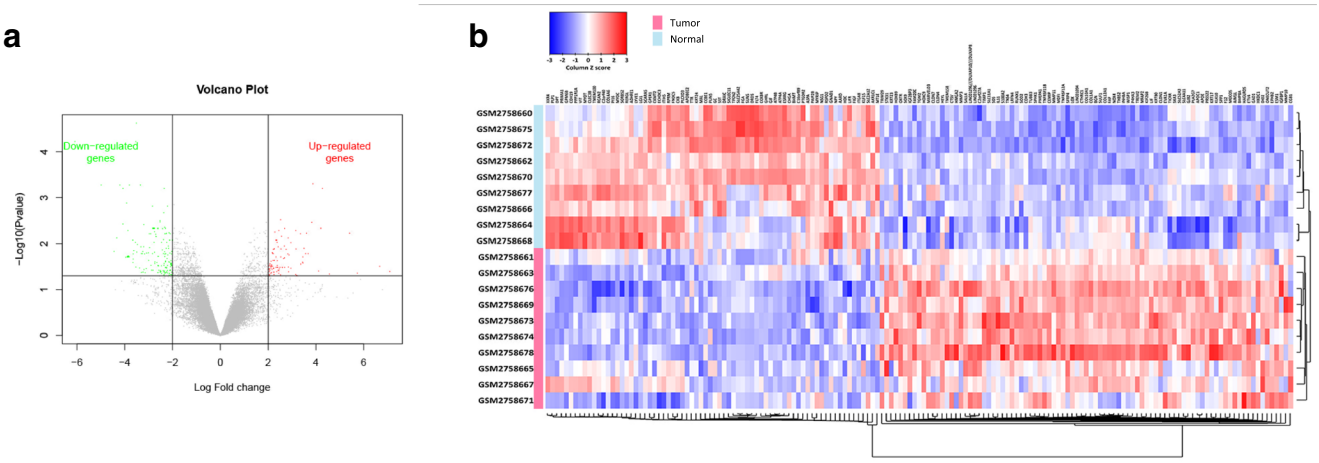

Fig. 1 Differentially expressed genes (DEGs) between tumor and normal tissues. **a** The volcano plot of DEGs; **b** the heat map of DEGs

Table 1 Differentially expressed genes (DEGs) between primary gastric cancer (PGC) and advanced gastric cancer (AGC)

ID	adj.P.Val	P.Value	logFC	Gene.symbol	Gene.title
23,421	0.328	0.0000215	−7.1918227	REG4	regenerating family member 4
29,875	0.328	0.0000218	2.6099073	ATP8B3	ATPase phospholipid transporting 8B3
42,478	0.359	0.0000319	−5.9789324	REG4	regenerating family member 4
4645	0.609	0.0000677	2.4691162	LIF	leukemia inhibitory factor
2359	0.723	0.0001002	2.2864412	VPS26A	VPS26, retromer complex component A
37,846	0.723	0.0001409	2.7886025	FAM83A	family with sequence similarity 83 member A
3337	0.723	0.0002225	3.2282741	IL1A	interleukin 1 alpha
4775	0.723	0.0002556	3.8082905	UTS2	urotensin 2
36,424	0.723	0.0002623	−2.2506787	STXBP6	syntaxin binding protein 6
18,140	0.723	0.0003578	−2.0936457	STXBP6	syntaxin binding protein 6

enriched in PGC compare to normal tissues at nominal *p* value of 0.01, i.e., angiogenesis (Nominal *p* value <0.001, FDR q-value =0.074, FWER *p* value =0.06) (Fig. 2b). 5 gene sets were significantly enriched in tumor compared to normal tissues, top three included G2M checkpoint (Nominal *p* value <0.001, FDR q-value =0.037, FWER p value =0.03), MYC targets V2 (Nominal *p* value <0.001, FDR q-value = 0.043, FWER *p* value =0.09) and Apoptosis (Nominal *p* value <0.001, FDR q-value =0.087, FWER *p* value =0.19) (Fig. 2c–e). Given the vacant DEGs results from PGC versus AGC with predefined cut-off values, GSEA strategy was also employed to address possible variances between the PGC and AGC. 8 gene sets were significantly enriched in AGC. Top three enriched gene sets include TNFA signaling via NFKB (Nominal *p* value <0.001, FDR q-value <0.001, FWER *p* value <0.001), KRAS signaling up (Nominal *p* value <0.001, FDR q-value =0.0388, FWER *p* value =0.05), MTORC1 signaling (Nominal *p* value <0.001, FDR q-value = 0.118, FWER *p* value =0.18) (Fig. 2f–h).

PPI Networks Construction and Modules

The PPI networks were constructed by the selected DEGs in the tumor versus normal group from STRING and visualized by Cytoscape (degree ≥1), including 79 nodes and 143 edges (Fig. 3a). The most scored cluster modules were also calculated by MCODE in Cytoscape (Fig. 3b). In module 1, gastric acid secretion was significantly enriched in KEGG with no terms in BP/CC/MF. In module 2, extracellular matrix organization, extracellular structure organization and extracellular matrix disassembly were found significantly enriched in BP with no terms in CC/MF and KEGG. In module 3, collagen catabolic process, multicellular organism catabolic process, and collagen metabolic process were the top three significantly enriched terms in BP while endoplasmic reticulum lumen and collagen trimer in were the significantly enriched terms in CC. However, no significant term was found in MF and KEGG analysis.

Analysis and External Validation of the mRNA Expressions of Hub Genes

The top 10 genes with the highest degrees were selected as hub genes, listed as collagen type I alpha 1 chain (*COL1A1*), tissue inhibitor of the metalloproteinases (*TIMP1*), Secreted Phosphoprotein 1 (*SPP1*), somatostatin (*SST*), neuropeptide Y (*NPY*), biglycan (*BGN*), matrix metallopeptidase 3 (*MMP3*), apolipoprotein E (*APOE*), ATPase H+/K+ transporting alpha subunit (*ATP4A*), lysyl oxidase (*LOX*) (Fig. 4a). The mRNA expressions of the hub genes between tumor and normal tissues were further externally investigated in the STAD of TCGA database (Fig. 4a), GSE13861 gene profile (Fig. 4b), GSE27342 gene profile (Fig. 4c) and qRT-PCT results (Fig. 4d). Moreover, eight independent microarray data (GSE19826, GSE33335, GSE63089, GSE27342, GSE56807, GSE54129, GSE26942 and GSE79973) were further incorporated for the comparison of DEG selection and hub genes in GSE103236. A total of 57 overlapped DEGs had been identified (Fig. 4e). The protein-protein-interaction (PPI) networks of the 57 DEGs indicated that 7 out of 10 hub genes (*COL1A1*, *SPP1*, *TIMP1*, *ATP4A*, *SST*, *MMP3* and *BGN*) were overlapped (Fig. 4f, Table 2).

In fact, the mRNA expressions of *COL1A1, TIMP1, SPP1, BGN, MMP3,* and *APOE* were consistently and significantly increased in tumor compared to normal tissues (Fig. 4a–c). *SST* and *ATP4A* were consistently and significantly reduced in tumor tissues from TCGA, GSE13861, and GSE27342, whereas *NPY* was reduced in tumor only in TCGA and GSE13861 (Fig. 4a–c). Our experimental qRT-PCR results indicated the abnormally increasing of all hub gene expressions (except *ATP4A* with decreased expression) in tumor compared to normal tissues (Fig. 4d). To elucidate the potential mechanism among the hub genes, we illustrated the expression correlation among the hub genes in both GSE103236 and TCGA cohorts. Of note, the highest positive correlation (cor = 0.87) existed between *SPP1* and *LOX* while the

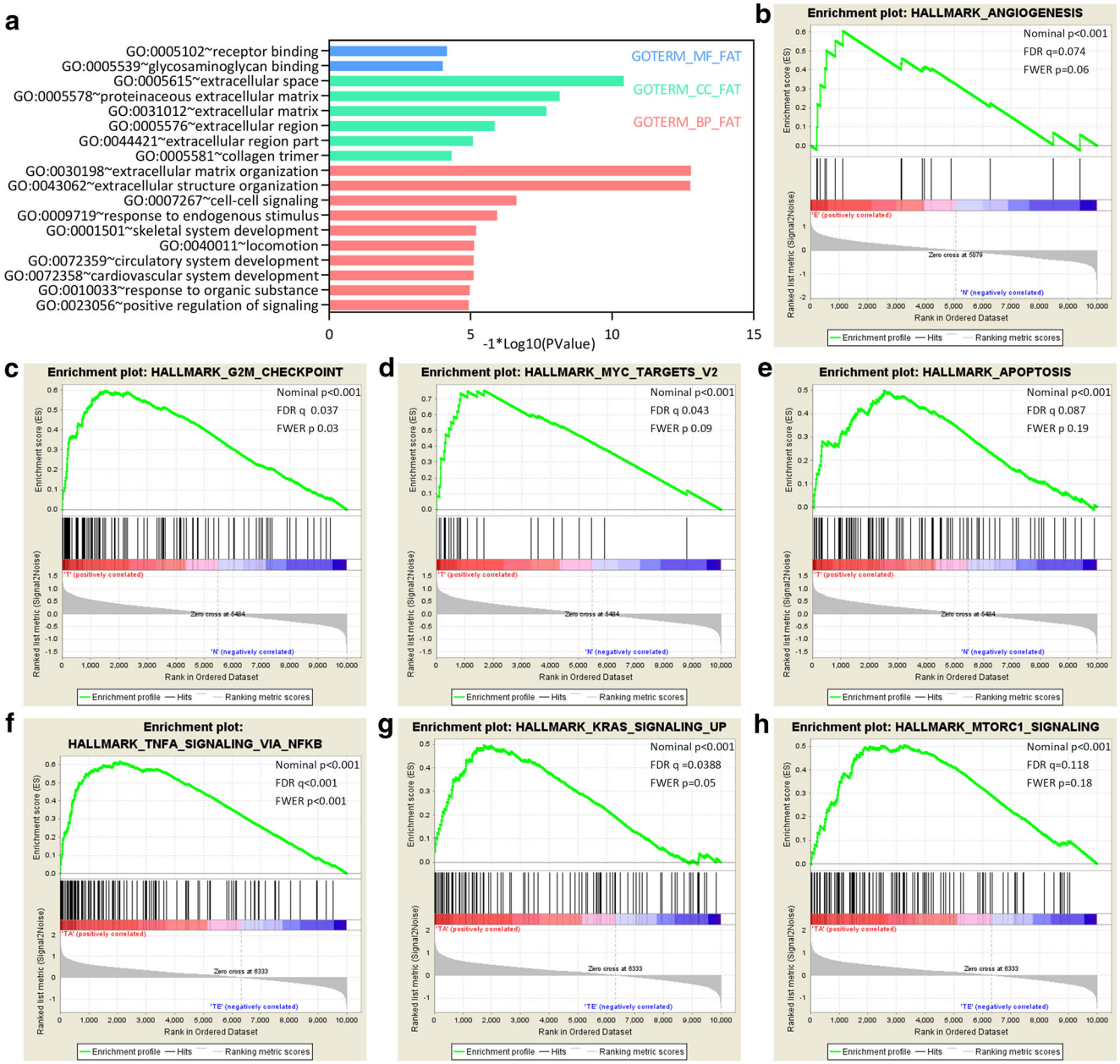

Fig. 2 Gene ontologies (GO) analysis of DEGs and gene sets enrichment analysis (GSEA). **a** Biological processes (BP), cellular component (CC) and molecular function (MF) of gene ontology in total DEGs; **b** GSEA result of primary gastric cancer (PGC) compared to normal; **c**–**e** GSEA results of total gastric cancer compared to normal; **f**–**h** GSEA results of advanced gastric cancer (AGC) compared to PGC

highest negative correlation (cor = −0.53) existed between *BGN* and *ATP4A* in GSE103236 (Fig. 5a). In the STAD of TCGA, the correlations were further validated. Interestingly, the top positive correlations were determined between *COL1A1* and *BGN, COL1A1* and *LOX* while negative correlation was comparably low (Fig. 5b).

Stage-Specific Expressions of the Hub Genes

To further delineate the expression pattern of each hub gene, the stage-specific expressions of hub gene were analyzed in GEPIA. *COL1A1* ($p = 0.0292$), *TIMP1* ($p = 0.000297$), *SPP1* ($p = 0.035$), *BGN* ($p = 0.000292$), *LOX* ($p = 0.00788$), and *APOE* ($p = 1.52e\text{-}05$) exhibited significantly stage-specific expressions (Fig. 6a). Furthermore, 176 (48%) cases had the genomic alterations of hub genes, including missense mutation (unknown significance), truncating mutation (unknown significance), amplification, deep deletion, and mRNA upregulation (Fig. 6b). However, no significant prognostic value was identified between cases with and without hub gene alterations (Fig. 6c, d).

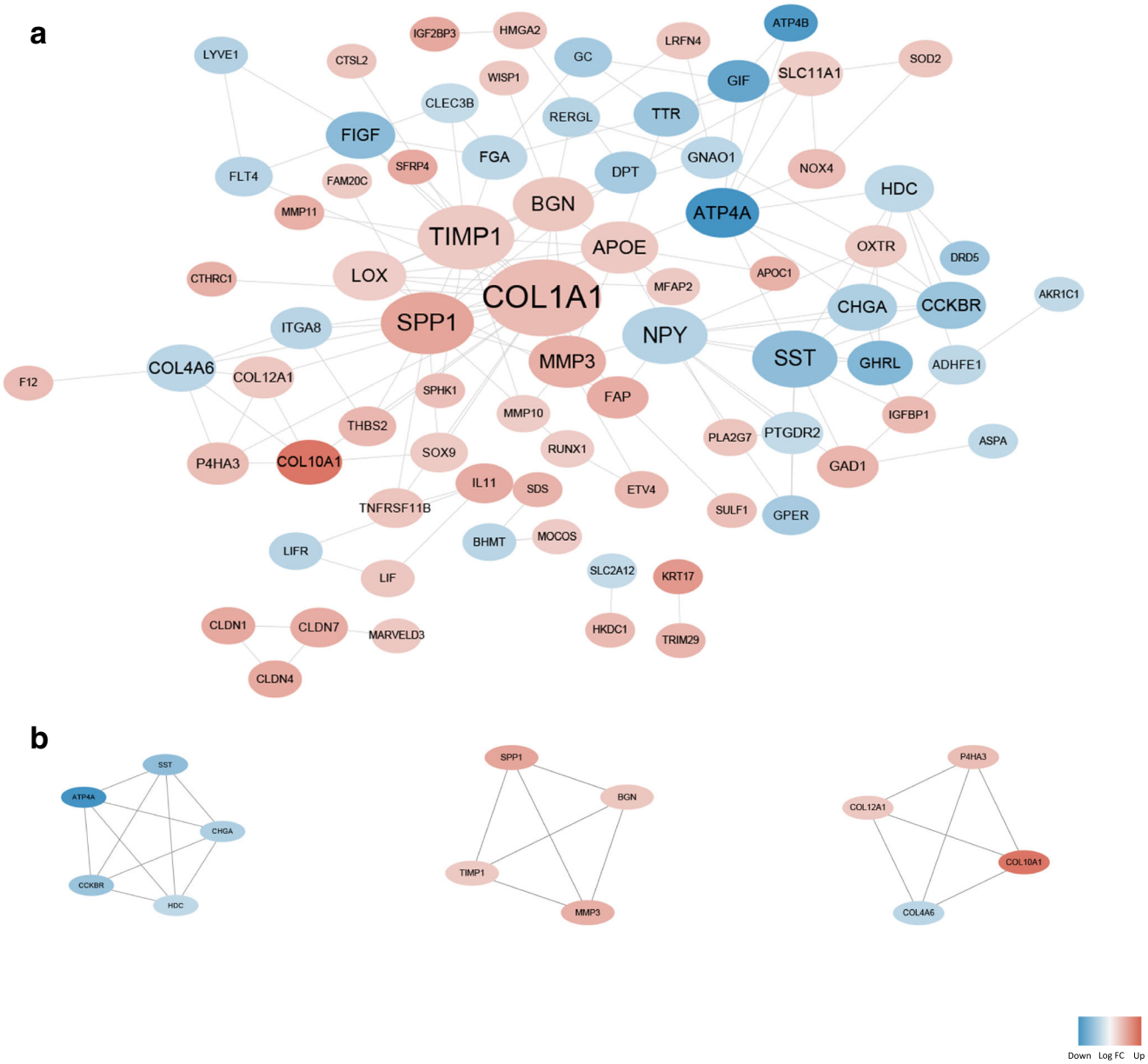

Fig. 3 Protein-protein interaction (PPI) networks of DEGs and top scored modules. **a** PPI network of DEGs with nodes representing genes and lines representing; (*Red*: up-regulation; *blue*: down-regulation) **b** the top three scored modules were shown from up to bottom

Protein Expressions of the Hub Genes

To further determine the protein expressions of the hub genes, the immunohistochemistry results of the hub genes in HPA (www.proteinatlas.org) were investigated. Consistent with the mRNAs results, COL1A1, TIMP1, SPP1, BGN, and APOE had the marked positive expressions in tumor than normal tissues while NPY and ATP4A had the obvious positive expressions in normal rather than tumor tissues. However, MMP3 had strong positive expression in normal rather than tumor tissues, which was contrary to its mRNA expression. Moreover, SST had indistinguishable expressions between normal and tumor tissues (Fig. 7). LOX was not available in HPA.

Survival Analysis of the Hub Genes

The prognostic values of the hub genes were further explored by the web-based tool, KM plotter. For total gastric cancer patients, two of the hub genes significantly associated with poor overall survival (OS) in gastric cancer (*SPP1* (Probe ID: 209875_s_at), log rank $p = 0.0048$, HR = 1.39 [1.1–1.75]; *MMP3* (Probe ID: 205828_ at), log rank $p < 0.0001$, HR = 1.77 [1.44–2.19]) while the rest

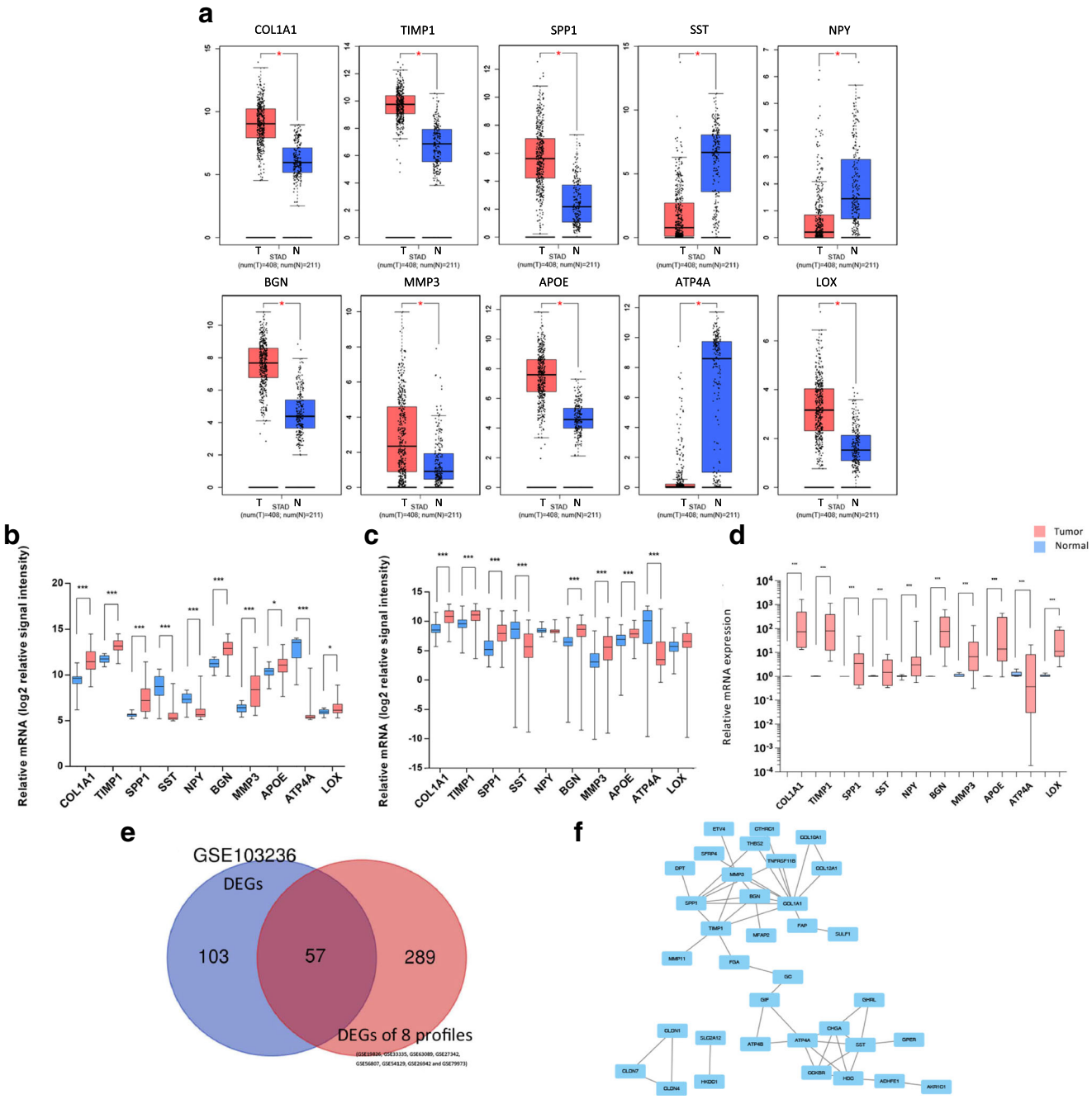

Fig. 4 The mRNA expression of hub genes between tumor and normal tissues. **a** The mRNA expression of ten hub genes in comparison between tumor and normal in STAD of TCGA; **b** the mRNA expression of ten hub genes in GSE13861 between tumor and normal; **c** the mRNA expression of ten hub genes in GSE27342 between tumor and normal; **d** the validation of hub genes by quantitative RT-PCR between gastric cancer and normal tissues; **e** Overlapped DEGs between eight independent microarray data (GSE19826, GSE33335, GSE63089, GSE27342, GSE56807, GSE54129, GSE26942 and GSE79973) and GSE103236; **f** PPI networks of the overlapped DEGs between eight independent microarray data (GSE19826, GSE33335, GSE63089, GSE27342, GSE56807, GSE54129, GSE26942 and GSE79973) and GSE103236. STAD: Stomach Adenocarcinoma; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ Red: tumor tissue; blue: normal tissue

eight hub genes, *COL1A1* (Probe ID: 202311_s_ at), *TIMP1* (Probe ID: 201666_ at), *SST* (Probe ID: 213921_s_ at), *NPY* (Probe ID: 206001_ at), *BGN* (Probe ID: 201261_x_ at), *APOE* (Probe ID: 203382_s_ at), *ATP4A* (Probe ID: 207139_ at), *LOX* (Probe ID: 215446_s_ at) did not show clear associations with prognosis (Fig. 8a–c), which indicating potential prognostic values of *SPP1* and *MMP3* in gastric cancer. Furthermore, the stage-specific prognostic values of hub genes were also investigated (Table 3). Interestingly,

Table 2 Comparison of the hub genes between GSE103236 and overlapped genes list

Hub genes of the overlapped DEGs			Hub genes of GSE103236		
Gene symbol	Gene name	Degree	Gene symbol	Gene name	Degree
COL1A1	Collagen type I alpha 1 chain	11	COL1A1	Collagen type I alpha 1 chain	18
SPP1	Secreted Phosphoprotein 1	7	TIMP1	Tissue inhibitor of the metalloproteinases	13
TIMP1	Tissue inhibitor of the metalloproteinases	6	SPP1	Secreted Phosphoprotein 1	12
ATP4A	ATPase H+/K+ transporting alpha subunit	6	SST	Somatostatin	10
SST	Somatostatin	6	NPY	Neuropeptide Y	10
MMP3	Matrix metallopeptidase 3	5	BGN	Biglycan	9
CHGA	Chromogranin A	5	MMP3	Matrix metallopeptidase 3	8
BGN	Biglycan	5	APOE	Apolipoprotein E	8
HDC	Histidine decarboxylase	5	ATP4A	ATPase H+/K+ transporting alpha subunit	7
CCKBR	Cholecystokinin B receptor	4	LOX	Lysyl oxidase	7

COL1A1 (Probe ID: 202311_s_at) showed significant correlation to prognosis in stage III (log rank $p = 0.048$, HR = 0.72 [0.52–1]) and IV (log rank $p = 0.048$, HR = 0.202 [0.99–4.09]) gastric cancer cohorts. *NPY* (Probe ID: 206001_at) showed significant correlation to prognosis in stage II (log rank $p = 0.0042$, HR = 3.31 [1.39–7.88]) gastric cancer cohort. *BGN* (Probe ID: 201261_x_at) showed significant correlation to prognosis in stage III (log rank $p = 0.0072$, HR = 1.57 [1.13–2.18]) and stage IV (log rank $p = 0.013$, HR = 2.02 [1.15–3.56]) gastric cancer cohorts. *LOX* (Probe ID: 215446_s_at) showed significant correlation to prognosis in stage I (log rank $p = 0.0078$, HR = 0.1 [0.02–0.08]) and stage II (log rank $p = 0.045$, HR = 2.35 [0.99–5.56]) gastric cancer cohorts. Noteworthy, *SPP1* did not exhibit significant correlation to prognosis in any stage cohort, respectively. *MMP3* only exhibited significant prognostic value in stage IV (log rank $p = 0.026$, HR = 0.52 [0.29–0.93]) gastric cancer cohort (Table 3). Furthermore, the DNA methylation levels of *SPP1* and *MMP3*, as well as the

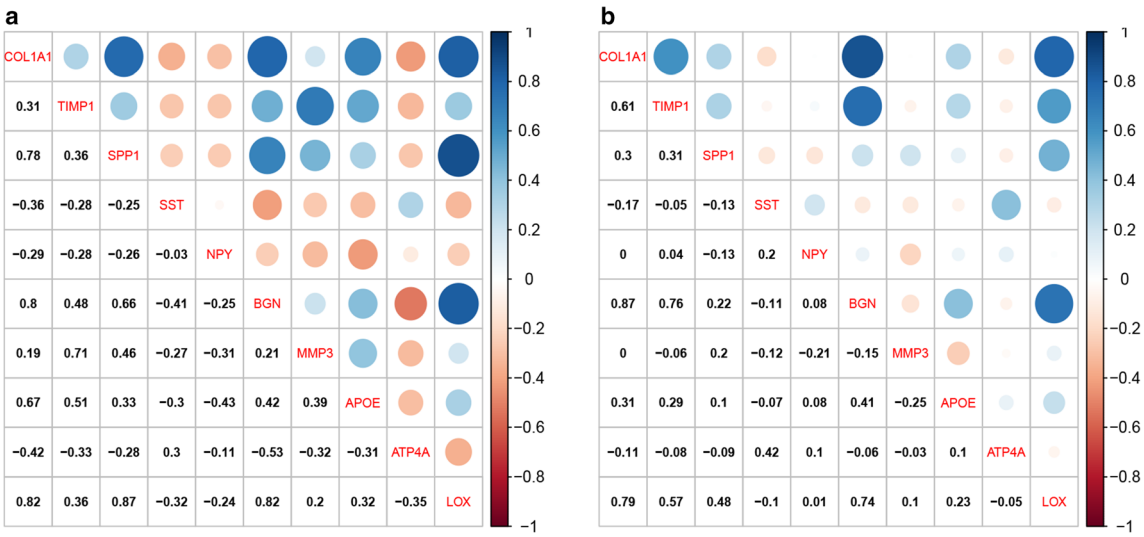

Fig. 5 The correlation of hub gene in GSE103236 and TCGA. **a** The correlation of hub genes by ball plot in GSE103236; **b** the correlation of hub genes by ball plot in TCGA. The red circle indicated a negative correlation, the blue circle indicated a positive correlation. The values of correlation coefficients were represented by the color bar aside. Color intensity and the circle size were proportional to the correlation coefficients

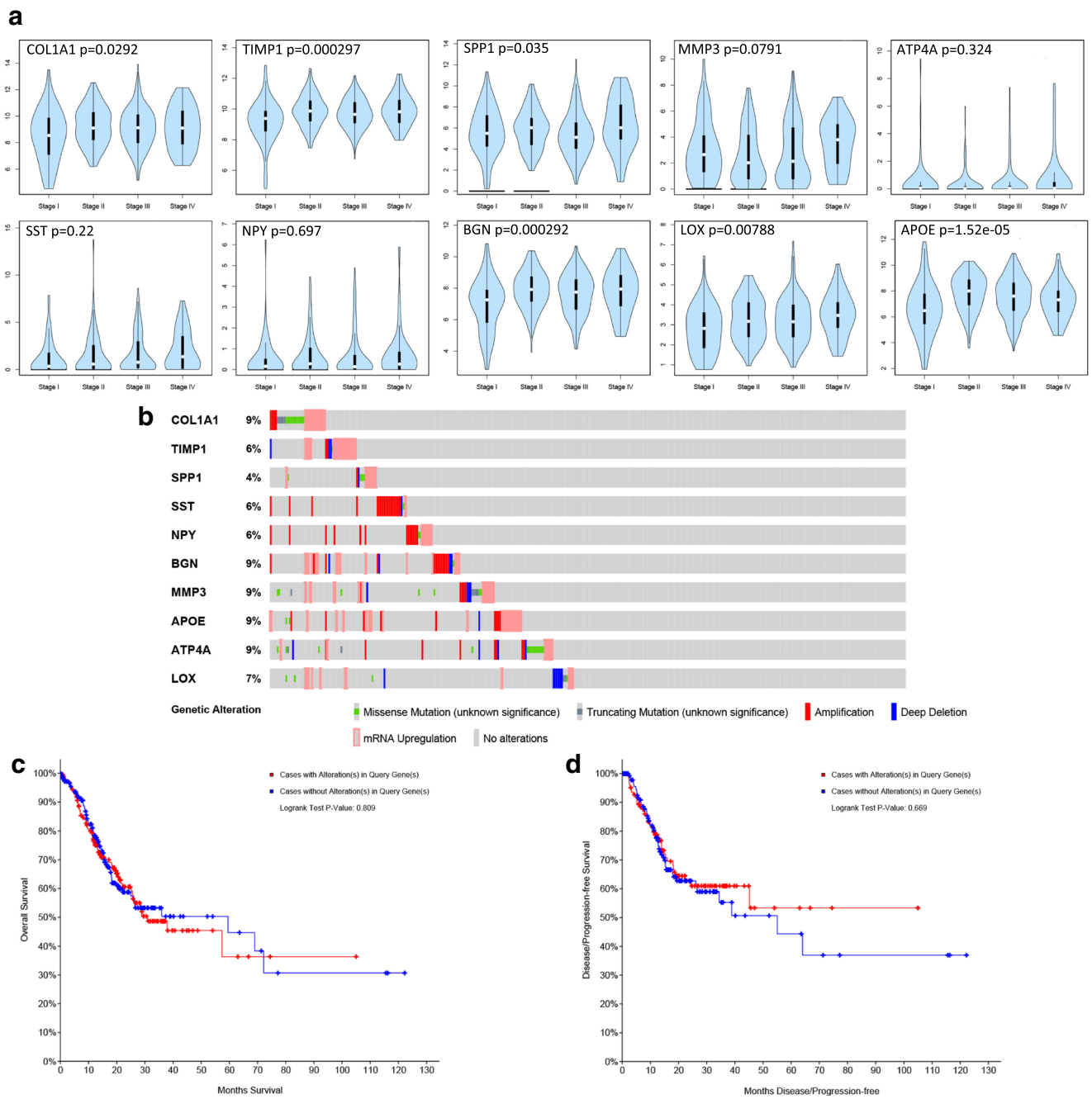

Fig. 6 The stage-specific expressions and genetic alterations of hub genes in TCGA. **a** The stage-specific expressions of hub genes in TCGA; **b** genomic alterations of hub genes in TCGA; **c** overall survival between cases with alterations in hub genes (red) and cases without alterations (blue); **d** disease-free survival between cases with alterations in hub genes (red) and cases without alterations (blue)

corresponding prognostic values of each CpG in TCGA, was retrieved and analyzed by MethSurv (Fig. 8c, d, e). In fact, cg00583003 of *SPP1* exhibited high methylation level and significant prognostic values (HR = 1.625, p = 0.013) (Fig. 8c, e). However, cg17145397, cg16466334 and cg18113270 also showed high methylation levels of *MMP3*, but only cg16466334 presented significant prognostic value (HR = 0.647, $p = 0.011$) (Fig. 8e).

Prognostic Values of the Hub Genes Signature Via SurvExpress

Given the prognostic values of gene signatures have been increasingly noticed, we further evaluated the prognostic values of hub genes-signature via SurvExpress platform. Based on the maximized risk group algorithm, the STAD of TCGA had been divided into high-risk ($n = 93$) and low-risk ($n =$

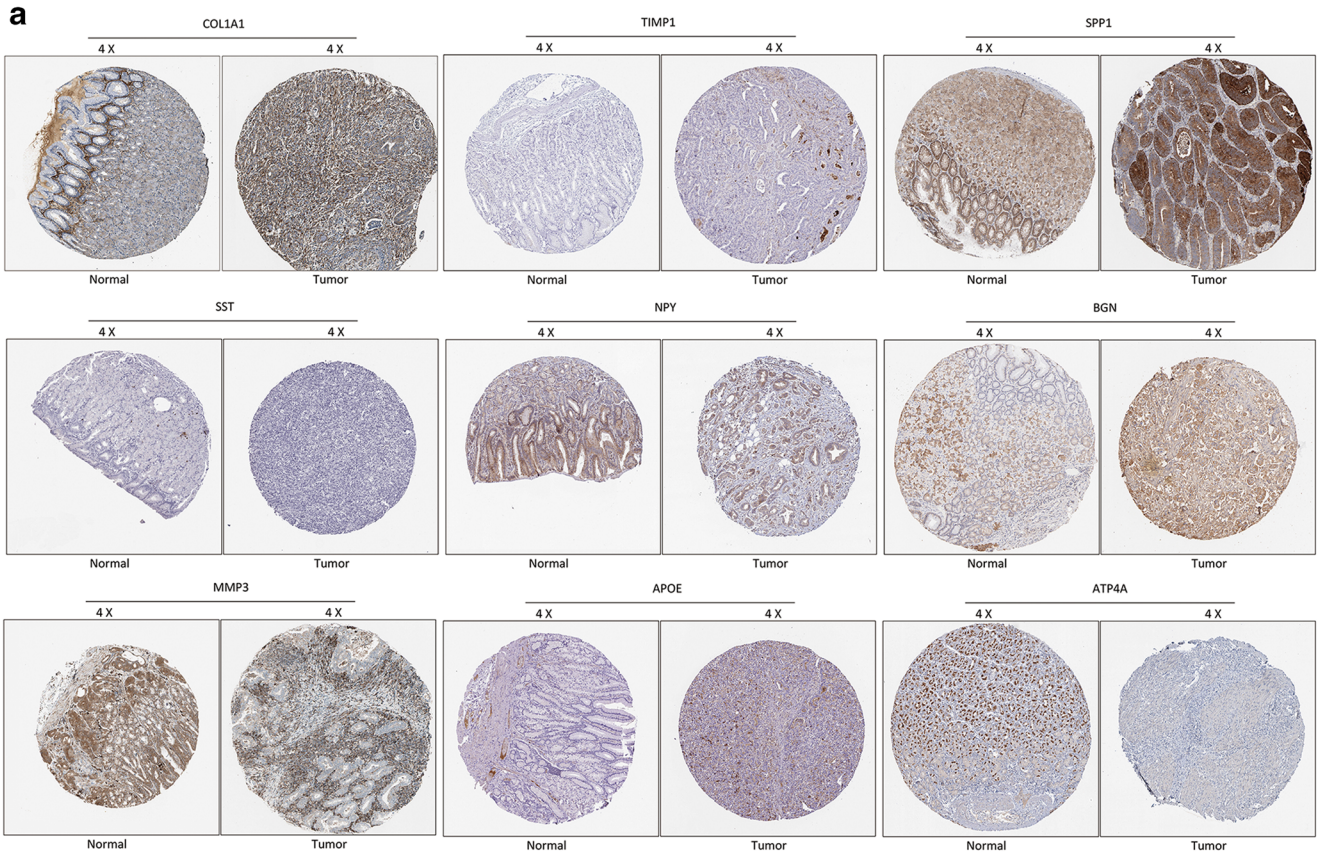

Fig. 7 The protein expression of hub genes. Images were retrieved from the Human Protein Atlas (HPA) (http://www.proteinatlas.org) database. LOX was excluded due to the absent protein results in HPA

259) groups. The low-risk group indicated a favorable clinical outcome (Fig. 9a–c, log-rank p value = 5.227e-05, hazard ratio = 2.04, 95%CI: 1.43–2.89). Moreover, *COL1A1*, *TIMP1*, *SPP1*, *SST*, *BGN*, *MMP3*, *APOE*, *ATP4A*, and *LOX* displayed significantly different expression between high- and low-risk groups (Fig. 9c). Intriguingly, the expression of *COL1A1*, *TIMP1*, *SPP1*, *BGN*, *APOE*, *LOX*, *SST*, and *ATP4A* in the low-risk group were consistent with previous results (Fig. 4a).

The Correlation of Hub Genes and the Tumor Immune Infiltrates

To further provide insights for immunotherapeutic managements in GC, this study analyzed the correlation of hub genes and the tumor immune infiltrates, including B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages and dendritic cells via the TIMER platform. In fact, *APOE* demonstrated the highest correlation with $CD8^+$ T cells (cor = 0.52), neutrophils (cor = 0.51) and dendritic cells (cor = 0.67), whereas *BGN* had the highest correlation with macrophages (cor =0.55) (Fig. 10). Furthermore, *TIMP1* and *LOX* also showed close correlations with several types of immune infiltrates (Fig. 10). Interestingly, given the drastic differential expression of *ATP4A* between normal and tumor tissues (Figs. 4a and 7), the correlation between *ATP4A* and all tumor infiltrates were low (Fig. 10). Moreover, contrary to other nine hub genes, *MMP3* exhibited a comparably negative correlation with tumor infiltrates (Fig. 10).

Discussion

Despite descending incidence and mortality rates, and the well-established endoscopy screening system, gastric cancer remains one of the leading common epithelial malignancies in Eastern Asia and Eastern Europe [28–30]. Specifically, both the tumor locations and pathological stages at diagnosis varied in an area-specific pattern [30, 31]. Gastric cancer patients diagnosed in Japan and Korea mainly are ECGs while AGCs are predominantly found in other areas. Therefore, the mechanisms underlying the tumorigenesis and progression in gastric cancer were among the most intensively investigated healthy issues worldwide.

Despite the fact that the GSE103236 profile had been thoroughly investigated from Chivu et al in 2010 in terms of biomarkers for PGC and AGC [7], many genomic features of this profile were yet to be fully characterized given the reshaped bioinformatics algorithms and resources. Thus, it is worthy of re-annotation and re-analysis by using updated

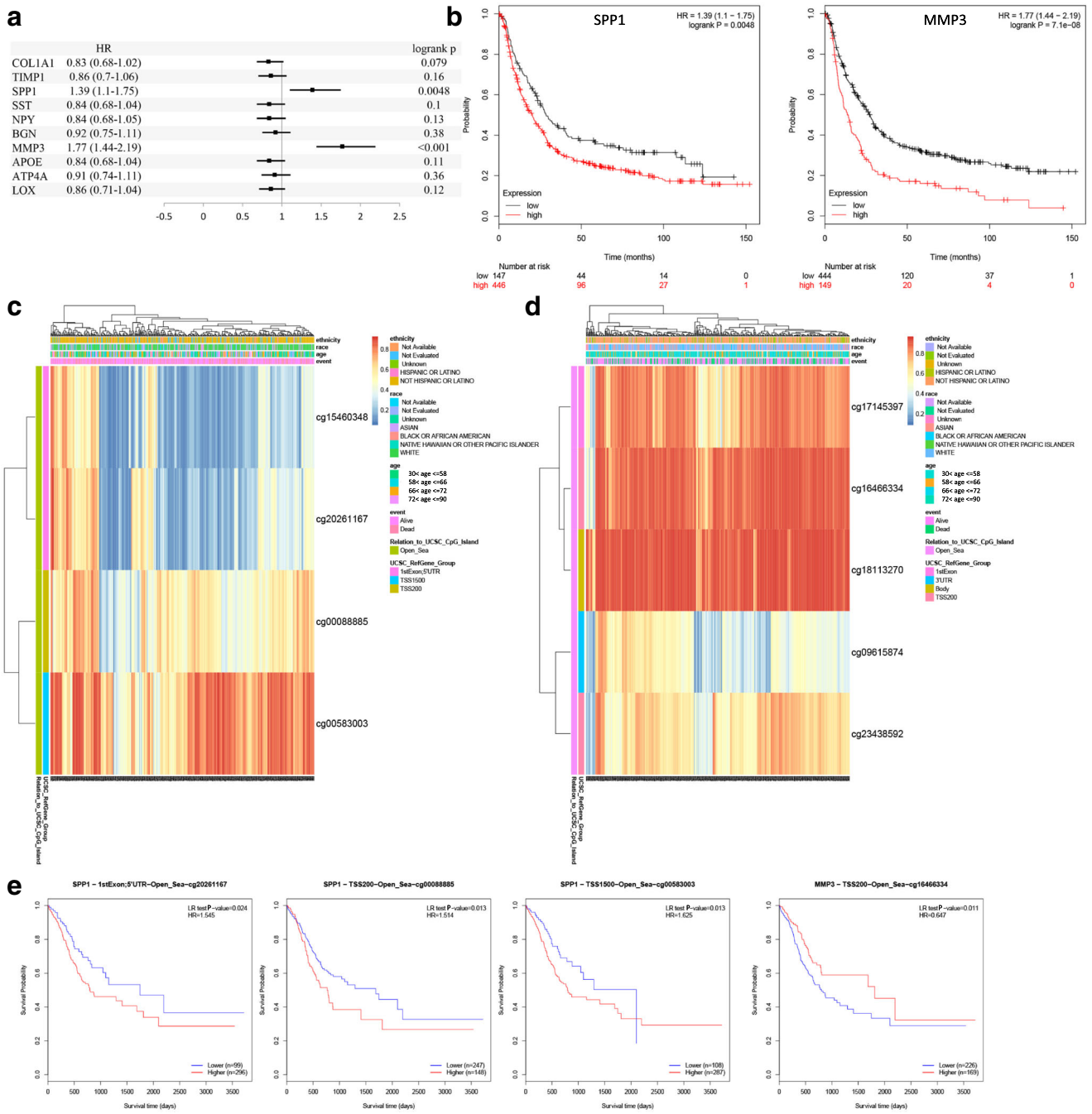

Fig. 8 The prognostic values of hub genes in Kaplan-Meier plotter (KM plotter). **a** The prognostic values of hub genes; **b** the survival curves of *SPP1* (Probe ID: 209875_s_at) and the survival curves of *MMP3* (Probe ID: 205828_ at); **c** heat map of the CpG methylation levels of *SPP1* gene (red represented fully methylated, blue represented fully unmethylated). Each CpG was shown in rows and patient in columns. **d** Heat map of the CpG methylation levels of *MMP3* gene (red represented fully methylated, blue represented fully unmethylated). **e** KM plots for cg20261167-*SPP1*, cg00088885-*SPP1*, cg00583003-*SPP1* and cg16466334-*MMP3* using STAD samples dichotomized by mean values of DNA methylation. HR: hazard ratio; STAD: stomach adenocarcinoma; KM: Kaplan-Meier; LR: log-likelihood ratio

bioinformatics analytic strategy to identify potential key target genes and signaling pathways for tumorigenesis and progression in gastric cancer [14–17, 22].

In this study, only the DEGs generated by the comparison of tumor and normal tissues were chosen for further GO/KEGG analysis and PPI network due to the limited number of DEGs found in other comparisons. Most significant enriched BP/CC/MF terms were the extracellular matrix-related. Noteworthy, no significant term was identified as significant enrichment in KEGG. Collectively, it indicated a close

Table 3 The stage-specific overall survival of hub genes based on Kaplan-Meier plotter (KM plotter)

Gene	Probe	Stage I HR (95%CI)	p	Stage II HR (95%CI)	p	Stage III HR (95%CI)	p	Stage IV HR (95%CI)	p
COL1A1	202311_s_at	0.25(0.06–1.15)	0.055	2.08(0.87–4.96)	0.092	0.72(0.52–1)	0.048	2.02(0.99–4.09)	0.048
TIMP1	201666_at	0.45(0.12–1.65)	0.22	0.44(0.17–1.16)	0.088	1.37(0.97–1.94)	0.075	1.65(0.92–2.95)	0.088
SPP1	209875_s_at	0.39(0.11–1.45)	0.15	0.6(0.25–1.46)	0.26	1.35(0.98–1.85)	0.069	2.03(0.95–4.34)	0.062
SST	213921_at	2.01(0.65–6.22)	0.22	1.8(00.73–4.44)	0.2	0.82(0.57–1.17)	0.27	0.65(0.35–1.22)	0.18
NPY	206001_at	2.03(0.68–6.08)	0.2	3.31(1.39–7.88)	0.0042	0.81(0.57–1.15)	0.23	0.71(0.4–1.24)	0.22
BGN	201261_x_at	0.38(0.12–1.23)	0.094	2.39(0.93–6.14)	0.063	1.57(1.13–2.18)	0.0072	2.02(1.15–3.56)	0.013
MMP3	205828_at	2.47(0.68–9)	0.16	1.6(0.65–3.94)	0.3	0.75(0.53–1.05)	0.095	0.52(0.29–0.93)	0.026
APOE	203382_s_at	0.43(0.14–1.33)	0.13	1.71(0.74–3.98)	0.21	1.27(0.88–1.83)	0.2	1.38(0.78–2.44)	0.27
ATP4A	207139_at	0.49(0.16–1.45)	0.19	0.64(0.27–1.52)	0.31	0.81(0.57–1.16)	0.25	0.74(0.42–1.32)	0.31
LOX	215446_s_at	0.1(0.01–0.8)	0.0078	2.35(0.99–5.56)	0.045	1.17(0.85–1.63)	0.34	1.71(0.97–2.99)	0.059

relationship between gastric cancer and the extracellular matrix-related signaling. In fact, the disruption in intercellular adhesion and the degradation in extracellular matrix enabled the initial locomotion of cancer cells featured by over-activated invasiveness and metastases [32]. Interestingly, this finding was also consistent with the subsequent PPI network and top hub genes.

The top 10 hub genes, *COL1A1*, *TIMP1*, *SPP1*, *BGN*, *MMP3*, *APOE*, *LOX*, *SST*, *NPY*, and *ATP4A*, were determined. Among them, *SPP1* and *MMP3* were significantly associated with poor OS in gastric cancer. Interestingly, compared to the hub genes reported by Chivu et al., *SPP1* and *MMP3* were also the only two overlapped with consistent results in mRNA expressions [7]. Next, the mRNA expressions of the hub

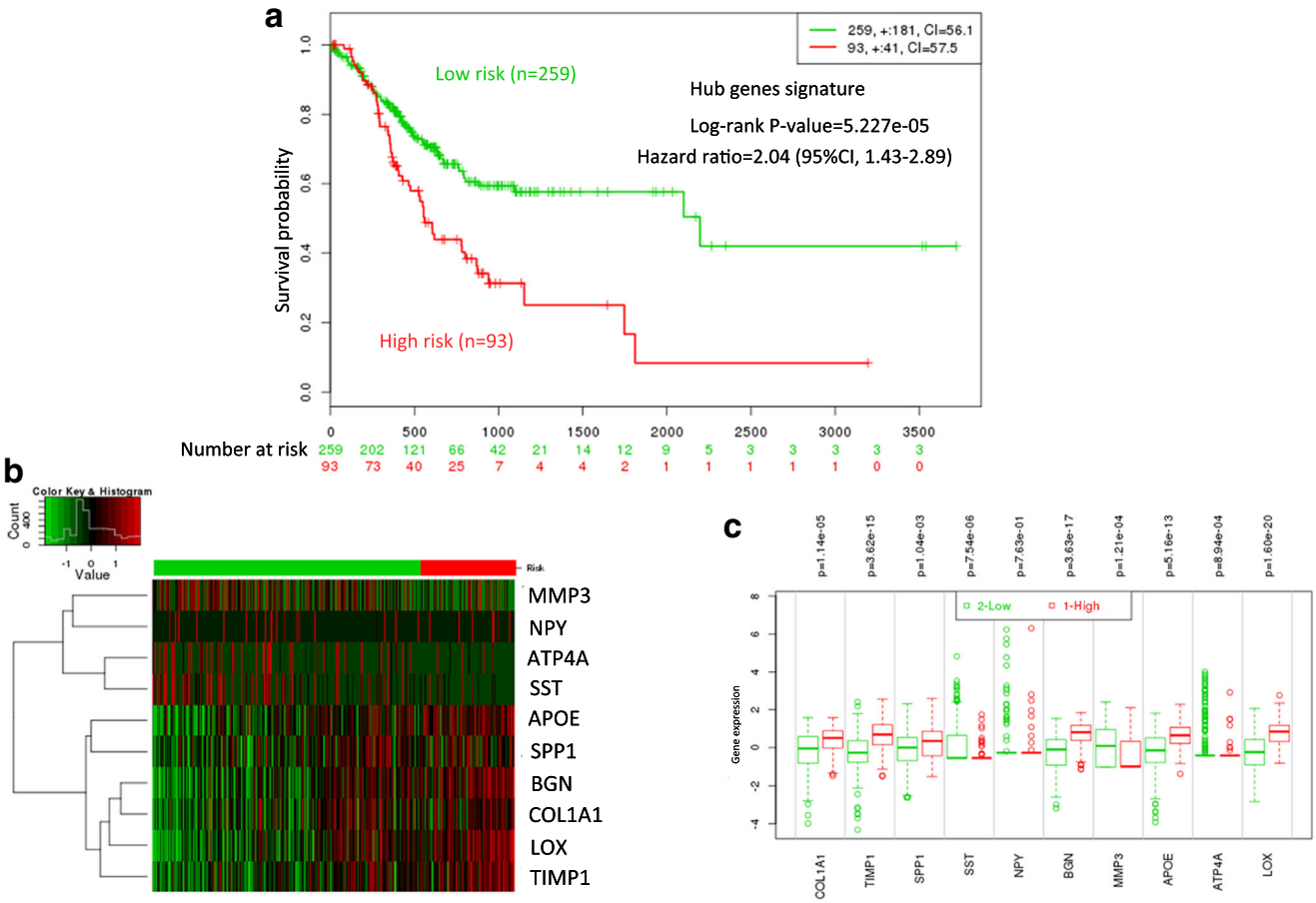

Fig. 9 Prognostic analysis of the hub genes signature via SurvExpress platform. **a** The Kaplan-Meier plot of the hub genes signature between high risk ($n = 93$, red) and low risk ($n = 259$, green) patients in TCGA; **b** the heat map of mRNA expression of the hub genes in TCGA divided into high risk (red) and low risk (green); **c** the mRNA expression of hub genes between high risk and low risk groups

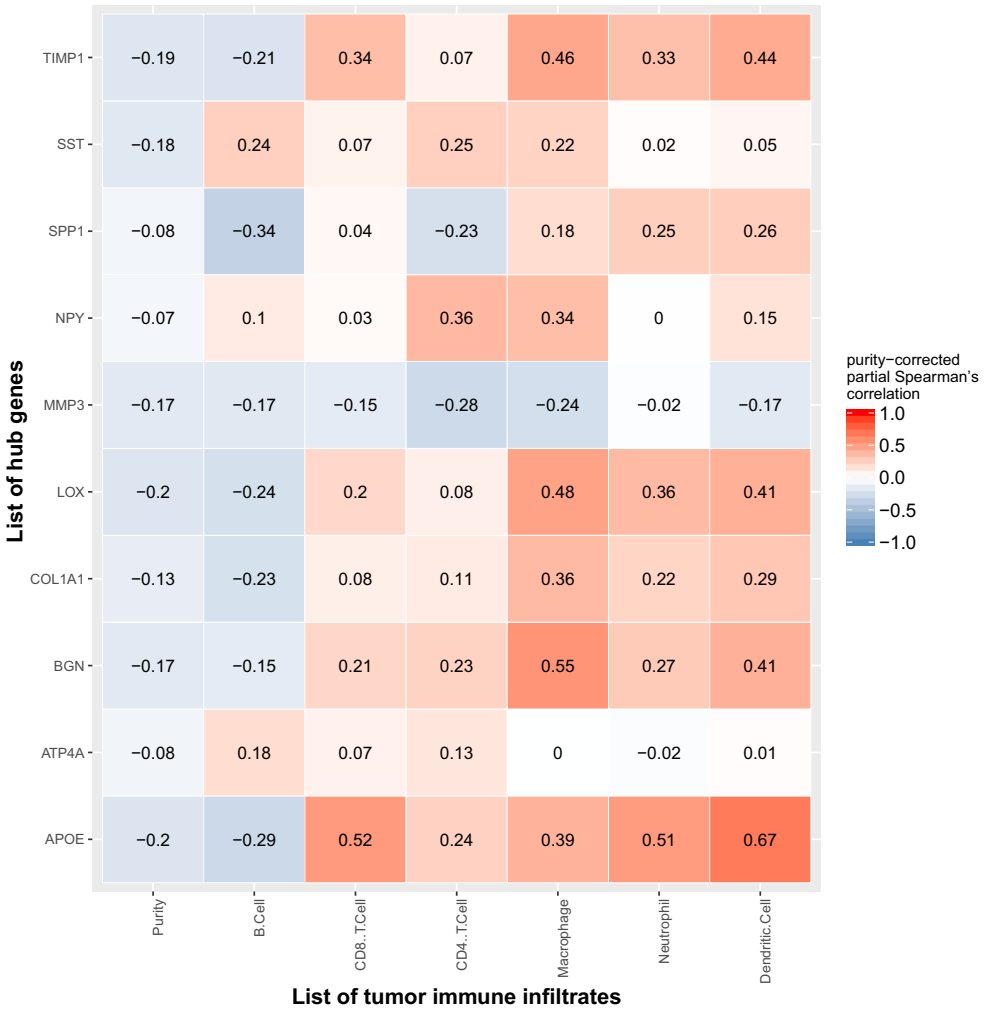

Fig. 10 The correlation of hub genes and tumor immune infiltrates. The correlation of hub genes and tumor immune infiltrates (B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells) was analyzed via the Tumor Immune Estimation Resource (TIMER) platform (https://cistrome.shinyapps.io/timer/). Red: positive correlation; blue: negative correlation

genes were investigated by TCGA and validated by two external gene expression profiles. Furthermore, in GSE103236, the potential pairwise correlations of the hub genes were determined and further validated by TCGA for large sample size. The correlations of *COL1A1* and *BGN*, *COL1A1* and *LOX*, *LOX* and *BGN* were consistently high between two profiles. In fact, TCGA data helped filtering out the significant negative correlations found in GSE103236.

Remarkably, the hub genes identified in GSE103236 could be reproducible and solid for cross-datasets validation and comparable quality of GSE103236 profile. Of note, the hub genes identified from the 57 DEGs of the eight independent microarrays were featured by a comparably less degrees than this manuscript (Table 2), which could be a potential cut-off confounding bias. In fact, clinical heterogeneity may also exist in each independent genomic profile deposited in the GEO. Theoretically, it is more convincing to combine genomic profiles for the DEG selection in order to pinpoint those commonly significantly dysregulated genes. However, this may lead to a comparable small amount of shared DEGs, potentially limited the subsequent bioinformatics enrichment and predictions. Meanwhile, some race-specific or clinical phenotype-specific genes may be involved. Moreover, to our knowledge, GSE103236 has not been fully explored in bioinformatics analysis. This is the first study focusing on the genomic profile of GSE103236 with multiple bioinformatics strategies compared to the original study [7].

SPP1 is primarily found in malignant pleural mesothelioma and bone giant cell tumor with a close association in focal adhesion, phospholipase-C pathway and extracellular matrix binding [33, 34]. *SPP1* had been intensively studied as a potential biomarker for gastric cancer [35]. Prior studies had demonstrated the expression of *SPP1* was altered in tumor tissues but did not affect the prognosis of gastric cancer [36]. Noteworthy, inconsistency with prior study, our study indicated that higher expression of *SPP1* was associated with poor prognosis, possibly due to the racial variance.

MMP3 was widely involved in both malignant and inflammation-related diseases, with close related pathways of degradation of the extracellular matrix, cell-matrix glycolconjugation and cell adhesion [37–39]. Remarkably, *MMP3* had been identified as a serum biomarker for the prognosis of

gastric cancer [40] and was negatively associated with the different stages [41]. However, the prognostic value of mRNA expression of *MMP3* in gastric cancer remains sparse. To our knowledge, this study provided a novel prognostic value analysis of *MMP3* based on public integrative resources. Although the protein expression of *MMP3* was contrary to the mRNA expression in this study based on HPA, it was partially because of limited samples used in HPA compared to large samples in mRNA expressions. Conspicuously, *MMP3* exhibited a comparably negative correlation with tumor infiltrates, which distincted from the other hub genes. Further study would focus on the clinic-pathological relevance and prognostic values of *MMP3* in gastric cancer with large samples as well as its potential role in immunotherapies.

Despite the significant prognostic values of *SPP1* and *MMP3*, the pairwise correlations of the hub genes indicated the key roles of *APOE*, *BGN*, *TIMP1*, *LOX*, and *COL1A1* in network regulation, rather than *SPP1* and *MMP3*. These different results also provide insights that multiple genes and signaling networks might be significantly involved in the multi-levels of tumorigenesis of gastric cancer. Intriguingly, *APOE* demonstrated the highest correlation with $CD8^+$ T cells, neutrophils, and dendritic cells whereas *BGN* had the highest correlation with macrophages. *TIMP1* and *LOX* also showed high correlation with several types of immune infiltrates. These results opened to another direction that *APOE*, *BGN*, *TIMP1* and *LOX* may be closely involved in the immune-regulation system in GC.

In addition, besides the analysis of DEGs of tumor versus normal set, this study also provided insightful GSEA results. Five gene sets were enriched in tumor compared to normal tissues, including G2M checkpoint, Myc targets V2, apoptosis, UV response up and angiogenesis. G2M checkpoint had previously been defined as an essential self-repair and DNA damage checkpoint program for eukaryotic cell cycle [42]. In consistent with prior investigations, G2M checkpoint was evidenced to be a major gene sets enrichment feature in this study, offering clues for further exploitation. Meanwhile, only angiogenesis was listed as a significant feature enriched in PGC compared to normal subgroups, indicating a potential role of angiogenesis in initiation stage of GC. Interestingly, comparing to the none DEGs results from the screen strategy (cut-off values of adjusted p value and $\log_2$FC), there were eight gene sets which were determined as significantly enriched in AGC compared to PGC, including TNFA signaling via NFKB, KRAS signaling up, MTORC1 signaling, inflammatory response, glycolysis, angiogenesis, apical junction, and hypoxia. These results indicated dynamic genomic alterations during the tumor progression between PGC and AGC and highlighted the individual features of GSEA and conventional screen strategy. Therefore, multiple bioinformatics analytic approaches were encouraged for sound annotation and analysis.

The genomic features, molecular stratification, and biomarkers of gastric cancer had been previously valued in a series of high-quality investigations [5–7]. TCGA described comprehensive molecular characteristics in 295 GC samples with four subtypes: a. tumors positive for Epstein-Barr virus (EBV); b. microsatellite unstable tumors (MSI); c. genomically stable tumors (GS); d. tumors with chromosomal instability (CIN) [5]. Moreover, Sun et al. investigated key genes in gastric cancer with GSE54129 profile, with ECM-receptor listed as one of the key findings. However, the prognostic roles of *BGN* and *COL1A1* were contradictory between Sun et al. and this study. It is because the GSE62254, which featured markedly prognostic values compared to other datasets, was excluded from prognostic analysis in this study while not in Sun et al. [43]. Meanwhile, Guo et al. and Rong et al. highlighted the roles of *COL1* family genes and *SPP1* in separate bioinformatics analysis in GSE29722, GSE27342 and GSE31789, which were consistent with our results [44, 45].

This study, based on both GSE genome files and TCGA, has multi-dimensionally uncovered novel features in progression of normal-PGC-AGC, identifing potential biomarkers in the networks and prognostic senses.

The limitation of this study was a comparably smaller size of AGC samples. Meanwhile, potential geological and racial characteristics in samples might also be other significant confounding factors involved. In addition, the protein expressions of the hub genes in larger clinical samples for clinical relevance and prognosis are further required.

Conclusion

This study systematically identified potential key genes and pathways involved in primary and advanced gastric cancer, further providing insightful targets for stage-specific therapeutic management.

Acknowledgements We would like to thank Dr. Meng-Kai Ge (Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Department of Pathophysiology, Shanghai Jiao Tong University School of Medicine) for experimental assistance. We would like to thank the academic supports from the biobank of Shanghai Minimally Invasive Surgery Center at Ruijin Hospital. We would like to thank Jiexuan Wang (Ruijin Hospital, Shanghai Jiao Tong University School of Medicine) for his contribution in samples preparation.

Author's Contributions CY, JC and JM carried out experiments and data analysis;
CY, JS, JC, LZ, and MZ drafted the manuscript;
CY, JS, LZ, and MZ participated in study design and data collection;
All authors read and approved the final manuscript.

Funding The study is financially supported by National Natural Science Foundation of China (NSFC) (81402423, 81572818, 81871984), Shanghai Municipal Commission of Health and Family Planning

(2017YQ062), as well as Shanghai Science and Technology Committee (18695841400).

Data Availability The datasets supporting the conclusion of this article were included within the article.

Compliance with Ethical Standards

Conflict of Interest All authors declare no conflict of interest in this study.

Ethics Approval and Consent to Participate All the subjects have given their written informed consent. The study protocol has been approved by the research institute's committee on human research. No animal experiment is applicable.

Consent for Publication Not applicable.

Abbreviations GC, Gastric cancer; DEGs, Differentially expressed genes; PGC, Primary gastric cancer; AJCC, American Joint Committee on Cancer; AGC, Advanced gastric cancer; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; GSEA, Gene set enrichment analysis; PPI, Protein-protein interaction; FC, Fold changes; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DAVID, Database for Annotation, Visualization and Integrated Discovery; FDR, False discovery rate; STRING, Search Tool for the Retrieval of Interacting Genes; MCODE, Molecular Complex Detection; GEPIA, Gene Expression Profiling Interactive Analysis; GTEx, Genotype-tissue expression; STAD, Stomach adenocarcinoma; HPA, Human protein atlas; qRT-PCR, Quantitative real-time PCR; KM, Kaplan-Meier; OS, Overall survival; HR, Hazard ratio; 95% CI, 95% confidence intervals; TIMER, Tumor IMmune Estimation Resource; BP, Biological process; CC, Cellular component; MF, Molecular function; COL1A1, Collagen type I alpha 1 chain; TIMP1, Tissue inhibitor of the metalloproteinases; SPP1, Secreted Phosphoprotein 1; SST, Somatostatin; NPY, Neuropeptide Y; BGN, Biglycan; MMP3, Matrix metallopeptidase 3; APOE, Apolipoprotein E; ATP4A, ATPase H+/K+ transporting alpha subunit; LOX, Lysyl oxidase

References

1. Siegel RL, Miller KD, Jemal A (2016) Cancer statistics, 2016[J]. CA Cancer J Clin 66(1):7–30
2. Chen W, Zheng R, Baade P D, et al. (2015) Cancer statistics in China[J]
3. Ajani JA, Bentrem DJ, Besh S, D'Amico TA, Das P, Denlinger C, Fakih MG, Fuchs CS, Gerdes H, Glasgow RE, Hayman JA, Hofstetter WL, Ilson DH, Keswani RN, Kleinberg LR, Korn WM, Lockhart AC, Meredith K, Mulcahy MF, Orringer MB, Posey JA, Sasson AR, Scott WJ, Strong VE, Varghese TK Jr, Warren G, Washington MK, Willett C, Wright CD, McMillian N, Sundar H, National Comprehensive Cancer Network (2013) Gastric cancer, version 2.2013[J]. J Natl Compr Cancer Netw 11(5):531–546
4. Edge SB, Compton CC (2010) The American joint committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM[J]. Ann Surg Oncol 17(6):1471–1474
5. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do IG, Liu S, Gong L, Fu J, Jin JG, Choi MG, Sohn TS, Lee JH, Bae JM, Kim ST, Park SH, Sohn I, Jung SH, Tan P, Chen R, Hardwick J, Kang WK, Ayers M, Hongyue D, Reinhard C, Loboda A, Kim S, Aggarwal A (2015) Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes[J]. Nat Med 21(5):449–456
6. Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of gastric adenocarcinoma[J]. Nature 513(7517):202–209
7. Chivu EM, Necula LG, Dragu D et al (2010) Identification of potential biomarkers for early and advanced gastric adenocarcinoma detection[J]. Hepato-Gastroenterol 57(104):1453
8. Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository[J]. Nucleic Acids Res 30(1):207–210
9. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A (2013) NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 41(Database issue):D991-5
10. Davis S, Meltzer PS (2007) GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor[J]. Bioinformatics 23(14):1846–1847
11. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A, Bacic A (2015) FunRich: an open access standalone functional enrichment and interaction network analysis tool. Proteomics 15(15):2597–2601
12. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology[J]. Nat Genet 25(1):25–29
13. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes[J]. Nucleic Acids Res 28(1):27–30
14. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources[J]. Nat Protoc 4(1):44–57
15. Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles[J]. Proc Natl Acad Sci 102(43): 15545–15550
16. Szklarczyk D, Franceschini A, Wyder S et al (2014) STRING v10: protein–protein interaction networks, integrated over the tree of life[J]. Nucleic Acids Res 43(D1):D447–D452
17. Tang Z, Li C, Kang B, et al. (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses[J]. Nucleic Acids Res
18. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N (2012) The cBio Cancer genomics portal: an open platform for exploring multidimensional Cancer genomics data. Cancer Discov 2(5):401–404
19. Gao J et al (2013) Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal 6(269):pl1–pl1
20. Cho JY, Lim JY, Cheong JH, Park YY, Yoon SL, Kim SM, Kim SB, Kim H, Hong SW, Park YN, Noh SH, Park ES, Chu IS, Hong WK, Ajani JA, Lee JS (2011) Gene expression signature-based prognostic risk score in gastric cancer. Clin Cancer Res 17(7):1850–1857
21. Cui J, Chen Y, Chou WC, Sun L, Chen L, Suo J, Ni Z, Zhang M, Kong X, Hoffman LL, Kang J, Su Y, Olman V, Johnson D, Tench DW, Amster IJ, Orlando R, Puett D, Li F, Xu Y (2011) An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer. Nucleic Acids Res 39(4):1197–1207
22. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, Wernerus

H, Björling L, Ponten F (2010) Towards a knowledge-based human protein atlas. Nat Biotechnol 28(12):1248–1250
23. Spandidos A, Wang X, Wang H, Seed B (2010) PrimerBank: a resource of human and mouse PCR primer pairs for gene expression detection and quantification. Nucleic Acids Res 38:D792–D799
24. Lanczky A, Nagy A, Bottai G, Munkacsy G, Paladini L, Szabo A, Santarpia L, Gyorffy B (2016) miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2,178 breast cancer patients. Breast Cancer Res Treat 160(3):439–446
25. Modhukur V, Iljasenko T, Metsalu T, Lokk K, Laisk-Podar T, Vilo J (2018 Mar) MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data. Epigenomics. 10(3):277–288
26. Aguirre-Gamboa R, Gomez-Rueda H, Martínez-Ledesma E, Martínez-Torteya A, Chacolla-Huaringa R, Rodriguez-Barrientos A, Tamez-Pena JG, Trevino V (2013) SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. PLoS One 8(9):e74250
27. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS (2017 Nov 1) TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res 77(21):e108–e110
28. Hwang SH, Park DJ, Jee YS, Kim MC, Kim HH, Lee HJ, Yang HK, Lee KU (2009) Actual 3-year survival after laparoscopy-assisted gastrectomy for gastric cancer[J]. Arch Surg 144(6):559–564
29. Hu Y, Ying M, Huang C, Wei H, Jiang Z, Peng X, Hu J, du X, Wang B, Lin F, Xu J, Dong G, Mou T, Li G, Chinese Laparoscopic Gastrointestinal Surgery Study (CLASS) Group (2014) Oncologic outcomes of laparoscopy-assisted gastrectomy for advanced gastric cancer: a large-scale multicenter retrospective cohort study from China[J]. Surg Endosc 28(7):2048–2056
30. Rugge M, Fassan M, Graham D Y (2015) Epidemiology of gastric cancer[M]//gastric Cancer. Springer International Publishing, 23–34
31. Strong VE, Song KY, Park CH, Jacks LM, Gonen M, Shah M, Coit DG, Brennan MF (2010) Comparison of gastric cancer survival following R0 resection in the United States and Korea using an internationally validated nomogram[J]. Ann Surg 251(4):640–646
32. Coussens LM, Werb Z (2002) Inflammation and cancer[J]. Nature 420(6917):860–867
33. Liaw L, Birk DE, Ballas CB, Whitsitt JS, Davidson JM, Hogan BL (1998) Altered wound healing in mice lacking a functional osteopontin gene (spp1)[J]. J Clin Investig 101(7):1468–1478
34. Spiegel S, Milstien S (2003) Sphingosine-1-phosphate: an enigmatic signalling lipid[J]. Nat Rev Mol Cell Biol 4(5):397–407
35. Junnila S, Kokkola A, Mizuguchi T, Hirata K, Karjalainen-Lindsberg ML, Puolakkainen P, Monni O (2010) Gene expression analysis identifies over-expression of CXCL1, SPARC, SPP1, and SULF1 in gastric cancer[J]. Genes Chromosom Cancer 49(1):28–39
36. Zhuo C, Li X, Zhuang H, Tian S, Cui H, Jiang R, Liu C, Tao R, Lin X (2016) Elevated THBS2, COL1A2, and SPP1 expression levels as predictors of gastric Cancer prognosis[J]. Cell Physiol Biochem 40(6):1316–1324
37. Sternlicht MD, Lochter A, Sympson CJ, Huey B, Rougier JP, Gray JW, Pinkel D, Bissell MJ, Werb Z (1999) The stromal proteinase MMP3/stromelysin-1 promotes mammary carcinogenesis[J]. Cell 98(2):137–146
38. Kubota E, Imamura H, Kubota T et al (1997) Interleukin 1β and stromelysin (MMP3) activity of synovial fluid as possible markers of osteoarthritis in the temporomandibular joint[J]. J Oral Maxillofac Surg 55(1):20–27
39. Stamenkovic I (2003) Extracellular matrix remodelling: the role of matrix metalloproteinases[J]. J Pathol 200(4):448–464
40. Yeh YC, Sheu BS, Cheng HC, Wang YL, Yang HB, Wu JJ (2010) Elevated serum matrix metalloproteinase-3 and-7 in H. pylori-related gastric cancer can be biomarkers correlating with a poor survival[J]. Dig Dis Sci 55(6):1649–1657
41. Xu J, Yao Y, Ren S et al (2016) Matrix metalloproteinase expression and molecular interaction network analysis in gastric cancer[J]. Oncol Lett 12(4):2403–2408
42. Fernandez-Capetillo O, Chen HT, Celeste A, Ward I, Romanienko PJ, Morales JC, Naka K, Xia Z, Camerini-Otero RD, Motoyama N, Carpenter PB, Bonner WM, Chen J, Nussenzweig A (2002) DNA damage-induced G2–M checkpoint activation by histone H2AX and 53BP1[J]. Nat Cell Biol 4(12):993–997
43. Sun C, Yuan Q, Wu D et al (2017) Identification of core genes and outcome in gastric cancer using bioinformatics analysis[J]. Oncotarget 8(41):70271
44. Guo L, Song C, Wang P, Dai L, Zhang J, Wang K (2015) A systems biology approach to detect key pathways and interaction networks in gastric cancer on the basis of microarray analysis[J]. Mol Med Rep 12(5):7139–7145
45. Rong L, Huang W, Tian S, Chi X, Zhao P, Liu F (2018) COL1A2 is a novel biomarker to improve clinical prediction in human gastric Cancer: integrating bioinformatics and meta-analysis[J]. Pathol Oncol Res 24(1):129–134