Original article

Identification of methylated-differentially expressed genes and pathways in esophageal squamous cell carcinoma

Liang Sang[a,e,1], Zhanwu Yu[d,1], Ang Wang[a,b,c], Hao Li[a,b,c], Xiantong Dai[a,b,c], Liping Sun[a,b,c], Hongxu Liu[d,**], Yuan Yuan[a,b,c,*]

[a] Cancer Etiology and Screening Department of Cancer Institute and General Surgery, the First Hospital of China Medical University, Shenyang 110001, China
[b] Key Laboratory of Cancer Etiology and Prevention in Liaoning Education Department, the First Hospital of China Medical University, Shenyang 110001, China
[c] Key Laboratory of GI Cancer Etiology and Prevention in Liaoning Province, the First Hospital of China Medical University, Shenyang 110001, China
[d] Department of Thoracic Surgery, Cancer Hospital of China Medical University, Liaoning Cancer Hospital & Institute, No. 44 Xiaoheyan Road, Shenyang, Liaoning 110042, China
[e] Ultrasound Department, the First Hospital of China Medical University, Shenyang 110001, China

ARTICLE INFO

ABSTRACT

Methylation, as an epigenetic modification, can affect gene expression and play a role in the occurrence and development of cancer. This research is devoted to discover methylated-differentially expressed genes (MDEGs) in esophageal squamous cell carcinoma (ESCC) and explore special associated pathways. We downloaded GSE51287 methylation profiles and GSE26886 expression profiles from GEO DataSets, and performed a comprehensive bioinformatics analysis. Totally, 19 hypermethylated, lowly expressed genes (Hyper-LGs) were identified, and involved in regulation of cell proliferation, phosphorus metabolic process and protein kinase activity. Meanwhile, 17 hypomethylated, highly expressed genes (Hypo-HGs) were participated in collagen catabolic process, metallopeptidase and cytokine activity. Pathway analysis determined that Hyper-LGs were enriched in arachidonic acid metabolism pathway, while Hypo-HGs were primarily associated with the cytokine-cytokine receptor interaction pathway. IL 6, MMP3, MMP9, SPP1 were identified as hub genes based on the PPI network that combined 7 ranked methods included in cytoHubba, and verification was performed in human tissues. Our integrated analysis identified many novel genetic lesions in ESCC and provides a crucial molecular foundation to improve our understanding of ESCC. Hub genes, including IL 6, MMP3, MMP9 and SPP1, could be considered for use as aberrant methylation-based biomarkers to facilitate the accurate diagnosis and therapy of ESCC.

1. Introduction

Esophageal cancer, is one of the most invasive cancers and, broadly, the seventh principal cause of cancer-related deaths for males. In China, it is the fourth-most fatal cancer, and includes two main subtypes, namely, esophageal squamous cell carcinoma (ESCC), and esophageal adenocarcinoma (EAC) [1,2]. ESCC is the primary histological classification that is broadly observed [3], and both its occurrence and progression are related to genetic factors, such as genomic amplifications, insertions, deletions, and mutations [4], as well as tumor epigenetics, which includes DNA methylation, histone acetylation and noncoding RNA [5]. DNA methylation commonly affects independent loci,

and different tumor types exhibit unique signatures of DNA methylation deregulation [6]. Aberrant DNA methylation, whether the hypermethylation of tumor suppressor genes or the hypomethylation of oncogenes, is considered to be a significant factor for carcinogenesis [7]. As a result, a profound understanding of methylated-differentially expressed genes (MDEGs) and their genetic characteristics is essential in the elucidation of the physiological and pathological processes of ESCC.

Earlier studies often analyzed gene expression or DNA methylation using array-based profiling [8–10]. However, the primary focus of many of these studies was on the relationship between the expression and methylation of solitary genes [11,12]. We can concurrently detect methylated expressed genes through the joint analysis of methylation

and gene expression microarray files and further determine their interrelated functions and biological characteristics [13,14]. To our knowledge, there are no published studies that have jointly examined gene expression and methylation profiling microarrays to investigate the carcinogenic processes of ESCC.

We aimed to profile the associations of interactions between differentially methylated genes (DMGs) and differentially expressed genes (DEGs), as well as signaling pathways in ESCC, through the bioinformatics analysis of gene methylation microarray profiles (GSE51287) and gene expression microarray profiles (GSE26866). The purpose of this study was to gain a novel perspective into the genetic features and biological pathways of MDEGs in ESCC, as well as to offer insights into the pathogenesis of ESCC.

2. Material and methods

2.1. Microarray data

We performed our analysis using methylation and mRNA microarray datasets to identify MDEGs between normal and ESCC tissue samples. The gene methylation dataset (GSE51287) and gene expression dataset (GSE26886) were downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), which is a database maintained by the National Center for Biotechnology Information (NCBI). The GSE51287 gene methylation microarray data comprise 98 ESCC and 93 normal tissue specimens (platform: GPL9183 Illumina GoldenGate Methylation Cancer Panel I); 9 ESCC and 19 normal squamous epithelial specimens are registered for GSE26886 (platform: GPL[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array).

2.2. Data process

Differential expression between normal and ESCC tissue samples was identified using the online tool GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) across experimental conditions in the GEO series. The multiple t-test was used to detect statistically significant genes, and the Benjamin & Hochberg false discovery rate method is selected. Fold change was used to express a difference in expression. In this study, we defined the cut-off criteria based on $P < 0.05$ and $|fold\ change| > 2$ to identify DMGs and DEGs. Additionally, we classified overlapping MDEGs by performing a "MATCH function". Finally, Hyper-LGs were identified by overlapping down-regulated and hypermethylated genes, while Hypo-HGs were identified by overlapping up-regulated and hypomethylated genes. GSE51287 methylation probes performed an exclusion criteria about filtering that located in sex chromosome.

2.3. Functional and pathway enrichment analysis

We performed functional annotation and enrichment analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/), which is an online tool used to determine the biological characteristics of large lists of genes [15]. Gene Ontology (GO) analysis is a robust bioinformatics tool for gene and gene product annotation, including biological processes, cellular components, and molecular functions [16]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database can be used to visualize molecular datasets including genomics, transcriptomics, proteomics, and metabolomics via the KEGG pathway to interpret the biological functions of these molecules [17]. The GO function and KEGG pathway analyses of MDEGs were performed using DAVID, and a P-value < 0.05 was considered to be statistically significant.

2.4. Protein–protein interaction (PPI) network construction and hub gene analysis

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, http://string-db.org/) is an online database that predicts PPIs [18]. To determine the underlying molecular principles of cell activity within the context of cancer progression, we constructed a PPI network using STRING for the Hyper-LGs and Hypo-HGs. An interaction score of 0.4 was set as the cut-off standard. We then visualized the PPI network using Cytoscape (http://www.cytoscape.org/) and ranked hub genes using cytoHubba within Cytoscape.

2.5. MDEGs analysis between ESCC and normal control tissues in TCGA database

The Cancer Genome Atlas (TCGA) database, a collaboration between the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI), has generated comprehensive, multi-dimensional maps of the key genomic changes in various types of cancers. In order to confirm the results, Hyper-LGs and Hypo-HGs were then analyzed in TCGA database. In this study, the nominal P values were adjusted by Benjamin & Hochberg false discovery rate method, and we defined the cut-off criteria based on $P < 0.05$ and $|fold\ change| \geq 1$ to identify DMGs and DEGs.

2.6. Verification of hub genes in human tissues

We further verified the hub genes using 20 pairs of tumor and adjacent non-tumor tissues in esophageal cancer that were matched according to age and sex (approved by the Human Ethics Review Committee of Cancer Hospital of China Medical University). All patients provided written informed consent for the clinical trial and using their data for research, this study was conducted in accordance with the Declaration of Helsinki.

Bisulfite genomic sequencing (BGS) or methylation-specific PCR (MSP) was used for the methylation detection, with deionized water as a negative control. The sodium bisulfite modification of genomic DNA was performed using the Zymo DNA Methylation-Gold kit (Zymo Research, US), and all procedures followed the manufacturer's protocol. Polymerase chain reaction was performed by HotStart Taq 2.0 Version (Takara, Japan), with the primers for BGS and the list of conditions included in Table 1 and Table 2. For BGS, the PCR products were used for forward sequence analysis after product purification. For MSP, the PCR products were separated and confirmed, along with methylated or unmethylated statues were defined in accordance with our previous study [13]. In addition, quantitative real-time PCR was performed to profile mRNA expression. Differences between groups were compared by the Mann-Whitney U test, and a P-value < 0.05 was considered statistically significant. All the primers and conditions are shown in Table 3 and Table 4.

Table 1
BGS or MSP primer sequence.

Name		Sequence	method
MMP9	F	GAATTTTGGGTTTTGGTTTTAG	MSP
	R	AAACTCTATCCTCTTTTTCCCTC	
MMP7	F	ATGGAGTTATTGGGTTAGAATT	BGS
	R	CCCTTCTTTAAATCTCCTTTTA	
MMP3	F	AATGGTTTTGTTGTTATTTGGAT	BGS
	R	TTACTCTAAAAACTCCCCACCT	
SPP1	F	TGGGGAAGTTTAAATTTTAAGG	BGS
	R	CCTACACAATCACCCACTAAA	
IL-6	F	TATTGAGATGTAAAGGAAGGAAGT	BGS
	R	TAAAAACATTAAAAATCCCCAAA	

Table 2
PCR reaction conditions.

	94 °C	5min
40cyc	94 °C	20 s
	55 °C	20 s
	72 °C	30 s
	72 °C	6min

Table 3
Realtime-PCR primer sequence.

Name		Sequence
MMP9	F	GGGACGCAGACATCGTCATC
	R	TCGTCATCGTCGAAATGGGC
MMP7	F	GAGTGAGCTACAGTGGGAACA
	R	CTATGACGCGGGAGTTTAACAT
MMP3	F	CGGTTCCGCCTGTCTCAAG
	R	CGCCAAAAGTGCCTGTCTT
SPP1	F	CTCCATTGACTCGAACGACTC
	R	CAGGTCTGCGAAACTTCTTAGAT
IL-6	F	ACTCACCTCTTCAGAACGAATTG
	R	CCATCTTTGGAAGGTTCAGGTTG

Table 4
Realtime-PCR reaction conditions.

	95 °C	30 s
40cyc	95 °C	10 s
	56 °C	20 s
	72 °C	30 s
	melting curve	
	4 °C	Hold

3. Results

3.1. Screening for MDEGs in ESCC

A total of 1468 and 54,675 gene records were obtained from the GSE51287 and GSE26886 profiling datasets, respectively. We performed an online analysis using GEO2R software to identify DMGs or DEGs. By comparing 256 DMGs (165 hypermethylated genes and 91 hypomethylated genes) with 8512 DEGs (4020 upregulated genes and 4492 downregulated genes), we categorized 19 Hyper-LGs and 17 Hypo-HGs that were subsequently examined in GO, KEGG, and PPI analyses. The flowchart for this process is shown in Fig. 1.

3.2. GO functional enrichment analysis

GO enrichment analysis is shown in Table 5. For Hyper-LGs, biological processes (BP) enriched in protein amino acid phosphorylation, regulation of cell proliferation and phosphorus metabolic process. Cell components (CC) mainly indicated endomembrane system and plasma membrane part. For molecular function (MF), there was an enrichment in protein kinase activity, transmembrane receptor protein tyrosine kinase activity, vascular endothelial growth factor receptor activity and ATP binding. For Hypo-HGs, BPs were enriched in the collagen catabolic process, multicellular organismal catabolic and metabolic process. CC were enriched in the extracellular space, the matrix region part, and in the proteinaceous extracellular matrix. Finally, MF enrichment was found in metalloendopeptidase activity, metallopeptidase activity, endopeptidase activity, calcium ion binding, and cytokine activity.

3.3. KEGG pathway analysis

For Hyper-LGs, KEGG pathway enrichment analysis demonstrated

enrichment in the arachidonic acid metabolism pathway. Hypo-HGs were significantly involved in the toll-like receptor signalling pathway and the cytokine-cytokine receptor interaction pathway (Table 6).

3.4. PPI network construction and cytoHubba analysis

We analyzed the MDEGs using the online PPI network tool STRING. Nineteen nodes and 3 edges in the Hyper-LGs networks, and 17 nodes and 54 edges in Hypo-HGs networks were found, respectively. The PPI network revealed significant interactions for the Hypo-HGs, with an enrichment P-value of 4.19e-09, whereas the PPI network did not detect any significant interactions for the Hyper-LGs ($P = 0.225$). The Hypo-HGs PPI network is shown in Fig. 2. We then visualized the Hypo-HGs network using Cytoscape, and the hub genes were identified by cyto-Hubba within Cytoscape. Finally, we identified 5 hub genes by over-lapping 7 ranked methods in cytoHubba (Table 7). These genes are annotated as Interleukin 6 (IL6), Matrix Metallopeptidase 9 (MMP9), MMP3, MMP7, and Secreted Phosphoprotein 1 (SPP1).

3.5. MDEGs analysis between ESCC and normal control tissues in TCGA database

There are 95 ESCC tissues, but only 3 normal control tissues in TCGA database including both DNA methylation and mRNA expression. We downloaded the data for MDEGs analysis, and found some MDEGs (Fig. 3, supplement Table 1), such as top 5 genes, CLDN18, CLIC6, KCNJ13, ME3, CKMT2, their expression changes caused by methylation may affect the occurrence and development of ESCC. Unfortunately, there is no common result with GEO data analysis. Due to the COVID-19 pandemic impact, we can not verify the analysis results of TCGA database by histology at present. We hope to have larger sample data in the future to make up for the current analysis.

3.6. Verification in human tissues

We next sought to verify the five identified hub genes in human tissues and found that gene expression levels of IL6, MMP9, MMP3, and SPP1 were higher in tumor tissues than in non-tumor tissues, though only SPP1 with a statistically significant. Furthermore, we identified gene hypomethylation in tumor samples that included all five hub genes. Correlation analysis showed a negative correlation between the expression of the IL6, MMP9, MMP3, and SPP1 genes and their methylation while excluding MMP7 (Table 8).

4. Discussion

ESCC goes through a multistage and complex process that involves multiple molecular changes comprised of increasing genetic, epigenetic, and endocrine aberrations [19]. We identified 19 Hyper-LGs and 17 Hypo-HGs through the analysis of gene methylation microarray data (GSE51287) and gene expression profiling data (GSE26886) for ESCC by utilizing public datasets and online bioinformatics tools. We found that linked genes could possibly be associated with the molecular guidance of vital pathways that are related to the pathogenesis of ESCC. Enrichment and functional analysis of the genes identified major pathways, and hub genes that are related to methylation offer a unique perspective into the pathogenesis of ESCC.

Based on DAVID analysis, GO enrichment of Hyper-LGs in ESCC revealed that BP included the protein amino acid phosphorylation, regulation of cell proliferation and phosphorus metabolic process. MF was enriched in protein kinase activity, transmembrane receptor protein tyrosine kinase activity, vascular endothelial growth factor receptor activity and ATP binding. KEGG enrichment analysis also indicated that the progression of ESCC may be affected by methylation via the arachidonic acid metabolism pathway.

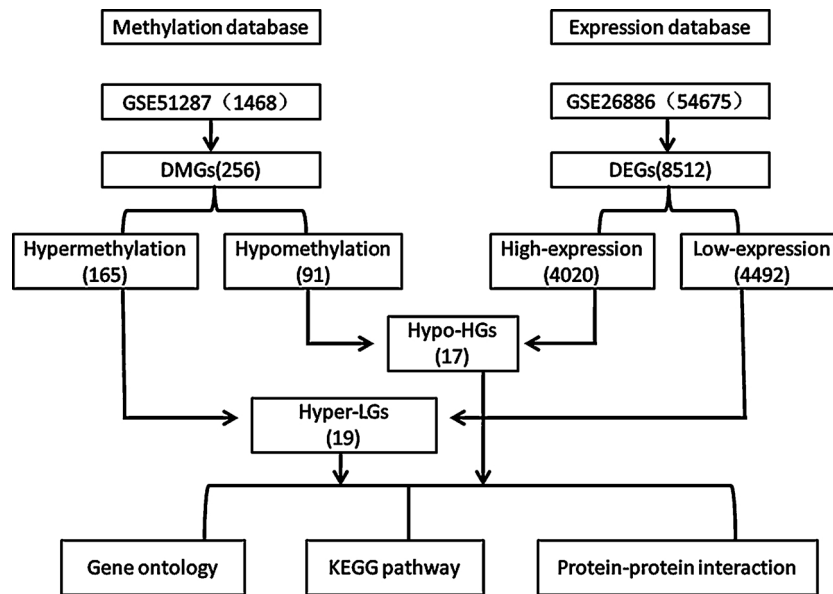Hypo-HGs in ESCC were enriched in BPs, including the collagen

Fig. 1. Flowchart of bioinformatics analyses. DMGs: differentially methylated genes; DEGs: differentially expressed genes; Hyper-LGs: hypermethylated, lowly expressed genes; Hypo-HGs: hypomethylated, highly expressed genes.

catabolic process, multicellular organismal catabolic and metabolic processes. GO analysis found that MF was enriched in metalloendopeptidase activity, metallopeptidase activity, endopeptidase activity, calcium ion binding, and cytokine activity. A previous study found that the cell cycle regulation pathway is ubiquitous in ESCC [20], and there are notable associations between cell proliferation, gene expression and signal transduction in ESCC incursion and metastasis [21]. A transcription factor that induces the cellular response to oxidative stress plays crucial role in the development of ESCC [22]. According to our GO analysis, MFs enrichment included metalloendopeptidase activity,

metallopeptidase activity, endopeptidase activity, calcium ion binding, and cytokine activity.

After constructing PPI networks for MDEGs, we observed significant interactions only for the Hypo-HGs network, with some of the MDEGs being implicated in the pathogenesis of ESCC. We visualized the Hypo-HGs network using Cytoscape and utilized cytoHubba within Cytoscape to identify the following 5 hub genes: IL6, MMP9, MMP3, MMP7, and SPP1. We then verified the observed interactions of these hub genes in human tissues and found that IL6, MMP9, MMP3, and SPP1 exhibited a negative correlation between high gene expression and DNA
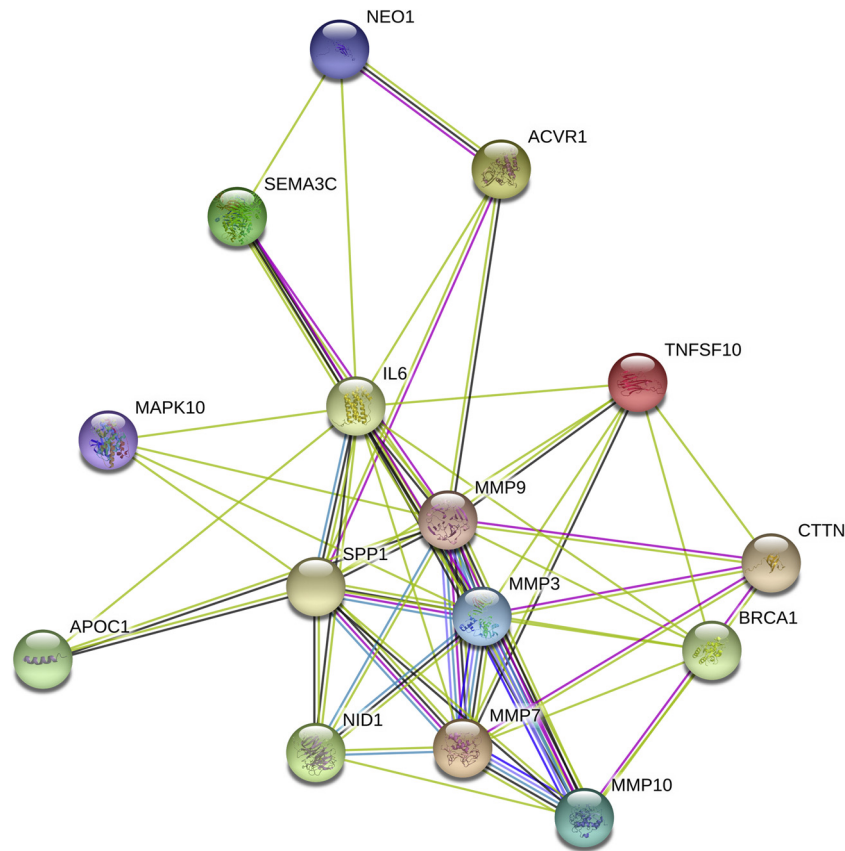
Table 5
GO enrichment analysis of MDEGs related to ESCC.

Category		Term	Count	%	PValue
Hyper-LGs	GOTERM_BP_FAT	GO:0006468 ~ protein amino acid phosphorylation	6	31.58	1.44E-03
	GOTERM_BP_FAT	GO:0042127 ~ regulation of cell proliferation	6	31.58	2.99E-03
	GOTERM_BP_FAT	GO:0016310 ~ phosphorylation	6	31.58	3.21E-03
	GOTERM_BP_FAT	GO:0046777 ~ protein amino acid autophosphorylation	3	15.79	5.59E-03
	GOTERM_BP_FAT	GO:0006793 ~ phosphorus metabolic process	6	31.58	7.41E-03
	GOTERM_CC_FAT	GO:0012505 ~ endomembrane system	4	21.05	4.12E-02
	GOTERM_CC_FAT	GO:0044459 ~ plasma membrane part	6	31.58	5.79E-02
	GOTERM_MF_FAT	GO:0004672 ~ protein kinase activity	6	31.58	8.44E-04
	GOTERM_MF_FAT	GO:0004713 ~ protein tyrosine kinase activity	4	21.05	1.22E-03
	GOTERM_MF_FAT	GO:0004714 ~ transmembrane receptor protein tyrosine kinase activity	3	15.79	3.39E-03
	GOTERM_MF_FAT	GO:0005021 ~ vascular endothelial growth factor receptor activity	2	10.53	1.04E-02
	GOTERM_MF_FAT	GO:0005524 ~ ATP binding	6	31.58	3.64E-02
Hypo-HGs	GOTERM_BP_FAT	GO:0030574 ~ collagen catabolic process	4	22.22	1.53E-06
	GOTERM_BP_FAT	GO:0044243 ~ multicellular organismal catabolic process	4	22.22	3.47E-06
	GOTERM_BP_FAT	GO:0032963 ~ collagen metabolic process	4	22.22	4.37E-06
	GOTERM_BP_FAT	GO:0044259 ~ multicellular organismal macromolecule metabolic process	4	22.22	5.98E-06
	GOTERM_BP_FAT	GO:0044236 ~ multicellular organismal metabolic process	4	22.22	1.03E-05
	GOTERM_CC_FAT	GO:0044421 ~ extracellular region part	10	55.56	1.03E-06
	GOTERM_CC_FAT	GO:0005615 ~ extracellular space	9	50.00	1.03E-06
	GOTERM_CC_FAT	GO:0005576 ~ extracellular region	10	55.56	4.21E-04
	GOTERM_CC_FAT	GO:0005578 ~ proteinaceous extracellular matrix	5	27.78	7.09E-04
	GOTERM_CC_FAT	GO:0031012 ~ extracellular matrix	5	27.78	9.39E-04
	GOTERM_MF_FAT	GO:0004222 ~ metalloendopeptidase activity	4	22.22	2.59E-04
	GOTERM_MF_FAT	GO:0008237 ~ metallopeptidase activity	4	22.22	1.35E-03
	GOTERM_MF_FAT	GO:0004175 ~ endopeptidase activity	4	22.22	1.01E-02
	GOTERM_MF_FAT	GO:0005509 ~ calcium ion binding	5	27.78	2.29E-02
	GOTERM_MF_FAT	GO:0005125 ~ cytokine activity	3	16.67	2.35E-02

The top five terms are listed based on *P* values over five terms in per category. Hyper-LGs: hypermethylated, lowly expressed genes; Hypo-HGs: hypomethylated, highly expressed genes.

Table 6
KEGG pathway analysis of MDEGs related to ESCC.

Category		Term	Count	%	PValue
Hyper-LGs	KEGG_PATHWAY	hsa00590:Arachidonic acid metabolism	2	10.53	9.49E-02
Hypo-HGs	KEGG_PATHWAY	hsa04620:Toll-like receptor signaling pathway	3	16.67	1.58E-02
	KEGG_PATHWAY	hsa04060:Cytokine-cytokine receptor interaction	3	16.67	9.06E-02

Hyper-LGs: hypermethylated, lowly expressed genes; Hypo-HGs: hypomethylated, highly expressed genes.

Fig. 2. Hypo-HGs PPI network. Disconnected nodes are hidden in the network. A total of 17 nodes and 54 edges were found in the Hypo-HGs networks.

Table 7
Hub genes for Hypo-HGs ranked in cytoHubba.

gene symbols	Ranking methods in cytoHubba						
	MCC	DMNC	MNC	Degree	EPC	Closeness	Radiality
IL6	1572	0.42	13	13	4.49	13.50	3.21
MMP9	1568	0.53	11	11	4.26	12.33	3.00
MMP3	1560	0.67	9	9	4.05	11.33	2.86
MMP7	1560	0.67	9	9	3.99	11.33	2.86
SPP1	1448	0.54	10	10	4.15	11.83	2.93

MCC = maximal cilque centrality, DMNC = density of maximum neighbourhood component, MNC = maximum neighbourhood component, Degree = node connect degree, EPC = edge percolated component.

hypomethylation, but the same was not observed for MMP7. IL6 is a proinflammatory cytokine associated with cancer development [23,24], including ESCC [25], and ESCC patients have been reported to have significantly higher serum expression levels of IL6 [26]. Furthermore, the methylation of IL6 has been associated with a range of cancers [27]. The autocrine loops of IL6 and IL6R are engaged in the development and progression of ESCC [28], which is inconsistent with our finding regarding the cytokine-cytokine receptor interaction pathway based on

KEGG enrichment analysis. Matrix metalloproteinases (MMPs), a zinc-dependent endopeptidases enzyme family with the ability to degrade extracellular matrix components, are considered to be involved in various stages of cancer progression [29–31]. Changes in MMPs expression may be due to DNA methylation since hypomethylation of a gene can result in transcriptional up-regulation [11,32]. MMP3 is one of the MMP genes involved in tumor initiation, has calcium ion binding and metallopeptidase activity, and is known to degrade basal membrane collagen and stimulate the production of other MMPs [33]. MMP9, categorized in the subgroup of gelatinases, was shown to participate in the growth and progression of many cancers [34]. MMP9 activity primarily degrades type IV collagen, which is regarded as the main basement membrane component [35], indicating an association with the progression of ESCC through the collagen catabolic process, which was found to be enriched in this study. Over-expression of MMP3 or MMP9 is closely related to the pathogenesis of many cancers, including ESCC [36–38], consistent with our findings. SPP1 is a protein coding gene with cytokine and extracellular matrix binding activity, that is involved in many biological processes, including cell proliferation, migration and invasion, and its over-expression is linked to tumor initiation and prognosis, including in ESCC [39–42]. A previous study showed that hypomethylation of SPP1 was strongly associated with
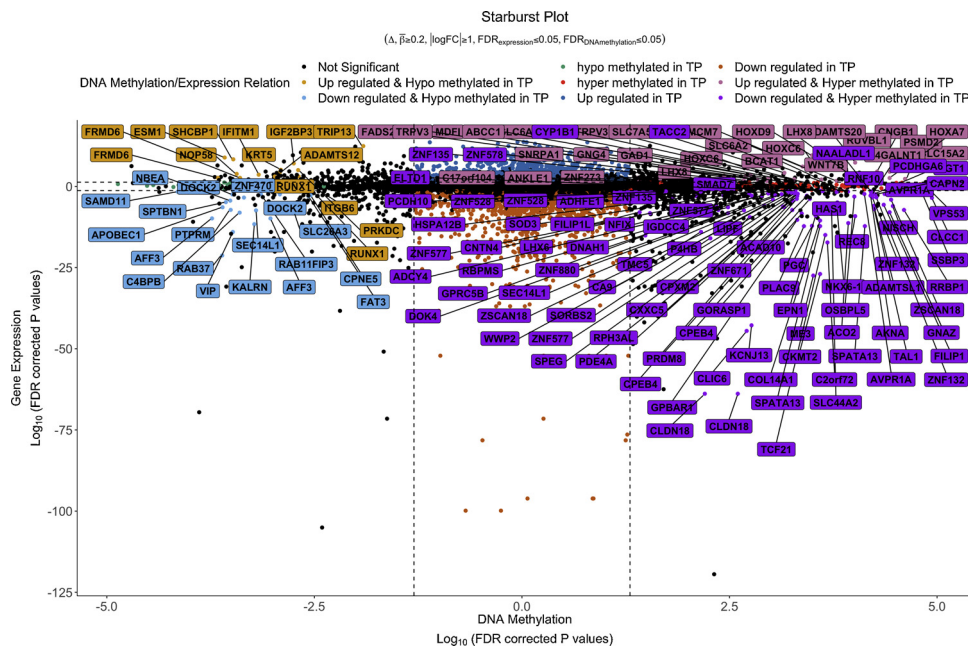
Fig. 3. MDEGs analysis between ESCC and normal control tissues in TCGA database.

Table 8
Human tissue verification for hub genes.

gene	mRNA expression		Methylation	
	Median(25 %,75 %)[a]	P value	Median(25 %,75 %)	P value
MMP3		0.290		0.885
Tumor	3.700(4.020, 2.809)		0.784(0.695, 0.825)	
Non-tumor	4.326(5.601, 2.996)		0.786(0.753, 0.870)	
MMP9		0.096		0.131
Tumor	2.757(3.192, 2.659)		0.191(0.120, 0.539)	
Non-tumor	3.265(3.574, 2.851)		0.455(0.407, 0.546)	
MMP7		0.510		0.261
Tumor	3.796(4.243, 3.219)		0.918(0.894, 0.959)	
Non-tumor	3.016(4.999, 2.363)		0.934(0.918, 0.951)	
SPP1		**0.033**		0.108
Tumor	2.411(3.244, 2.297)		0.786(0.722, 0.866)	
Non-tumor	3.552(4.330, 3.315)		0.849(0.824, 0.882)	
IL-6		0.345		0.520
Tumor	4.083(4.533, 3.779)		0.847(0.834, 0.910)	
Non-tumor	4.608(5.253, 3.883)		0.892(0.834, 0.932)	

[a] Take -log10 as standardization.

cancer progression, though an inverse relationship with its mRNA expression was found for gastrointestinal stromal tumors [43]. This finding is also consistent with the results of our ESCC analysis. Taken together, the identified hubs are closely related to metallopeptidase activity, cytokine activity, and the degradation of extracellular matrix components during physiological and pathological processes.

In our previous studies, we found that cancer-related genes were hypermethylated - lowly expressed or hypomethylated - highly expressed in the tumor tissues of cancer patients [13,44,45]. In this study, our findings indicate that the MDEGs in ESCC may exert regulatory effects composed of molecular functions and biological processes, as reliably characterized by enrichment analysis. The Hypo-HGs PPI network had significantly more interactions than expected, though the same was not seen in the Hyper-LGs network. DNA hypomethylation of the identified hub genes generally leads to increased gene expression and may promote the progression of ESCC. However, further investigation of the novel genes and pathways identified in this study that have not been previously considered as targets during ESCC pathogenesis are required.

Several limitations of our study should be mentioned. First, our study did not investigate the clinical parameters and prognosis due to a lack of data availability in the bioinformatics databases and tools. Second, the sample size was small since only two microarray profiles were analyzed, and only 20 pairs of human tumor and adjacent non-tumor tissue were verified, this may have led to no significant difference in the methylation status of hub genes in tumor tissue compared with non-tumor tissue. Third, although we analyzed the GEO database and TCGA database at the same time, due to the bias of the samples, especially the small number of normal tissue samples in TCGA database, we did not obtain the key genes of ESCC common MDEGs in the two databases. Thus, larger sample sizes are necessary to validate our findings. In addition, future molecular experiments on the identified target genes and pathways in ESCC should be performed.

5. Conclusion

In conclusion, we elucidated the biological characteristics of ESCC by constructing interaction networks for DMGs and DEGs using several bioinformatics databases and tools. Our results suggest that the interactions of MDEGs are related to the development and progression of ESCC through specialized functions and signaling pathways. Hub genes for Hypo-HGs in ESCC, including IL6, MMP9, MMP3, and SPP1, may serve as specific biomarkers for the accurate diagnosis and treatment of ESCC based on aberrant methylation. This study provides hypothetical insight into the pathogenesis of ESCC. Further molecular studies are required to validate the identified genes in ESCC and to investigate their underlying mechanisms.

CRediT authorship contribution statement

Liang Sang: Methodology, Writing - original draft. **Zhanwu Yu:** Methodology, Writing - review & editing. **Ang Wang:** Resources, Formal analysis. **Hao Li:** Resources, Formal analysis. **Xiantong Dai:** Resources, Formal analysis. **Liping Sun:** Investigation, Resources. **Hongxu Liu:** Conceptualization, Writing - review & editing. **Yuan Yuan:** Supervision, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.prp.2020.153050.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer Statistics, 2017, CA Cancer J. Clin. 67 (2017) 7–30.

[2] W. Chen, R. Zheng, P.D. Baade, S. Zhang, H. Zeng, F. Bray, A. Jemal, X.Q. Yu, J. He, Cancer statistics in China, 2015, CA Cancer J. Clin. 66 (2016) 115–132.

[3] G.C. Ligthart-Melis, P.J. Weijs, N.D. te Boveldt, S. Buskermolen, C.P. Earthman, H.M. Verheul, E.S. de Lange-de Klerk, S.J. van Weyenberg, D.L. van der Peet, Dietician-delivered intensive nutritional support is associated with a decrease in severe postoperative complications after surgery in patients with esophageal cancer, Dis. Esophagus 26 (2013) 587–593.

[4] Y. Sasaki, M. Tamura, R. Koyama, T. Nakagaki, Y. Adachi, T. Tokino, Genomic characterization of esophageal squamous cell carcinoma: insights from next-generation sequencing, World J. Gastroenterol. 22 (2016) 2284–2293.

[5] C. Kandoth, M.D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J.F. McMichael, M.A. Wyczalkowski, M.D.M. Leiserson, C.A. Miller, J.S. Welch, M.J. Walter, M.C. Wendl, T.J. Ley, R.K. Wilson, B.J. Raphael, L. Ding, Mutational landscape and significance across 12 major cancer types, Nature 502 (2013) 333–339.

[6] R.H. Wilting, J.H. Dannenberg, Epigenetic mechanisms in tumorigenesis, tumor cell heterogeneity and drug resistance, Drug Resist. Updat. 15 (2012) 21–38.

[7] X. Yang, L. Gao, S. Zhang, Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns, Brief. Bioinformatics 18 (2017) 761–773.

[8] J.W. Yang, Y.L. Choi, Genomic profiling of esophageal squamous cell carcinoma (ESCC)-Basis for precision medicine, Pathol. Res. Pract. 213 (2017) 836–841.

[9] W. Pu, C. Wang, S. Chen, D. Zhao, Y. Zhou, Y. Ma, Y. Wang, C. Li, Z. Huang, L. Jin, S. Guo, J. Wang, M. Wang, Targeted bisulfite sequencing identified a panel of DNA methylation-based biomarkers for esophageal squamous cell carcinoma (ESCC), Clin. Epigenetics 9 (2017) 129.

[10] H. Peng, S. Wang, L. Pang, L. Yang, Y. Chen, X.B. Cui, Comprehensive bioinformation analysis of methylated and differentially expressed genes in esophageal squamous cell carcinoma, Mol. Omics 15 (2019) 88–100.

[11] M. Esteller, CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future, Oncogene 21 (2002) 5427–5440.

[12] H. Makino, T. Tajiri, M. Miyashita, K. Sasajima, R. Anbazhagan, J. Johnston, E. Gabrielson, Differential expression of TCEAL1 in esophageal cancers by custom cDNA microarray analysis, Dis. Esophagus 18 (2005) 37–40.

[13] H. Li, J.W. Liu, S. Liu, Y. Yuan, L.P. Sun, Bioinformatics-based identification of methylated-differentially expressed genes and related pathways in gastric Cancer, Dig. Dis. Sci. 62 (2017) 3029–3039.

[14] S.R. Chen, K. Zhou, L.G. Yang, G.H. Ding, H. Li, Racial differences in esophageal squamous cell carcinoma: incidence and molecular features, Biomed Res. Int. (2017).

[15] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.

[16] M.A. Harris, J.I. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Collins, K. Eilbeck, S. Lewis, C. Mungall, J. Richter, G.M. Rubin, S.Q. Shu, J.A. Blake, C.J. Bult, A.D. Diehl, M.E. Dolan, H.J. Drabkin, J.T. Eppig, D.P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, G. Binkley, J.M. Cherry, K.R. Christie, M.C. Costanzo, Q. Dong, S.R. Engel, D.G. Fisk, J.E. Hirschman, B.C. Hitz, E.L. Hong, C. Lane, S. Miyasato, R. Nash, A. Sethuraman, M. Skrzypek, C.L. Theesfeld, S.A. Weng, D. Botstein, K. Dolinski, R. Oughtred, T. Berardini, S. Mundodi, S.Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, N. Mulder, R. Chisholm, P. Fey, P. Gaudet, W. Kibbe, K. Pilcher, C.A. Bastiani, R. Kishore, E.M. Schwarz, P. Sternberg, K. Van Auken, M. Gwinn, L. Hannick, J. Wortman, M. Aslett, M. Berriman, V. Wood, S. Bromberg, C. Foote, H. Jacob, D. Pasko, V. Petri, D. Reilly, K. Seiler, M. Shimoyama, J. Smith, S. Twigger, P. Jaiswal, T. Seigfried, C. Collmer, D. Howe, M. Westerfield, The Gene Ontology (GO) project in 2006, Nucleic Acids Res. 34 (2006) D322–D326.

[17] M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes, Nucleic Acids Res. 28 (2000) 27–30.

[18] C. Di Mauro, R. Rosa, V. D'Amato, P. Ciciola, A. Servetto, R. Marciano, R.C. Orsini,

[19] L. Formisano, S. De Falco, V. Cicatiello, M. Di Bonito, M. Cantile, F. Collina, A. Chambery, B.M. Veneziani, S. De Placido, R. Bianco, Hedgehog signalling pathway orchestrates angiogenesis in triple-negative breast cancers, Br. J. Cancer 116 (2017) 1425–1435.

[19] Y. Zhang, T.C. Pan, X.X. Zhong, C. Cheng, Androgen receptor promotes esophageal cancer cell migration and proliferation via matrix metalloproteinase 2, Tumor Biol. 36 (2015) 5859–5864.

[20] L. Zhang, Y. Zhou, C. Cheng, H. Cui, L. Cheng, P. Kong, J. Wang, Y. Li, W. Chen, B. Song, F. Wang, Z. Jia, L. Li, Y. Li, B. Yang, J. Liu, R. Shi, Y. Bi, Y. Zhang, J. Wang, Z. Zhao, X. Hu, J. Yang, H. Li, Z. Gao, G. Chen, X. Huang, X. Yang, S. Wan, C. Chen, B. Li, Y. Tan, L. Chen, M. He, S. Xie, X. Li, X. Zhuang, M. Wang, Z. Xia, L. Luo, J. Ma, B. Dong, J. Zhao, Y. Song, Y. Ou, E. Li, L. Xu, J. Wang, Y. Xi, G. Li, E. Xu, J. Liang, X. Yang, J. Guo, X. Chen, Y. Zhang, Q. Li, L. Liu, Y. Li, X. Zhang, Y. Yang, D. Lin, X. Cheng, Y. Guo, J. Wang, Q. Zhan, Y. Cui, Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma, Am. J. Hum. Genet. 96 (2015) 597–611.

[21] M. Qixing, D. Gaochao, X. Wenjie, W. Anpeng, C. Bing, M. Weidong, X. Lin, J. Feng, Microarray analyses reveal genes related to progression and prognosis of esophageal squamous cell carcinoma, Oncotarget 8 (2017) 78838–78850.

[22] G. Sawada, A. Niida, R. Uchi, H. Hirata, T. Shimamura, Y. Suzuki, Y. Shiraishi, K. Chiba, S. Imoto, Y. Takahashi, T. Iwaya, T. Sudo, T. Hayashi, H. Takai, Y. Kawasaki, T. Matsukawa, H. Eguchi, K. Sugimachi, F. Tanaka, H. Suzuki, K. Yamamoto, H. Ishii, M. Shimizu, H. Yamazaki, M. Yamazaki, Y. Tachimori, Y. Kajiyama, S. Natsugoe, H. Fujita, K. Mafune, Y. Tanaka, D.P. Kelsell, C.A. Scott, S. Tsuji, S. Yachida, T. Shibata, S. Sugano, Y. Doki, T. Akiyama, H. Aburatani, S. Ogawa, S. Miyano, M. Mori, K. Mimori, Genomic landscape of esophageal squamous cell carcinoma in a japanese population, Gastroenterology 150 (2016) 1171–1182.

[23] B.E. Lippitz, R.A. Harris, Cytokine patterns in cancer patients: a review of the correlation between interleukin 6 and prognosis, Oncoimmunology 5 (2016) e1093722.

[24] M. Lukaszewicz-Zajac, B. Mroczko, M. Kozlowski, J. Niklinski, J. Laudanski, M. Szmitkowski, Higher importance of interleukin 6 than classic tumor markers (carcinoembryonic antigen and squamous cell cancer antigen) in the diagnosis of esophageal cancer patients, Dis. Esophagus 25 (2012) 242–249.

[25] M.F. Chen, P.T. Chen, M.S. Lu, P.Y. Lin, W.C. Chen, K.D. Lee, IL-6 expression predicts treatment response and outcome in squamous cell carcinoma of the esophagus, Mol. Cancer 12 (2013) 26.

[26] Y. Qiao, C. Zhang, A. Li, D. Wang, Z. Luo, Y. Ping, B. Zhou, S. Liu, H. Li, D. Yue, Z. Zhang, X. Chen, Z. Shen, J. Lian, Y. Li, S. Wang, F. Li, L. Huang, L. Wang, B. Zhang, J. Yu, Z. Qin, Y. Zhang, IL6 derived from cancer-associated fibroblasts promotes chemoresistance via CXCR7 in esophageal squamous cell carcinoma, Oncogene 37 (2018) 873–883.

[27] B.T. Joyce, T. Gao, L. Liu, Y. Zheng, S. Liu, W. Zhang, F. Penedo, Q. Dai, J. Schwartz, A.A. Baccarelli, L. Hou, Longitudinal study of DNA methylation of inflammatory genes and Cancer risk, Cancer Epidemiol. Biomarkers Prev. 24 (2015) 1531–1538.

[28] M. Oka, K. Yamamoto, M. Takahashi, M. Hakozaki, T. Abe, N. Iizuka, S. Hazama, K. Hirazawa, H. Hayashi, A. Tangoku, K. Hirose, T. Ishihara, T. Suzuki, Relationship between serum levels of interleukin 6, various disease parameters and malnutrition in patients with esophageal squamous cell carcinoma, Cancer Res. 56 (1996) 2776–2780.

[29] X. Guan, X. Wang, H. Luo, J. Wu, X. Zhang, J. Wu, Matrix metalloproteinase 1, 3, and 9 polymorphisms and esophageal squamous cell carcinoma risk, Med. Sci. Monit. 20 (2014) 2269–2274.

[30] Y. Chen, T. Jiang, A. Mao, J. Xu, Esophageal cancer stem cells express PLGF to increase cancer invasion through MMP9 activation, Tumour Biol. 35 (2014) 12749–12755.

[31] K. Kessenbrock, V. Plaks, Z. Werb, Matrix metalloproteinases: regulators of the tumor microenvironment, Cell 141 (2010) 52–67.

[32] K. Campos, C.C. Gomes, L.C. Farias, R.M. Silva, A. Letra, R.S. Gomez, DNA methylation of MMP9 is associated with high levels of MMP-9 messenger RNA in periapical inflammatory lesions, J. Endod. 42 (2016) 127–130.

[33] J. Zhang, X. Jin, S. Fang, Y. Li, R. Wang, W. Guo, N. Wang, Y. Wang, D. Wen, L. Wei, G. Kuang, Z. Dong, The functional SNP in the matrix metalloproteinase-3 promoter modifies susceptibility and lymphatic metastasis in esophageal squamous cell carcinoma but not in gastric cardia adenocarcinoma, Carcinogenesis 25 (2004) 2519–2524.

[34] M. Groblewska, M. Siewko, B. Mroczko, M. Szmitkowski, The role of matrix metalloproteinases (MMPs) and their inhibitors (TIMPs) in the development of esophageal cancer, Folia Histochem. Cytobiol. 50 (2012) 12–19.

[35] Z.D. Gu, J.Y. Li, M. Li, J. Gu, X.T. Shi, Y. Ke, K.N. Chen, Matrix metalloproteinases expression correlates with survival in patients with esophageal squamous cell carcinoma, Am. J. Gastroenterol. 100 (2005) 1835–1843.

[36] R. Zeng, L. Duan, Y. Kong, Y. Liang, X. Wu, X. Wei, K. Yang, Clinicopathological and prognostic role of MMP-9 in esophageal squamous cell carcinoma: a meta-analysis, Chin. J. Cancer Res. 25 (2013) 637–645.

[37] Y. Li, H. Guo, D. Dong, H. Wu, E. Li, Expression and prognostic relevance of cyclophilin A and matrix metalloproteinase 9 in esophageal squamous cell carcinoma, Diagn. Pathol. 8 (2013) 207.

[38] L. Zhang, R.X. Xi, X.Z. Zhang, Matrix metalloproteinase variants associated with risk and clinical outcome of esophageal cancer, Genet. Mol. Res. 14 (2015) 4616–4624.

[39] S. Viana Lde, R.J. Affonso Jr, S.R. Silva, M.V. Denadai, D. Matos, C. Salinas de Souza, J. Waisberg, Relationship between the expression of the extracellular matrix genes SPARC, SPP1, FN1, ITGA5 and ITGAV and clinicopathological parameters of

tumor progression and colorectal cancer dissemination, Oncology 84 (2013) 81–91.

[40] A.B. Tuck, A.F. Chambers, A.L. Allan, Osteopontin overexpression in breast cancer: knowledge gained and possible implications for clinical management, J. Cell. Biochem. 102 (2007) 859–868.

[41] C.F. Huang, G.T. Yu, W.M. Wang, B. Liu, Z.J. Sun, Prognostic and predictive values of SPP1, PAI and caveolin-1 in patients with oral squamous cell carcinoma, Int. J. Clin. Exp. Pathol. 7 (2014) 6032–6039.

[42] Y. Li, Z. Lu, Y. Che, J. Wang, S. Sun, J. Huang, S. Mao, Y. Lei, Z. Chen, J. He, Immune signature profiling identified predictive and prognostic factors for esophageal squamous cell carcinoma, Oncoimmunology 6 (2017) e1356147.

[43] F. Haller, J.D. Zhang, E.A. Moskalev, A. Braun, C. Otto, H. Geddert, Y. Riazalhosseini, A. Ward, A. Balwierz, I.M. Schaefer, S. Cameron, B.M. Ghadimi, A. Agaimy, J.A. Fletcher, J. Hoheisel, A. Hartmann, M. Werner, S. Wiemann, O. Sahin, Combined DNA methylation and gene expression profiling in gastro-intestinal stromal tumors reveals hypomethylation of SPP1 as an independent prognostic factor, Int. J. Cancer 136 (2015) 1013–1023.

[44] J. Liu, H. Li, L. Sun, Z. Wang, C. Xing, Y. Yuan, Aberrantly methylated-differentially expressed genes and pathways in colorectal cancer, Cancer Cell Int. 17 (2017) 75.

[45] L. Sang, X.M. Wang, D.Y. Xu, W.J. Zhao, Bioinformatics analysis of aberrantly methylated-differentially expressed genes and pathways in hepatocellular carcinoma, World J. Gastroenterol. 24 (2018) 2605–2616.