An integrated methylation and gene expression microarray analysis reveals significant prognostic biomarkers in oral squamous cell carcinoma

CHENGUANG ZHAO[1*], HUIRU ZOU[2*], JUN ZHANG[3], JINHUI WANG[1] and HAO LIU[3]

[1]Department of Emergency; [2]Central Laboratory; [3]Department of Oral and Maxillofacial Surgery,
Tianjin Stomatological Hospital, Tianjin 300041, P.R. China

Abstract. Oral squamous cell carcinoma (OSCC) is a life-threatening disease with a poor prognosis. Although previous studies have reported that the methylation of certain genes is associated with the pathogenesis of OSCC, the methylation of genes that have relevance to OSCC progression is not clearly documented. The present study aimed to gain insights into the mechanisms underlying DNA methylation regulation associated with OSCC progression and to identify potential prognostic markers for OSCC treatment. DNA methylation dataset GSE41114 and gene expression dataset GSE74530 were downloaded from the Gene Expression Omnibus database. The global methylation status of OSCC tumor samples and normal control samples was determined, and differentially methylated genes (DMGs) in OSCC samples compared with control samples were identified. The mRNA expression data were then integrated to identify differentially expressed genes (DEGs) in OSCC samples compared with control samples. Overlapping genes between DEGs and DMGs were identified, and functional enrichment analysis was performed. In addition, survival analysis of the overlapping genes was performed to screen genes with prognostic significance in OSCC. A total of 40,115 differential methylation CpG sites spanning 3,360 DMGs were identified; CpG sites in the promoter, gene body and intergenic regions were generally highly hypermethylated or hypomethylated. Additionally, 508 DEGs in OSCC samples were identified, including 332 upregulated and 176 downregulated genes. A total of 82 overlapping genes between DEGs and DMGs were found, which were mainly involved in protein metabolism, regulation of the metabolic process and the immune system. Additionally, differential methylation or expression of several genes, including fibroblast activation protein α (*FAP*), interferon α inducible protein 27 (*IFI27*), laminin subunit γ2 (*LAMC2*), matrix metallopeptidase 1 (*MMP1*), serine peptidase inhibitor Kazal-type 5 (*SPINK5*) and zinc finger protein 662 (*ZNF662*), was significantly associated with the survival of OSCC patients, and their differential expression in OSCC patients was further confirmed by reverse transcription-quantitative polymerase chain reaction in OSCC and normal oral cell lines. Overall, *FAP*, *IFI27*, *LAMC2*, *MMP1*, *SPINK5* and *ZNF662* genes caused by epigenetic changes via DNA methylation may be associated with the development and progression of OSCC, and should be valuable OSCC therapeutic biomarkers.

Introduction

Oral squamous cell carcinoma (OSCC), the most prevalent type of SCC of the head and neck (HNSCC), typically behaves in an aggressive manner, frequently leading to local invasion and early lymph node metastasis (1). In addition, ~60% of head and neck cancer cases are diagnosed with advanced stage disease with a high lethality rate (2). Despite improvements in the treatment of OSCC, including surgery, radiotherapy and chemotherapy, its survival has not markedly improved and OSCC remains a life-threatening illness with a poor prognosis due to frequent development of local-regional recurrence or/and distant organ metastasis (3). It would be of great value to find useful biomarkers and prognostic molecular signatures to aid in the development of novel therapeutic strategies or chemopreventive agents.

DNA methylation serves an important role in cancer initiation, progression and metastasis, partially by transcriptional silencing of tumor suppressor genes (TSGs) (4). Khor *et al* (5) demonstrated 33 promoter hypermethylated genes such as dimethylarginine dimethylaminohydrolase 2 and dual specificity phosphatase 1 that were significantly silenced in OSCC, which may be used as hypermethylated-based biomarkers. Basu *et al* (6) reported a unique set of differentially methylated immune genes in OSCC patients. Furthermore, Clausen *et al* (7) revealed that WNT1 inducible signaling

Correspondence to: Dr Chenguang Zhao, Department of Emergency, Tianjin Stomatological Hospital, 75 Dagu Road, Tianjin 300041, P.R. China
E-mail: chenguangzhao237@163.com

*Contributed equally

pathway protein 1 hypomethylation contributed to lymph node metastasis in OSCC. Although previous studies have reported that the methylation of certain genes is associated with the pathogenesis of OSCC, the methylation of genes that have relevance to OSCC progression is not clearly documented.

High throughput genome-wide methylation studies offer novel ways to understand the significance of DNA methylation and its impact on gene regulation (8-10). Numerous studies have used the integration of DNA methylation data and gene expression data for identifying novel epigenetically deregulated genes involved in cancer development/progression (11,12). Li et al (13) identified several biomarkers for the early detection of buccal OSCC using the Illumina GoldenGate Methylation Cancer Panel. In the present study, comprehensive analyses of transcriptome microarray and methylation microarray data downloaded from a public database were performed. Differentially expressed genes (DEGs) and differentially methylated genes (DMGs) in OSCC samples were identified, and functional enrichment analysis was performed for overlapping genes between DEGs and DMGs to investigate their potential roles in OSCC progression. Survival analysis of the overlapping genes was performed to screen genes with prognostic significance in OSCC. This systematic approach should provide novel insights into the understanding of mechanisms underlying DNA methylation regulation associated with OSCC pathogenesis and progression, and contribute to the development of prognostic markers with potential clinical significance in OSCC treatment.

Materials and methods

Gene expression and DNA methylation datasets. The Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih. gov/geo/) is a gene expression/molecular abundance repository from the National Center for Biotechnology Information that archives and freely distributes microarray, next-generation sequencing and other functional genomics data submitted by the scientific community. In the present study, DNA methylation data from the study by Pickering et al (15) was retrieved from the GEO database with accession number GSE41114; this dataset includes data from 42 OSCC tumor samples and 4 normal control samples. The available clinical factors are provided in Table I. The β-value that estimates a ratio of DNA methylation signal intensity to the sum of the methylated and unmethylated intensities at each position was detected based on the GPL13534 Illumina HumanMethylation 450 BeadChip platform (Illumina, Inc., San Diego, CA, USA). Additionally, gene expression data from tumor tissues and adjacent non-tumor tissues from 6 clinical OSCC patients was downloaded from the GEO database with accession number GSE74530 (16). The platform of GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix; Thermo Fisher Scientific, Inc., Waltham, MA, USA) was used for the quantification of transcriptome expression profiles.

Assessment of genome-wide DNA methylation levels. The downloaded methylation data was preprocessed using the Illumina Methylation Analyzer (IMA) package in R (http://ima.r-forge.r-project.org/), which was designed to automate the pipeline for analyzing site (methylation locus)-level and region (all loci in a gene)-level methylation changes in epigenetic studies. Methylation sites with P-values >0.05 in >75% of the samples were filtered out, and samples with P-values $>1x10^{-5}$ at >75% of CpG sites were excluded from the analysis. Limma method (http://www.bioconductor.org/packages/release/bioc/html/limma.html) in IMA was used to identify differentially methylated CpG (dmCpG) sites with the cut-off points of $|\Delta\beta| >0.2$ and a Benjamini and Hochberg (BH)-corrected P-value (P_{BH}) <0.05 (17). OmicCircos is an R software package for generating high-quality circular plots and illustrates genomic data analyses. In the present study, OmicCircos was used to visualize the heatmap of the top 1,000 dmCpG sites.

Differential gene expression analysis. Raw data was normalized using RMA in the Bioconductor R package Affy (14). Probe annotations were obtained by using the Bioconductor hgu133plus2.db package and Limma package was used for the identification of DEGs. $|\log_2$fold change (FC)$| >1.5$ and adjusted P-value <0.05 [corrected by BH method (17)] were chosen as the threshold values for the DEGs. The heatmap of these DEGs was visualized with the OmicCircos package.

Association of DNA methylation and mRNA expression. The overlapping genes between DEGs and DMGs (genes containing dmCpG sites) were abbreviated as OSCC genes, which may be more relevant to OSCC than using DEGs or DMGs alone. In this study, further functional analysis was performed on this set of genes.

Functional enrichment analyses. Gene ontology (GO) enrichment analysis of OSCC genes was performed using the bioinformatics analysis tool WebGestalt (http://www. webgestalt.org/), and the GO biological process (BP) terms were identified. P-value adjustment was performed by BH multiple testing correction (17) and enrichment was considered significant only if P_{BH}<0.05. To investigate and understand the interactions among significantly enriched GO BP terms, cross-talk analysis of GO BP terms was conducted. A Cytoscape plug-in, EnrichmentMap (18), was used for the visualization of the GO BP enrichment map. GO BP terms were connected according to genes that overlapped and were grouped by functional similarity.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (19) enrichment analysis of OSCC genes was performed using the bioinformatics analysis tool KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/index.php) (20) with the criterion of P_{BH}<0.05.

Survival analysis of OSCC genes. Survival analysis of OSCC genes was performed using a novel and powerful web-based tool, Gene Expression Profiling Interactive Analysis (http://gepia.cancer-pku.cn/), which is based on The Cancer Genome Atlas (TCGA; https://cancergenome.nih.gov/) and Genotype-Tissue Expression data (http://commonfund.nih. gov/GTEx/). The associated disease 'HNSCC' was selected (OSCC was categorized into HNSCC in TCGA). The other parameters were set as the default.

Table I. Clinical factors of samples in the DNA methylation dataset.

Sample ID	Sex	Site
GSM1008735	Male	Tongue
GSM1008736	Male	Tongue
GSM1008737	Female	Tongue
GSM1008738	Male	Tongue
GSM1008739	Male	Tongue
GSM1008740	Female	Tongue
GSM1008741	Male	Tongue
GSM1008742	Male	Tongue
GSM1008743	Male	Tongue
GSM1008744	Male	Tongue
GSM1008745	Male	Tongue
GSM1008746	Female	Tongue
GSM1008747	Female	Tongue
GSM1008748	Female	Tongue
GSM1008749	Male	Tongue
GSM1008750	Male	Tongue
GSM1008751	Male	Tongue
GSM1008752	Male	Tongue
GSM1008753	Male	Tongue
GSM1008754	Male	Tongue
GSM1008755	Female	Tongue
GSM1008756	Male	FOM
GSM1008757	Male	Tongue
GSM1008758	Female	FOM
GSM1008759	Male	FOM
GSM1008760	Male	Tongue
GSM1008761	Male	FOM
GSM1008762	Female	Tongue
GSM1008763	Female	Tongue
GSM1008764	Female	Tongue
GSM1008765	Male	Tongue
GSM1008766	Female	FOM
GSM1008767	Female	Tongue
GSM1008768	Male	Buccal cavity
GSM1008769	Male	FOM
GSM1008770	Male	Alveolus
GSM1008771	Male	FOM
GSM1008772	Male	Tongue
GSM1008773	Male	FOM
GSM1008774	Male	Buccal cavity
GSM1008775	Male	Alveolus
GSM1008776	Male	Tongue
GSM1008777	Male	Tongue
GSM1008778	Male	Tongue
GSM1008779	NA	Blood
GSM1008780	NA	Blood

FOM, floor of the mouth; NA, not available.

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR). OSCC SCC25 and human normal oral epithelial HIOEC cell lines were purchased from the American Type Culture Collection (ATCC; Manassas, VA, USA) and cultured in RPMI-1640 (Sigma-Aldrich; Merck KGaA, Darmstadt, Germany) at 37°C in 5% CO_2, and supplemented with 10% FBS (PAA Laboratories GmbH, Cölbe, Germany) and 20 mM HEPES (Sigma-Aldrich; Merck KGaA). For RT-qPCR analysis, the total RNA of the SCC25 and HIOEC cells was isolated with TRIzol reagent (Invitrogen; Thermo Fisher Scientific, Inc.). The cDNA was synthesized by Transcriptor First strand cDNA Synthesis kit (Roche Diagnostics, Basel, Switzerland) using random primers and subjected to PCR amplification using rTaq polymerase (Takara Bio, Inc., Shiga, Japan). For each target gene, the PCR mixtures (Applied Biosystems; Thermo Fisher Scientific, Inc.) containing 2 μl of the diluted cDNA were prepared in a final volume of 20 μl. The PCR was performed with a total volume of 20 ml reaction mixture using FastStart Universal SYBR Green Master kit (Roche Diagnostics). qPCR assays were conducted in polypropylene 96-well plates on an ABI Prism 7000 sequence detection system (Applied Biosystems; Thermo Fisher Scientific, Inc.). Non-template controls were used to detect non-specific amplification. The quantitation cycle (Cq) value of each target product was determined and ΔCq between target and endogenous control was calculated. The difference in ΔCq values of the 2 groups (ΔΔCq) was used to calculate the fold increase ($F=2^{-\Delta\Delta Cq}$) (21) and to determine the changes in target gene expression between control and sample groups. β-actin was used as the control gene. The primer pairs used for amplification are shown in Table II. GraphPad Prism Version 5.0 software for windows (GraphPad Software Inc., La Jolla, CA, USA) was used for statistical analysis.

Statistical analysis. All statistical analyses in this study were conducted based on R 3.4.3 (https://www.r-project.org/) and GraphPad Prism 5.0 (https://www.graphpad.com/scientific-software/prism/). |Log2FC|>1.5 and |Δβ|>0.2 with an adjusted P-value of <0.05 was applied for screening of the DEGs and dmCpGs, respectively. Data analysis was performed by one-way analysis of variance followed by Tukey's test for multiple group comparisons, and Student's t-test was used for pairwise comparisons. The Kaplan-Meier method was used for patient survival estimations and the log-rank test was used for survival comparisons between different groups. P<0.05 was used to indicate a statistically significant difference.

Results

DNA methylation profile analysis between OSCC samples and control samples. Subsequent to data preprocessing for all samples, 461,304 out of the original 485,577 CpG sites (95.0%) were retained, which were located in different regions, including that within 1,500 bps of a transcription start site (TSS1500), within 200 bp of the TSS (TSS200), the 5'-untranslated region (UTR), the first exon, the body, the 3'UTR, the island, the N shelf, the N shore and the S shelf. β-values of the 461,304 CpG sites are presented in Fig. 1A. It was found that the global methylation level of CpG sites in the normal control samples was higher compared with those in the OSCC samples. Additionally, the regions at close proximity to the promoter (TSS200) and 1,500 bp upstream of the promoter (TSS1500) were designated as the promoter region. A total of 140,003 CpG sites were found

Table II. Primer sequences for reverse transcription-quantitative polymerase chain reaction.

Genes	Forward primer (5'-3')	Reverse primer (5'-3')	Product length, bp	Temperature, ˚C
FAP	ATGAGCTTCCTCGTCCAATTCA	AGACCACCAGAGAGCATATTTTG	215	58
IFI27	TGCTCTCACCTCATCAGCAGT	CACAACTCCTCCAATCACAACT	115	60
LAMC2	GACAAACTGGTAATGGATTCCGC	GACAAACTGGTAATGGATTCCGC	98	60
MMP1	AAAATTACACGCCAGATTTGCC	GGTGTGACATTACTCCAGAGTTG	82	58
SPINK5	ATAGCCACAGTGTCAGTGCTT	TGTTGCGTAAGGCTTGTGTTC	202	58
ZNF662	CAAACCTGATTCGTCACCAGA	TGTTGCGTAAGGCTTGTGTTC	100	60
ACTB	AGCGAGCATCCCCCAAAGTT	GGGCACGAAGGCTCATCATT	285	60

FAP, fibroblast activation protein α; *IFI27*, interferon α inducible protein 27; *LAMC*, laminin subunit γ2; *MMP1*, matrix metallopeptidase 1; *SPINK5*, serine peptidase inhibitor Kazal-type 5; *ZNF662*, zinc finger protein 662; *ACTB*, β-actin.

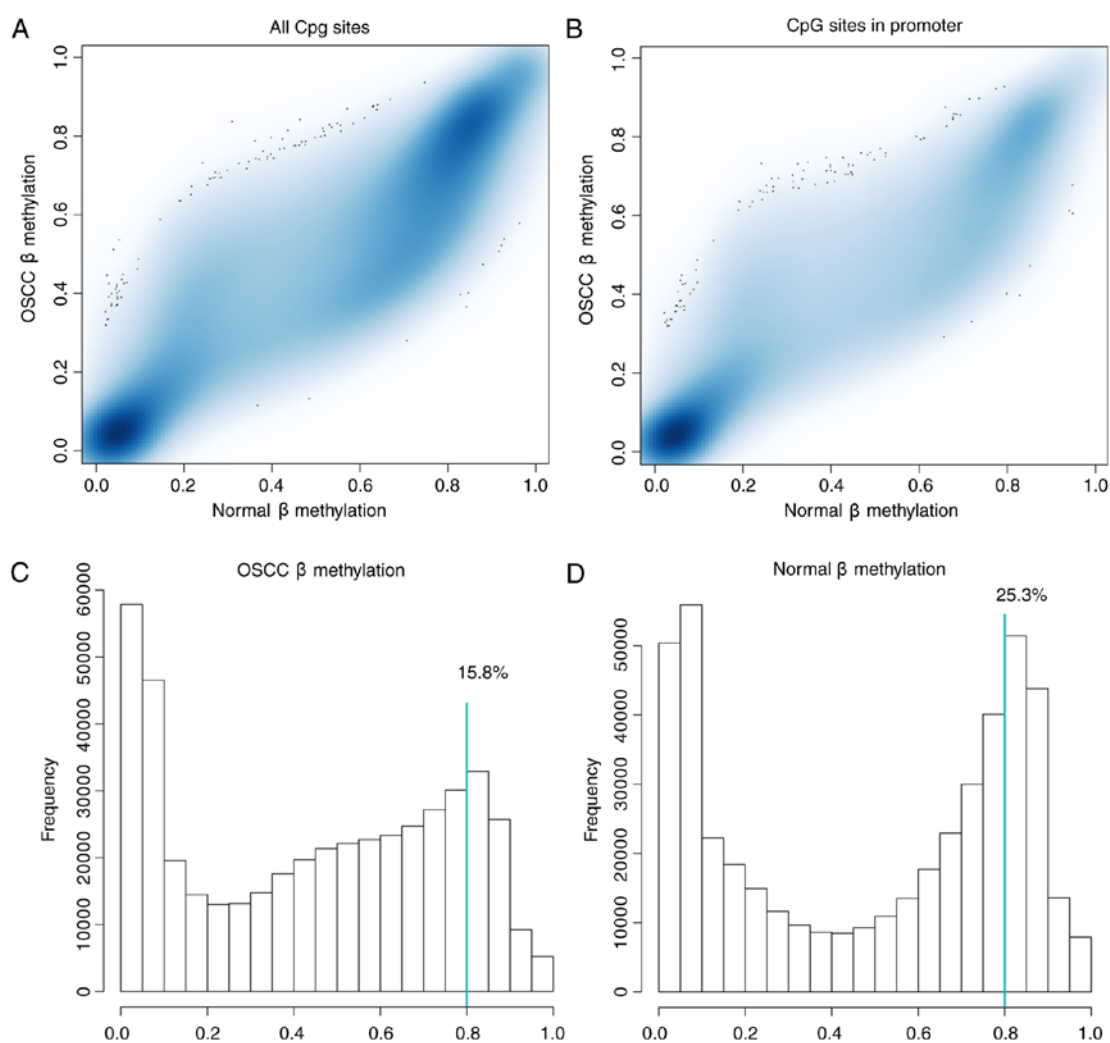Figure 1. β-methylation distribution. (A) β-values of the 461,304 CpG sites. The vertical axis indicates the methylation β-value of CpG sites in the OSCC samples and the horizontal axis indicates the methylation β-value of the CpG sites of the normal control samples. (B) β-methylation in the promoter region. The vertical axis represents the methylation β-value of the CpG sites in the promoter region of the OSCC sample, and the horizontal axis represents the methylation β-value of the CpG sites in the normal control samples. (C) Histogram representing the β-value distribution in different bins in OSCC samples. (D) Histogram representing the β-value distribution in different bins in normal samples. OSCC, oral squamous cell carcinoma.

in the promoter region (Fig. 1B) and the level of methylation at the CpG sites in the promoter region showed a higher level in the OSCC group than that in the control samples. Of those 461,304 CpG sites, 15.8% CpG sites had β-value >0.8 in the OSCC samples (Fig. 1C), and 25.3% CpG sites were hypermethylated (β-value >0.8) in the normal control samples (Fig. 1D).
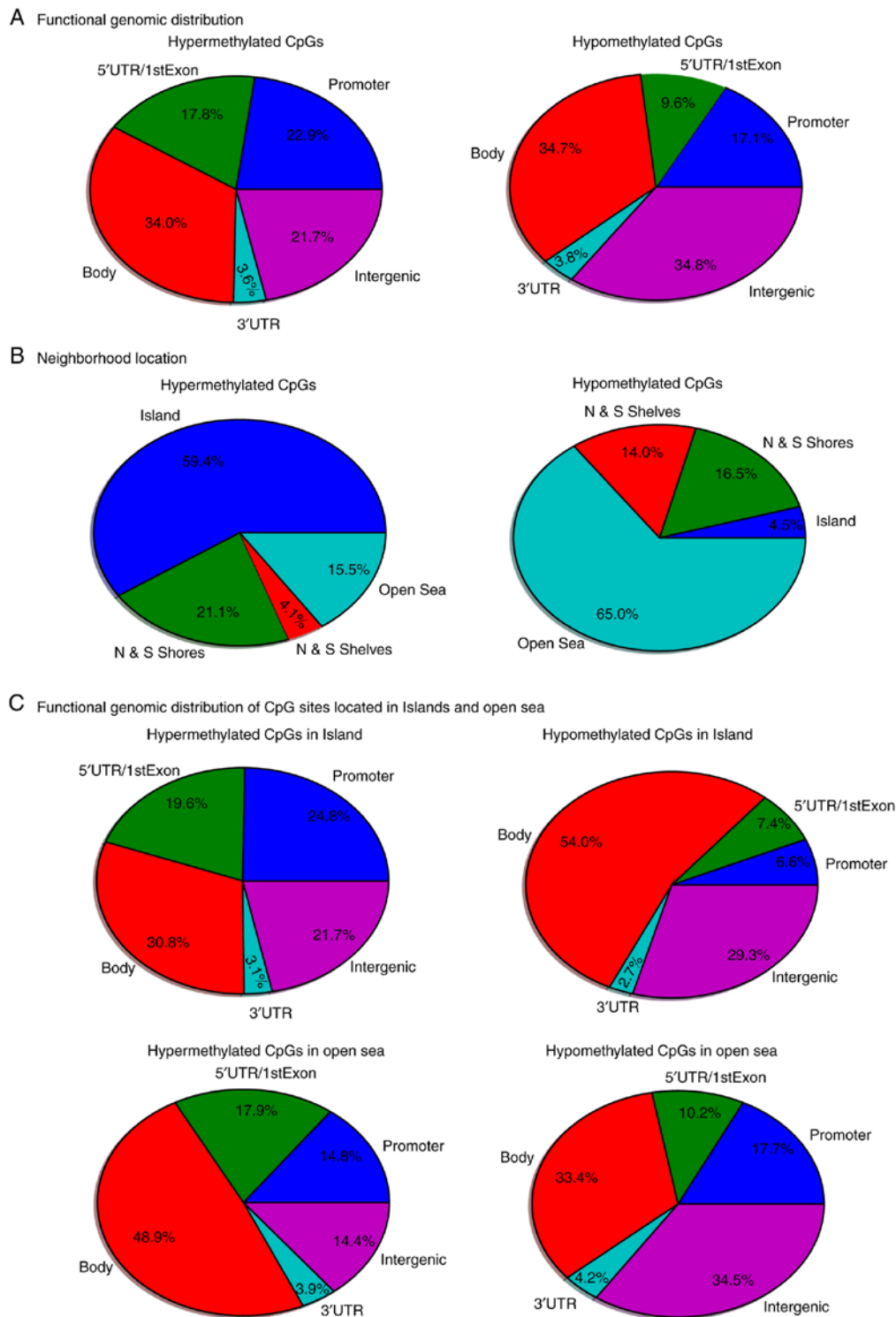
Figure 2. Functional genomic and neighborhood location distribution of differentially methylated CpG sites. (A) Functional genomic distribution and (B) neighborhood location of hypermethylated and hypomethylated CpG sites in OSCC samples compared with those in controls. Promoter region is defined as the combination of TSS200 and TSS1500, which represent sites that are located 200 and 1,500 bp, respectively, from a TSS. Intergenic regions are defined as the remainder of locations located between genes. Shores and shelves are composed of CpG methylation sites located 0-2 and 2-4 kb, respectively, from the nearest CpG island; open sea is defined as CpG methylation sites located >4 kb from a CpG island. (C) Functional genomic distribution of CpG sites located in islands and open sea. OSCC, oral squamous cell carcinoma; TSS, transcription start site; UTR, untranslated region.

Identification of dmCpG sites. In the comparison of DNA methylation between the OSCC group and the control group, 40,115 dmCpGs were observed to reach a liberal significance threshold of methylation difference of at least 20% ($|\Delta\beta|>0.2$) and $P_{BH}<0.05$. These 40,115 dmCpGs covered 3,360 genes, i.e., DMGs. In addition, 6,736 dmCpGs were hypermethylated and 33,379 were hypomethylated in the OSCC tissue samples. The functional genomic distribution of the dmCpG sites in the OSCC samples is shown in Fig. 2A. CpG sites in the promoter, gene body and intergenic regions were
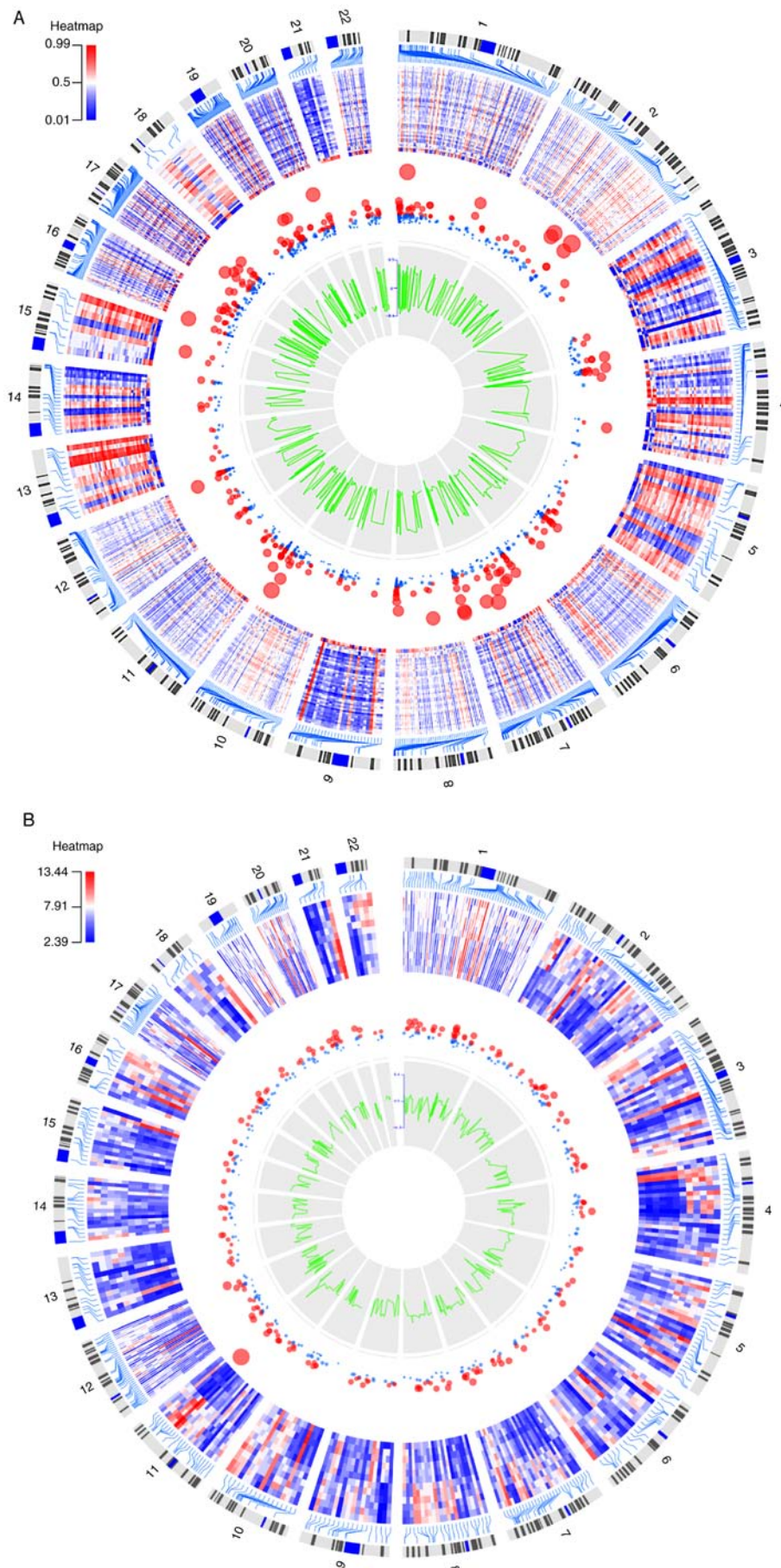
Figure 3. Circular plots of (A) the top 1,000 differentially methylated CpG sites and (B) the differentially expressed genes. The tracks from outside to inside are the genome positions by chromosomes, heatmap, adjusted P-value (larger nodes indicate smaller adjusted P-values), and Δβ or log2fold-change. Colors from green to red correspond to low to high expression values.

Table III. Kyoto Encyclopedia of Genes and Genomes pathways enriched by overlapping genes between differentially expressed genes and differentially methylated genes.

Pathway ID	Description	P-value	Corrected P-value
hsa00350	Tyrosine metabolism	1.30×10^{-6}	1.27×10^{-4}
hsa00010	Glycolysis/gluconeogenesis	1.46×10^{-5}	7.14×10^{-4}
hsa05204	Chemical carcinogenesis	3.11×10^{-5}	1.02×10^{-3}
hsa01100	Metabolic pathways	2.51×10^{-4}	6.14×10^{-3}
hsa00830	Retinol metabolism	3.82×10^{-4}	7.39×10^{-3}
hsa00982	Drug metabolism-cytochrome P450	4.52×10^{-4}	7.39×10^{-3}
hsa00980	Metabolism of xenobiotics by cytochrome P450	5.30×10^{-4}	7.42×10^{-3}
hsa05323	Rheumatoid arthritis	9.84×10^{-4}	1.21×10^{-2}
hsa05146	Amoebiasis	1.28×10^{-3}	1.40×10^{-2}
hsa04668	TNF signaling pathway	1.67×10^{-3}	1.64×10^{-2}
hsa05219	Bladder cancer	3.58×10^{-3}	3.19×10^{-2}
hsa00071	Fatty acid degradation	4.08×10^{-3}	3.34×10^{-2}
hsa05202	Transcriptional misregulation in cancer	6.47×10^{-3}	4.64×10^{-2}
hsa05150	*Staphylococcus aureus* infection	6.64×10^{-3}	4.64×10^{-2}
hsa04062	Chemokine signaling pathway	7.17×10^{-3}	4.69×10^{-2}

generally highly hypermethylated or hypomethylated. The neighborhood locations of all dmCpG sites are shown in Fig. 2B; 59.4% of the hypermethylated CpG sites were in the island and 65.0% of the hypomethylated CpG sites were in the open sea (located >4 kb from a CpG island), whereas only 4.5% hypomethylated CpG sites were in the island and 15.5% of the hypermethylated CpG sites were in the open sea. This result prompted the comparison of the functional genomic distribution of the dmCpG sites in islands and open sea (Fig. 2C). Among the 6,736 hypermethylated CpG sites, 4,001 CpG sites were located in the island and 1,044 CpG sites in the open sea. Among the 33,379 hypomethylated CpG sites, 1,487 CpG sites were located in the islands and 21,710 CpG sites in the open sea. The heatmap of the top 1,000 dmCpGs with BH-adjusted P-value and β-value is shown in Fig. 3A.

DEG screening between OSCC samples and control samples. With the criteria of |log$_2$FC|>1.5 and an adjusted P-value of <0.05, a total of 508 DEGs were identified, consisting of 332 upregulated genes and 176 downregulated genes. The heatmap of these 508 DEGs with log$_2$FC and adjusted P-value is shown in Fig. 3B.

Association between DNA methylation and mRNA expression. There were 82 overlapping genes between the 3,360 DMGs and 508 DEGs, termed the OSCC genes; this set of genes may be more relevant to OSCC. Further functional analysis was performed on these 82 OSCC genes.

GO and KEGG pathway enrichment analysis of OSCC genes. In total, 48 significant GO BP terms were identified (Fig. 4A). Using GO enrichment map view (Fig. 4B), these 48 significant GO BP terms were mainly grouped into 3 clusters: A cluster associated with protein metabolism (including 'proteolysis', 'negative regulation of proteolysis', 'regulation of proteolysis', 'negative regulation of peptidase activity' and 'regulation of peptidase activity'), a cluster associated with the regulation of the metabolic process (including 'extracellular matrix disassembly', 'extracellular matrix organization' and 'extracellular structure organization') and a cluster associated with immune system (including 'immune response', 'cell migration', 'regulation of defense response', 'regulation of inflammatory response', 'defense response' and leukocyte migration').

In addition, 15 KEGG terms were enriched (Table III), including 'tyrosine metabolism' and 'glycolysis/gluconeogenesis'.

Survival analysis for OSCC genes. To investigate the associations between OSCC overall survival and the expression of these OSCC genes, Kaplan-Meier curve analysis was performed to obtain the prognostic signature (Fig. 5). As a result, 6 genes were found, namely fibroblast activation protein α (FAP; upregulated), interferon α inducible protein 27 (*IFI27*; upregulated), laminin subunit γ2 (*LAMC*; upregulated), matrix metallopeptidase 1 (*MMP1*; upregulated), serine peptidase inhibitor Kazal-type 5 (*SPINK5*; downregulated) and zinc finger protein 662 (*ZNF662*; downregulated), were significantly associated with the survival of patients with OSCC.

RT-qPCR analysis. Differences in the expression of *FAP*, *IFI27*, *LAMC2*, *MMP1*, *SPINK5* and *ZNF662* between OSCC SCC25 cells and the human normal oral epithelial HIOEC cell line were investigated through RT-qPCR. Consistent with the results from the microarray analysis, the expression of *FAP*, *IFI27*, *LAMC2* and *MMP1* was significantly upregulated, while the expression of *SPINK5* and *ZNF662* was significantly downregulated in OSCC cells compared with that in normal cells, as shown in Fig. 6.
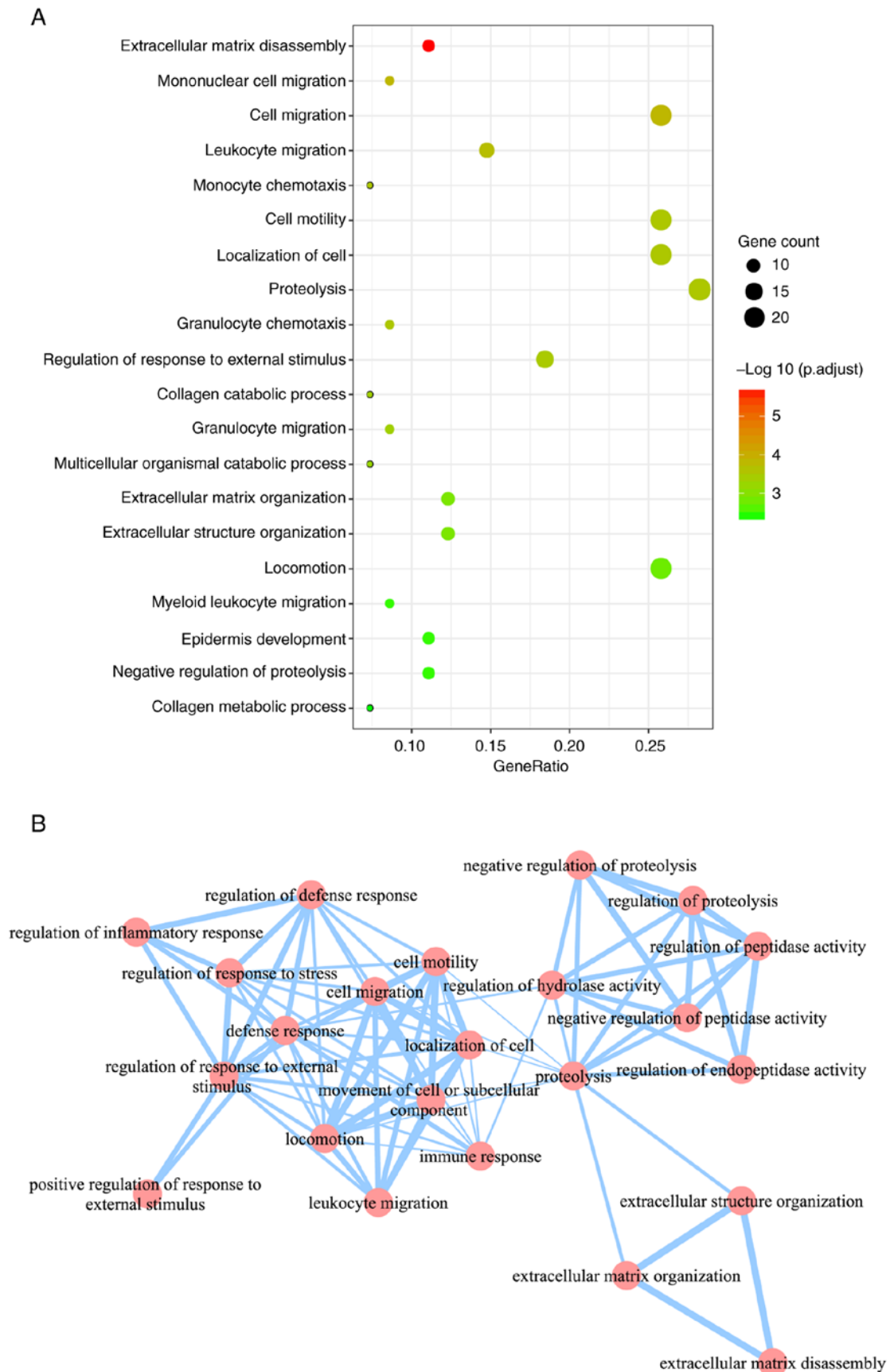
Figure 4. GO enrichment analysis of overlapping genes. (A) The top 20 enriched GO BP terms. The size of the dot represents the number of OSCC genes included in the GO BP term; the abscissa GeneRatio indicates the ratio of the number of OSCC genes mapped to a GO BP term to the total number of OSCC genes. (B) Cross-talk analysis of significantly enriched GO terms. Nodes and edges represent GO BP terms and associations between two terms respectively, larger node size and thicker edge indicates more genes contained in the GO BP terms and more overlapping genes between two GO BP terms. GO, Gene Ontology; BP, Biological Process; OSCC genes, overlapping genes between differentially expressed genes and differentially methylated genes; OSCC, oral squamous cell carcinoma.
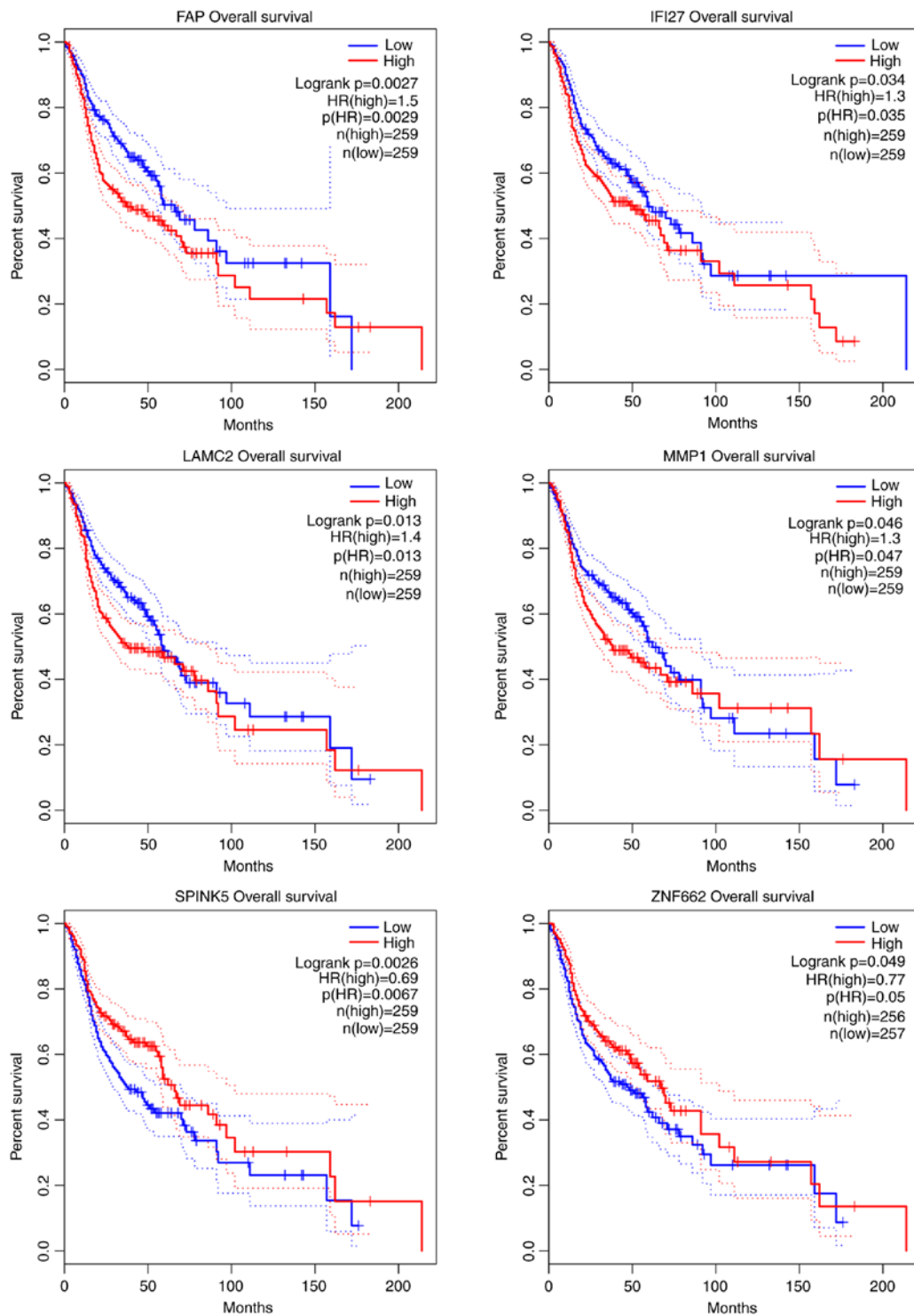
Figure 5. Kaplan-Meier curve analysis of genes for the overall survival in OSCC patients. *OSCC*, oral squamous cell carcinoma. *FAP*, fibroblast activation protein α; *IFI27*, interferon α inducible protein 27; LAMC, laminin subunit γ2; *MMP1*, matrix metallopeptidase 1; *SPINK5*, serine peptidase inhibitor Kazal-type 5; *ZNF662*, zinc finger protein 662; HR, hazard ratio; p(HR), log-rank test P-value.

Discussion

In the current study, it was found that 40,115 dmCpGs spanning 3,360 DMGs were aberrantly methylated in OSCC samples compared with those in normal samples; CpG sites in the promoter, gene body and intergenic regions were generally highly hypermethylated or hypomethylated. Additionally, 508 DEGs were identified in the OSCC samples, including 332 upregulated and 176 downregulated genes. A total of 82 overlapping genes between DEGs and DMGs were found, which were mainly involved in protein metabolism, regulation of the metabolic process and the immune system. Additionally, differential methylation or expression of several genes, namely *FAP, IFI27, LAMC2, MMP1, SPINK5* and *ZNF662,* were significantly associated with the survival of patients with OSCC.
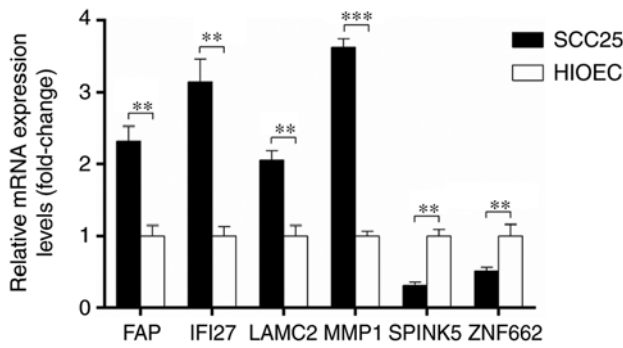
Figure 6. Reverse transcription-quantitative polymerase chain reaction analysis of *FAP, IFI27, LAMC2, MMP1 SPINK5* and *ZNF662* between oral squamous cell carcinoma SCC25 cells and the human normal oral epithelial HIOEC cell line. FAP, fibroblast activation protein α; *IFI27*, interferon α inducible protein 27; *LAMC*, laminin subunit γ2; *MMP1*, matrix metallopeptidase 1; *SPINK5*, serine peptidase inhibitor Kazal-type 5; *ZNF662*, zinc finger protein 662. **P<0.01 and ***P<0.001.

DNA methylation in different genomic regions can alter gene expression. A previous study showed a causal association between gene body DNA methylation and gene expression (22). Hypermethylation of CpG islands within the promoter regions of TSGs is believed to serve a crucial role in the development of carcinogenesis (23). In the present study, to better address the function of dmCpGs sites, their neighborhood locational distribution was determined. As a result, it was found that >50% of hypermethylated sites were located in CpG islands, while >50% of the hypomethylated CpG sites were in the open sea. However, regardless of the neighborhood location of the promoter methylation sites, evidence has shown that DNA methylation in the promoter region is most often associated with transcriptional downregulation (24).

In addition, the present study found that methylation and expression changes in the genome of patients with OSCC, according to functional enrichment analysis, interfere with protein metabolism, glycolysis/gluconeogenesis and the immune system. Proteolytic processing of E-cadherin that suppresses cell-cell adhesion of OSCC cells may facilitate the progression of OSCC (25). Increased glucose transport and metabolism has been reported to be associated with poor prognosis in patients with OSCC (26). Moreover, immune cell dysfunction in OSCC patients is an important factor influencing tumor growth (27). One previous study indicated that the modulation of functional dynamics of regulatory T cells may be useful for immunotherapeutic strategy for patients with OSCC (27). Regarding the biological processes and KEGG pathways likely to be impaired by methylation or expression changes in the present study, we suggested involvement of significant methylated genes in the progression of OSCC.

The present study found 82 genes that were overlapping between the 3,360 DMGs and 508 DEGs, of which 6 genes (*FAP, IFI27, LAMC2, MMP1, SPINK5* and *ZNF662*) were predicted to be significantly associated with the survival of OSCC patients. Wang *et al* (28) demonstrated that the downregulation of FAP suppresses cell proliferation and metastasis in OSCC. In a study by Li *et al* (29), OSCC was closely allied to certain key genes, including *IFI27*. Additionally, *IFI27* was

found to be dysregulated in patients with tongue squamous cell carcinoma (30). *LAMC2* was found to be significantly associated with lymph node metastasis in OSCC (31), while *MMP1* is a potential oral cancer marker (32,33). One previous study reported *SPINK5* as one of the genes that were downregulated in HNSCC (34). Therefore, we hypothesize that the dysregulation of these genes caused by epigenetic changes may be a suitable mechanism linked to the development of OSCC and that these 6 genes may serve as prognostic biomarkers in OSCC treatment. However, the potential clinical significance of these genes should be verified by further experiments.

In summary, genome-wide DNA methylation profiling revealed a set of DMGs in OSCC patients. Moreover, by integration of gene expression data, it was suggested that the dysregulation of *FAP, IFI27, LAMC2, MMP1, SPINK5* and *ZNF662* genes caused by epigenetic changes via DNA methylation may be associated with the development and progression of OSCC, and that these genes may be useful prognostic markers with potential clinical significance in OSCC treatment, thus deserving further investigation.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Gene Expression Omnibus repository of the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41114).

Authors' contributions

CZ and HZ conceived and designed the project; JZ acquired the data; CZ, HL and JW analyzed and interpreted the data; and HZ and HL wrote the paper and revised it for important intellectual content. CZ and HZ agree to be accountable for all aspects of the work in ensuring that questions associated with the accuracy or integrity of any part of the study are appropriately investigated and resolved. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Ehtesham H, Safdari R, Mansourian A, Tahmasebian S, Mohammadzadeh N, Ghazisaeedi M and Bashiri A: Clinical decision support system, a potential solution for diagnostic accuracy improvement in oral squamous cell carcinoma: A systematic review. J Oral Health Oral Epidemiol 6: 187-195, 2017.

2. Marur S and Forastiere AA: Head and neck squamous cell carcinoma: Update on epidemiology, diagnosis, and treatment. Mayo Clin Proc 91: 386-396, 2016.

3. Velmurugan BK, Yeh KT, Lee CH, Lin SH, Chin MC, Chiang SL, Wang ZH, Hua CH, Tsai MH, Chang JG and Ko YC: Acidic leucine-rich nuclear phosphoprotein-32A (ANP32A) association with lymph node metastasis predicts poor survival in oral squamous cell carcinoma patients. Oncotarget 7: 10879-10890, 2016.

4. Chi HC, Tsai CY, Tsai MM and Lin KH: Impact of DNA and RNA methylation on radiobiology and cancer progression. Int J Mol Sci 19: E555, 2018.

5. Khor GH, Froemming GR, Zain RB, Abraham MT, Omar E, Tan SK, Tan AC, Vincent-Chong VK and Thong KL: DNA methylation profiling revealed promoter hypermethylation-induced silencing of p16, DDAH2 and DUSP1 in primary oral squamous cell carcinoma. Int J Med Sci 10: 1727-1739, 2013.

6. Basu B, Chakraborty J, Chandra A, Katarkar A, Baldevbhai JR, Dhar Chowdhury D, Ray JG, Chaudhuri K and Chatterjee R: Genome-wide DNA methylation profile identified a unique set of differentially methylated immune genes in oral squamous cell carcinoma patients in India. Clin Epigenetics 9: 13, 2017.

7. Clausen MJ, Melchers LJ, Mastik MF, Slagter-Menkema L, Groen HJ, van der Laan BF, van Criekinge W, de Meyer T, Denil S, Wisman GB, et al: Identification and validation of WISP1 as an epigenetic regulator of metastasis in oral squamous cell carcinoma. Genes Chromosomes Cancer 55: 45-59, 2016.

8. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, Shu J, Chen X, Waterland RA and Issa JP: Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet 3: 2023-2036, 2007.

9. Garcia MP and Garcia-Garcia A: Epigenome and DNA methylation in oral squamous cell carcinoma. Methods Mol Biol 863: 207-219, 2012.

10. Jithesh PV, Risk JM, Schache AG, Dhanda J, Lane B, Liloglou T and Shaw RJ: The epigenetic landscape of oral squamous cell carcinoma. Br J Cancer 108: 370-379, 2013.

11. Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Siegmund KD, Koss MN, Hagen JA, Lam WL, et al: Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. Genome Res 22: 1197-1211, 2012.

12. Yoo S, Takikawa S, Geraghty P, Argmann C, Campbell J, Lin L, Huang T, Tu Z, Foronjy RF, Spira A, et al: Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. PLoS Genet 11: e1004898, 2015.

13. Li YF, Hsiao YH, Lai YH, Chen YC, Chen YJ, Chou JL, Chan MW, Lin YH, Tsou YA, Tsai MH and Tai CK: DNA methylation profiles and biomarkers of oral squamous cell carcinoma. Epigenetics 10: 229-236, 2015.

14. Gautier L, Cope L, Bolstad BM and Irizarry RA: Affy-analysis of affymetrix genechip data at the probe level. Bioinformatics 20: 307-315, 2004.

15. Pickering CR, Zhang J, Yoo SY, Bengtsson L, Moorthy S, Neskey DM, Zhao M, Ortega Alves MV, Chang K, Drummond J, et al: Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. Cancer Discov 3: 770-781, 2013.

16. Oghumu S, Knobloch TJ, Terrazas C, Varikuti S, Ahn-Jarvis J, Bollinger CE, Iwenofu H, Weghorst CM and Satoskar AR: Deletion of macrophage migration inhibitory factor inhibits murine oral carcinogenesis: Potential role for chronic pro-inflammatory immune mediators. Int J Cancer 139: 1379-1390, 2016.

17. Benjamini Y and Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Royal Statistical Society Series B 57: 289-300, 1995.

18. Merico D, Isserlin R, Stueker O, Emili A and Bader GD: Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. PLoS One 5: e13984, 2010.

19. Kanehisa M and Goto S: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30, 2000.

20. Wu J, Mao X, Cai T, Luo J and Wei L: KOBAS server: A web-based platform for automated annotation and pathway identification. Nucleic Acids Res 34: W720-W724, 2006.

21. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. Methods 25: 402-408, 2001.

22. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA and Liang G: Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell 26: 577-590, 2014.

23. Khatami F, Larijani B, Heshmat R, Keshtkar A, Mohammadamoli M, Teimoori-Toolabi L, Nasiri S and Tavangar SM: Meta-analysis of promoter methylation in eight tumor-suppressor genes and its association with the risk of thyroid cancer. PLoS One 12: e0184892, 2017.

24. Jones PA: Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat Rev Genet 13: 484-492, 2012.

25. Hayashido Y, Hamana T, Yoshioka Y, Kitano H, Koizumi K and Okamoto T: Plasminogen activator/plasmin system suppresses cell-cell adhesion of oral squamous cell carcinoma cells via proteolysis of E-cadherin. Int J Oncol 27: 693-698, 2005.

26. Kunkel M, Reichert TE, Benz P, Lehr HA, Jeong JH, Wieand S, Bartenstein P, Wagner W and Whiteside TL: Overexpression of Glut-1 and increased glucose metabolism in tumors are associated with a poor prognosis in patients with oral squamous cell carcinoma. Cancer 97: 1015-1024, 2003.

27. Aggarwal S, Sharma SC and N Das S: Dynamics of regulatory T cells (T_{regs}) in patients with oral squamous cell carcinoma. J Surg Oncol 116: 1103-1113, 2017.

28. Wang H, Wu Q, Liu Z, Luo X, Fan Y, Liu Y, Zhang Y, Hua S, Fu Q, Zhao M, et al: Downregulation of FAP suppresses cell proliferation and metastasis through PTEN/PI3K/AKT and Ras-ERK signaling in oral squamous cell carcinoma. Cell Death Dis 5: e1155, 2014.

29. Li G, Li X, Yang M, Xu L, Deng S and Ran L: Prediction of biomarkers of oral squamous cell carcinoma using microarray technology. Sci Rep 7: 42105, 2017.

30. Boldrup L, Gu X, Coates PJ, Norberg-Spaak L, Fahraeus R, Laurell G, Wilms T and Nylander K: Gene expression changes in tumor free tongue tissue adjacent to tongue squamous cell carcinoma. Oncotarget 8: 19389-19402, 2017.

31. Zanaruddin SN, Saleh A, Yang YH, Hamid S, Mustafa WM, Khairul Bariah AA, Zain RB, Lau SH and Cheong SC: Four-protein signature accurately predicts lymph node metastasis and survival in oral squamous cell carcinoma. Hum Pathol 44: 417-426, 2013.

32. Hashimoto T, Uchida K, Okayama N, Imate Y, Suehiro Y, Hamanaka Y, Ueyama Y, Shinozaki F, Yamashita H and Hinoda Y: Association of matrix metalloproteinase (MMP)-1 promoter polymorphism with head and neck squamous cell carcinoma. Cancer Lett 211: 19-24, 2004.

33. Yen CY, Chen CH, Chang CH, Tseng HF, Liu SY, Chuang LY, Wen CH and Chang HW: Matrix metalloproteinases (MMP) 1 and MMP10 but not MMP12 are potential oral cancer markers. Biomarkers 14: 244-249, 2009.

34. Jayakumar A, Kang Ya, Mitsudo K, Henderson Y, Frederick MJ, Wang M, El-Naggar AK, Marx UC, Briggs K and Clayman GL: Expression of LEKTI domains 6-9' in the baculovirus expression system: Recombinant LEKTI domains 6-9' inhibit trypsin and subtilisin A. Protein Expr Purif 35: 93-101, 2004.