World Journal of Surgical Oncology

RESEARCH   Open Access

Identification of potential CpG sites for oral squamous cell carcinoma diagnosis via integrated analysis of DNA methylation and gene expression

Le Chen and Dong Wang[*]

Abstract

Background: Oral squamous cell carcinoma (OSCC) accounts for more than 90% of the oral carcinomas and has a high fatality rate. This study aimed to identify potentially diagnostic biomarkers of OSCC through integrated analysis of DNA methylation and gene expression profiles.

Methods: The DNA methylation profiles of OSCC patients from The Cancer Genome Atlas (TCGA) were analyzed to screen patients with CpG island methylator phenotype (CIMP) and investigate the relationship between CIMP and survival probability of OSCC patients. Differential methylation and expression analyses of the paired OSCC tumor and paracancerous samples from TCGA were performed. Logistic regression model was established, and the accuracy of this diagnostic model for OSCC was evaluated in validation sets from Gene Expression Omnibus (GEO).

Results: OSCC patients with CIMP had lower survival probability than those without CIMP. The cg02860732 and cg04342955 were determined as candidate diagnostic methylation sites for OSCC. Logistic regression model was established based on cg02860732 and cg04342955 showed relatively high diagnostic accuracy in OSCC.

Conclusions: A diagnostic model for OSCC was identified based on the methylation sites cg02860732 and cg04342955, which might be favorable for the diagnosis of OSCC.

Keywords: OSCC, Diagnosis, DNA methylation, Gene expression, Logistic regression

Introduction

Oral squamous cell carcinoma (OSCC) is a malignant tumor occurring in the oral cavity accounts for more than 90% of the oral carcinomas [1]. As a frequent cancer, there are approximately 500,000 new cases diagnosed with OSCC annually [2]. In spite of the advances in therapeutic strategies including chemotherapy, radiotherapy, and surgery, OSCC still has a relatively high mortality of 48%, remaining to be a severe health burden worldwide [3, 4]. Although biopsy and histopathological exams were gold standard for OSCC diagnosis, a large percentage of OSCC cases are diagnosed at advanced stage, which results in the grave prognosis, it is generally believed that early and accurate diagnosis plays a pivotal role in ameliorating the survival rate and prognosis of OSCC [5]. Genetic molecular mechanism exploration of occurrent and development of OSCC was critical for monitoring disease progression earlier.

An increasing number of studies have been investigating the OSCC-related biomarkers. A previously reported study showed that cytokeratins (CKs) like CK17, which exhibited the most pronounced upregulation among the identified CKs in OSCC samples compared with the

* Correspondence: tjwangdong1@outlook.com
Department of Stomatology, The Second Hospital of Tianjin Medical University, 23 Pingjiang Road Intersection, Hexi District, Tianjin 300211, China

normal controls, were a possible diagnostic biomarker in OSCC [6]. Ishiwata et al. found that heat shock factor 1 (HSF1) expression level in nuclear had close association with OSCC development, including the tumor size and histopathologic types, which might act as a potential tool for OSCC diagnosis [7]. Tang et al. indicated that some long non-coding RNAs (lncRNAs) were abnormally expressed in OSCC, presented detectable amounts in saliva, and could be adopted as non-aggressive diagnostic biomarker in OSCC [8]. Except for the aberrant expressions of mRNA and lncRNA, epigenetics has been a promising field in cancer research, including DNA methylation which occurs in the CpG islands near gene transcription start sites [9]. Alteration of DNA methylation is able to affect the gene expression, as well as diverse molecular mechanisms; thus, aberrantly methylated CpG sites are regarded as promising biomarkers in various cancers including OSCC [10, 11]. For example, Foy et al. found that the abnormal methylation of genes *FOXI2*, *AGTR1*, *PENK*, and *LINE1* is a possible indicator of OSCC development [12]. Langevin et al. identified 7 methylation sites which were related to the prognosis of OSCC patients and might be used as non-invasive detection tools in clinical practice [13]. However, study of epigenetics focuses on the mechanism of OSCC progression. To obtain the reliable diagnosis signature, study of differential expression of genes should be analyzed between cancer samples and paracancerous normal samples. After comprehensively analyzing the two expression profiles, the downstream diagnosis could be found more accurately and reasonably based on the methylation molecular mechanism of OSCC.

In this study, the DNA methylation and gene expression profiles of OSCC samples from The Cancer Genome Atlas (TCGA) were comprehensively analyzed to identify potentially diagnostic methylation sites. The diagnostic value of the screened biomarkers was further validated in two validation sets from the Gene Expression Omnibus (GEO) by logistic regression analysis. Our study might provide novel promising signatures for the diagnosis of OSCC.

Material and methods

Data source

The gene expression profiles of 75 OSCC samples were obtained from The Cancer Genome Atlas (TCGA, www.cancergenome.nih.gov), including 65 OSCC tumor samples and 5 pairs of tumor and paracancerous samples. The HTSeq-Counts data were downloaded for subsequent analysis.

The DNA methylation data were downloaded from TCGA and Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). The TCGA dataset consisted of the DNA methylation data and clinical information of 75 OSCC samples, including 65 OSCC tumor samples and 5 pairs of tumor and paracancerous samples. The GEO dataset GSE87053 included 11 OSCC tumor samples and 10 normal controls, and GSE123781 included 23 OSCC tumor samples and 18 normal controls. The DNA methylation profiles were all measured by Illumina Human Methylation 450 (HM450) arrays.

Identification of CpG island methylator phenotype (CIMP)

The methylation sites with standard deviation (SD) > 0.2 in 70 OSCC patients and $\beta < 0.05$ in 5 paracancerous samples were selected [14, 15]. Based on these methylation sites, the 70 OSCC patients were clustered using K-means method; then, the patients with CIMP were identified according to the methylation level of each cluster.

Differential methylation and expression analyses

Differential methylation analysis was performed for the 5 pairs of OSCC and paracancerous samples. The methylation sites with missing values or those on the sex chromosomes were firstly removed; then, paired t test was used for differential methylation analysis, with $|\triangle\beta| > 0.2$ and $p < 0.05$ as thresholds for significantly differential methylation.

The differential expression analysis was performed using edgeR package in the R software, with |log2 Fold Change| > 1 and $p < 0.05$ as thresholds for significantly differential expression.

Identification of candidate diagnostic methylation sites

The hypermethylated loci were mapped to corresponding genes based on the annotation of Illumina Human Methylation 450 arrays. Intersection of the downregulated genes and the genes that hypermethylated loci were mapped to were selected; then, hypermethylated loci that were not distributed within the promoter region were filtered out. The candidate diagnostic methylation sites were identified via an interaction between the methylation sites above and the hypomethylated sites in normal samples ($\beta < 0.05$).

Assessment of diagnostic methylation sites

The logistic regression model for OSCC diagnosis was established based on the β values of the identified methylation sites in OSCC and paracancerous samples, using TCGA dataset as the training set. The specificity and sensitivity of this model was further evaluated using GEO datasets (GSE87053 and GSE123781) as the validation sets.

Statistical analysis

The survival analysis was conducted using survival package in the R software, and then, Kaplan-Meier curve was

plotted. Fisher's exact test was adopted for calculation of the *p* value.

Results

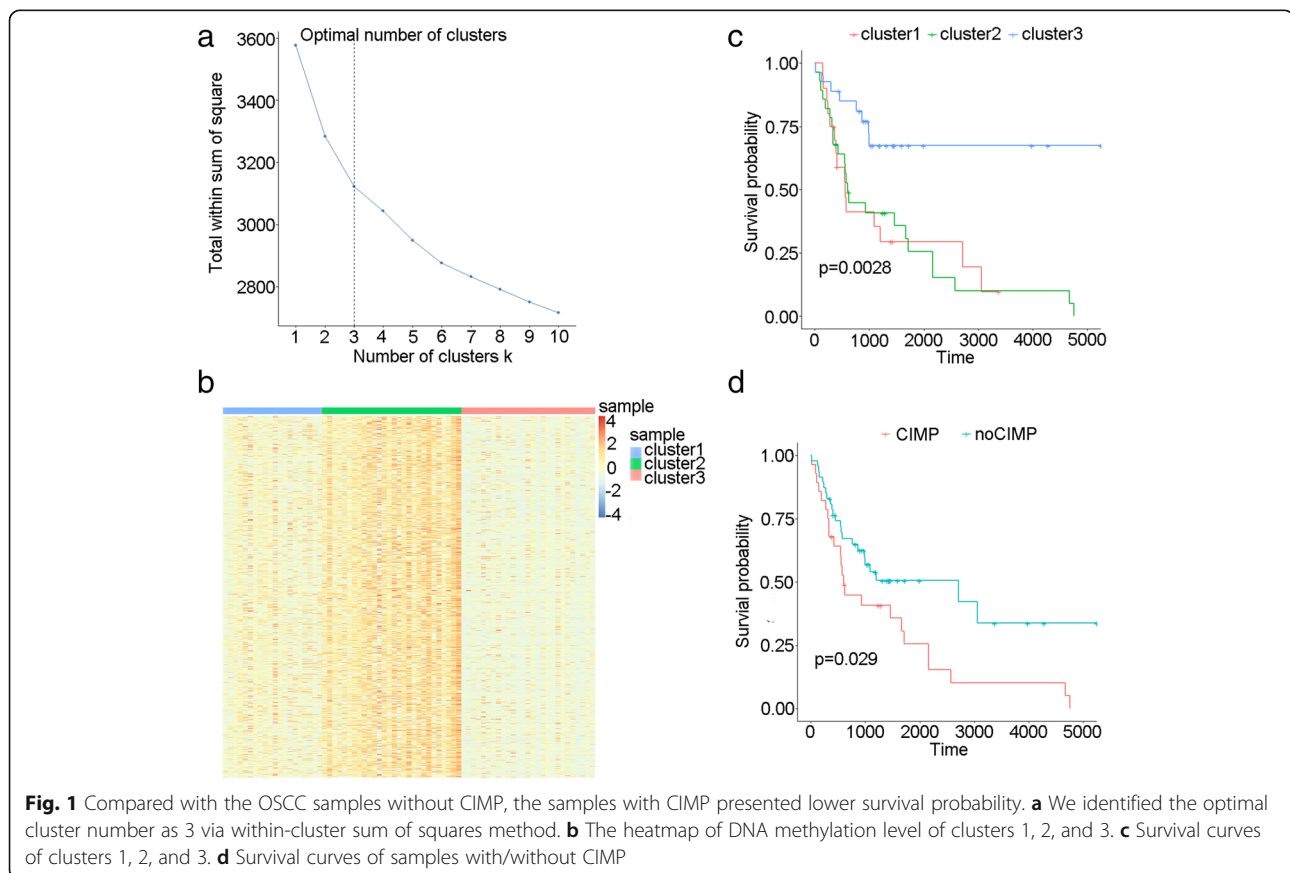### DNA methylation overview and its effect on OSCC prognosis

A total of 1243 methylation sites were selected with the criteria of SD > 0.2 in 70 OSCC patients and $\beta < 0.05$ in 5 paracancerous samples. The optimal cluster number was determined as 3 by using within-cluster sum of squares (wss) method (Fig. 1a). Then, the samples were clustered by using the Euclidean distance metric, and cluster 1, cluster 2, and cluster 3, which consisted of 20, 25, and 25 samples respectively, were identified. The heatmap of DNA methylation level showed that cluster 3 had the highest methylation level among these clusters (Fig. 1b); thus, we inferred that cluster 3 was probably characterized by CIMP.
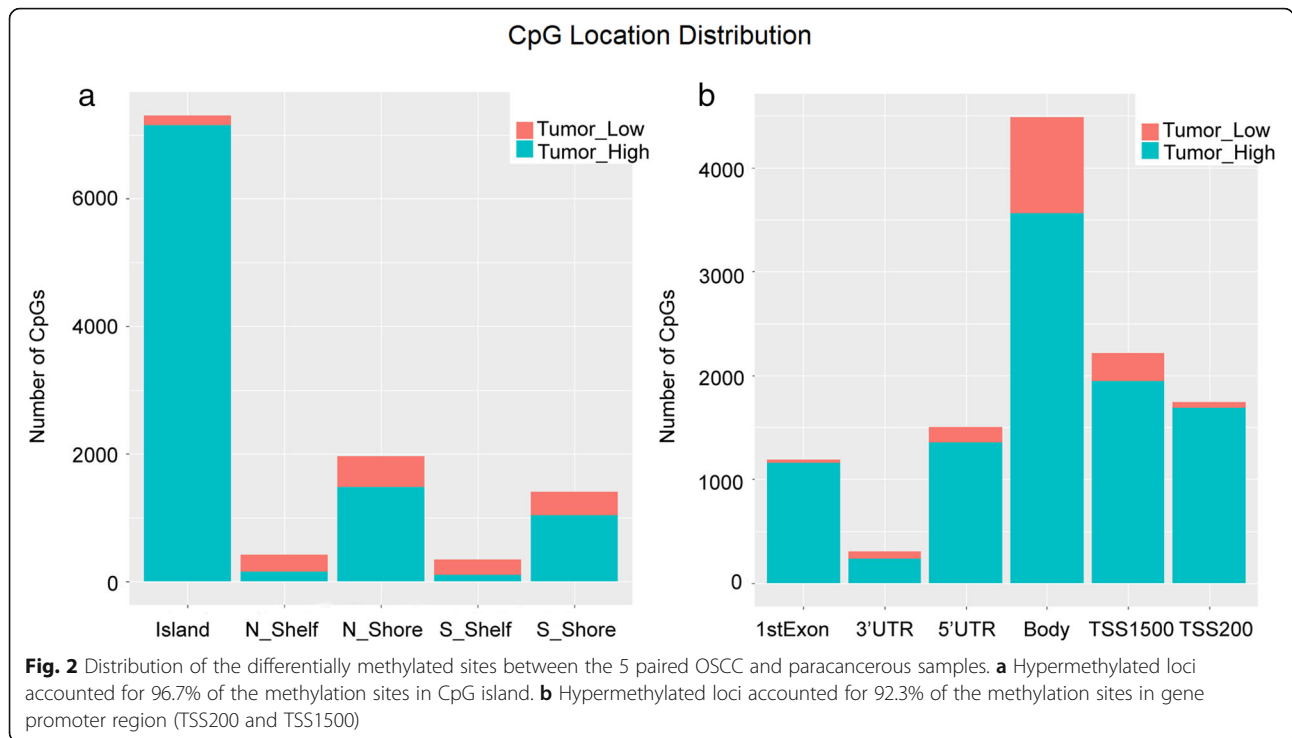
Subsequently, survival analysis of the clusters was performed. As shown in Fig. 1c, cluster 2 had significantly decreased survival probability compared with clusters 1 and 3 ($p = 0.0028$). After merging of clusters 1 and 3, cluster 2 still presented lower survival probability ($p = 0.029$). The result indicated that the OSCC samples with CIMP had lower survival rate than those without CIMP.

The 160 differentially expressed genes were screened in the group with good prognosis as compared with the group with poor prognosis (Figure S1), and 80 genes of them were upregulated and 80 were downregulated, as shown in Table S1.

cg02860732 and cg04342955 were identified as candidate methylation sites for OSCC diagnosis

A total of 23,315 differentially methylated sites between the 5 paired OSCC and paracancerous samples were obtained. As shown in Fig. 2, the hypermethylated loci were mainly distributed in CpG island (96.7%) and gene promoter region (TSS200 and TSS1500, 92.3%). These differentially methylated sites included 8090 hypomethylated loci and 15,225 hypermethylated loci, which corresponded to 2860 and 3253 genes respectively. There were 131 differentially expressed genes between the 5 paired OSCC and paracancerous samples, and all of them showed downregulated expressions (Fig. 3a). It is known that the hypermethylation of tumor suppressor gene, especially for the hypermethylation within the gene promoter region, could lead to the downregulation of the tumor suppressor gene and carcinogenesis [16, 17]. Therefore, interaction of the genes that hypermethylated loci were mapped to and the downregulated genes were

Fig. 1 Compared with the OSCC samples without CIMP, the samples with CIMP presented lower survival probability. **a** We identified the optimal cluster number as 3 via within-cluster sum of squares method. **b** The heatmap of DNA methylation level of clusters 1, 2, and 3. **c** Survival curves of clusters 1, 2, and 3. **d** Survival curves of samples with/without CIMP

Fig. 2 Distribution of the differentially methylated sites between the 5 paired OSCC and paracancerous samples. **a** Hypermethylated loci accounted for 96.7% of the methylation sites in CpG island. **b** Hypermethylated loci accounted for 92.3% of the methylation sites in gene promoter region (TSS200 and TSS1500)
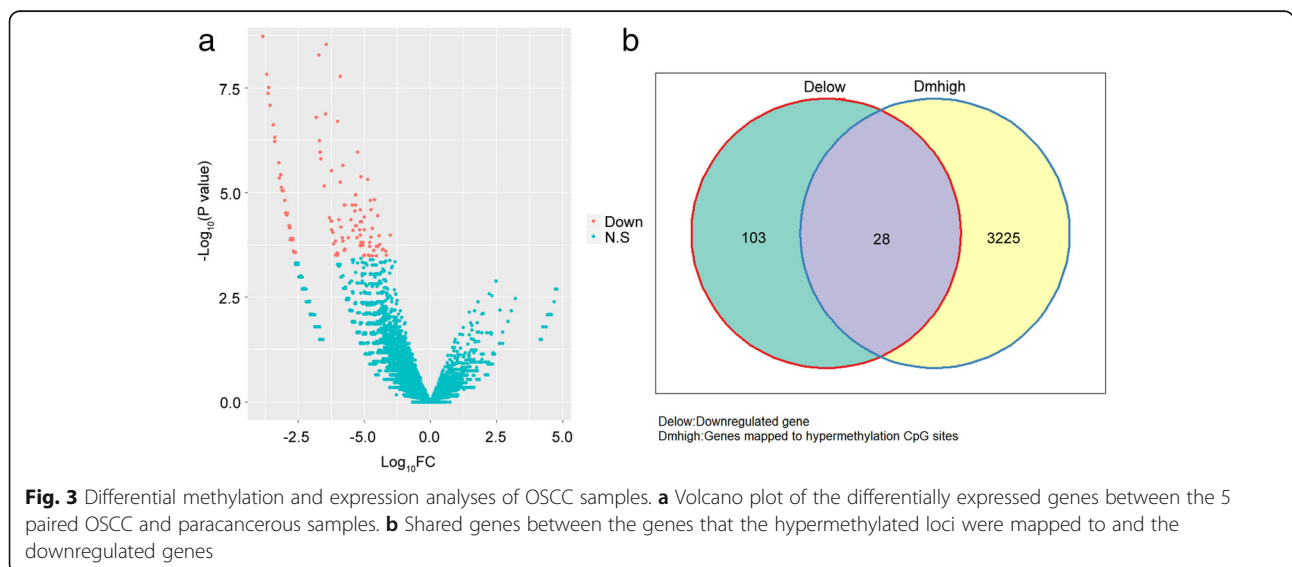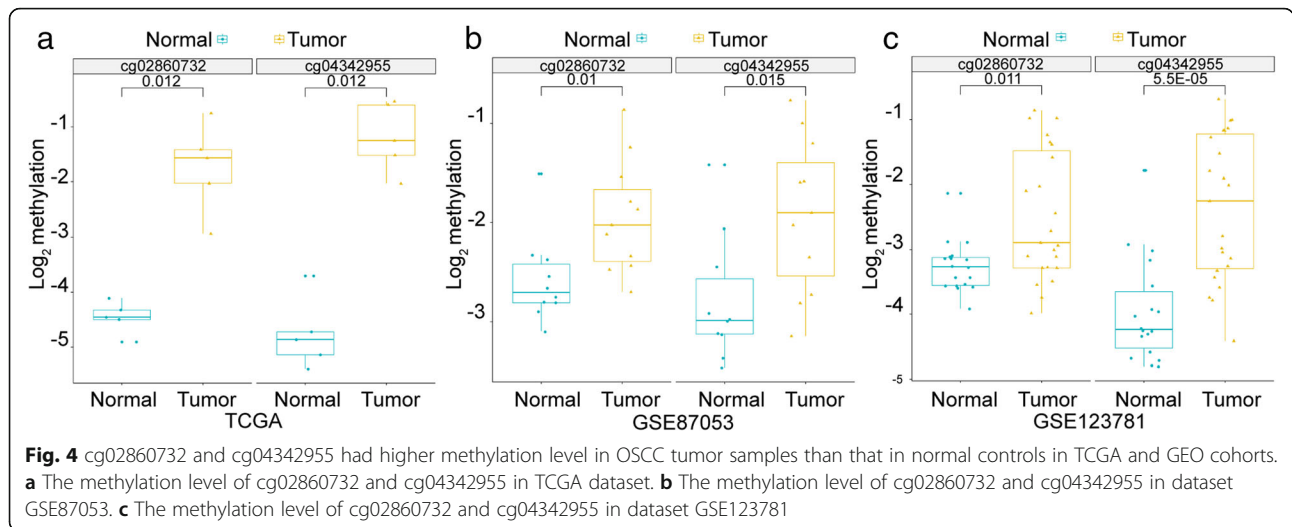
obtained. As shown in Fig. 3b, there were 28 common genes between them, which contained 134 hypermethylated loci. After removal of 54 hypermethylated loci that were not located in the promoter region, an interaction of the remaining hypermethylated loci and the hypomethylated loci in normal samples ($\beta < 0.05$) was obtained. cg02860732 and cg04342955 were preliminarily screened as candidate diagnostic methylation sites for OSCC.

Establishment and evaluation of the diagnostic model of OSCC

To validate the accuracy of cg02860732 and cg04342955, their methylation level in GEO datasets GSE87053 and GSE123781 were investigated. As shown in Fig. 4, both of the two methylation sites presented higher methylation level in OSCC tumor samples than that in normal controls in GEO datasets ($p < 0.05$), which was consistent with the tendency in TCGA dataset.
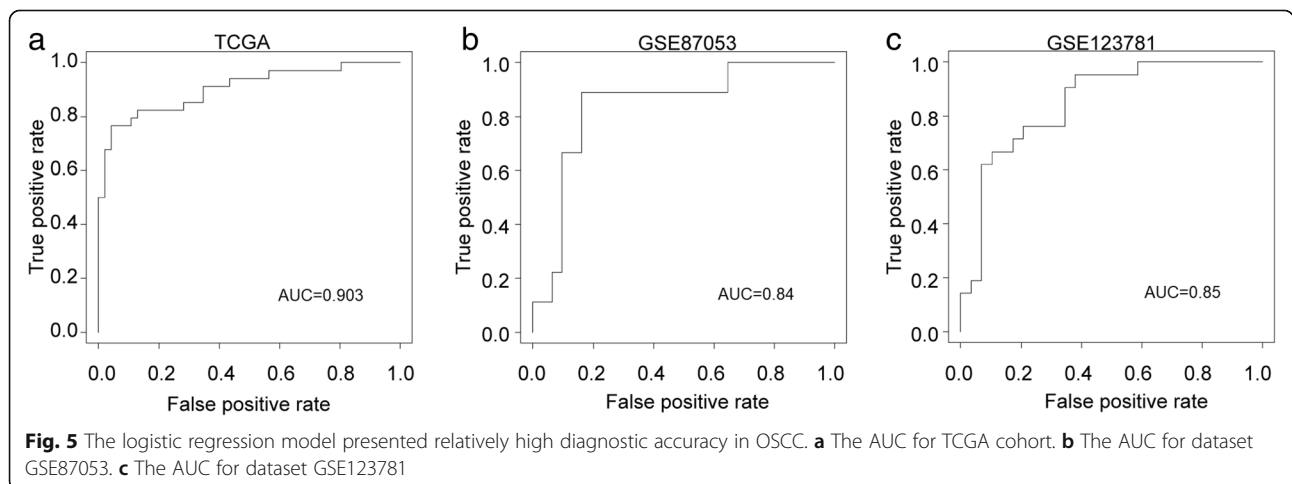

Fig. 3 Differential methylation and expression analyses of OSCC samples. **a** Volcano plot of the differentially expressed genes between the 5 paired OSCC and paracancerous samples. **b** Shared genes between the genes that the hypermethylated loci were mapped to and the downregulated genes

Fig. 4 cg02860732 and cg04342955 had higher methylation level in OSCC tumor samples than that in normal controls in TCGA and GEO cohorts. **a** The methylation level of cg02860732 and cg04342955 in TCGA dataset. **b** The methylation level of cg02860732 and cg04342955 in dataset GSE87053. **c** The methylation level of cg02860732 and cg04342955 in dataset GSE123781

Then, the logistic regression model was established using the TCGA dataset as training set. As shown in Fig. 5a, the area under ROC curve (AUC) was 0.903 for the TCGA cohort. We validated the accuracy of this model in GEO datasets, achieving AUC of 0.84 for dataset GSE87053 and 0.85 for dataset GSE123781 (Fig. 5b and c). These results suggested that the logistic regression model we established with cg02860732 and cg04342955 presented relatively high diagnostic accuracy in OSCC.

Discussion

Abnormal DNA methylation patterns, including hypermethylation of gene promoter region which is accompanied by gene silencing, play a key role in a large number of cancers [18]. Previous research revealed that the hub-methylated sites were related the diagnosis and prognosis for OSCC by weighted gene comethylation network analysis (WGCNA) [19]. To identify samples with CIMP, patients were clustered

by K-means method in our study, and the results suggested that samples with CIMP had lower survival rate. Subsequently, a total of 131 downregulated genes and 23,315 differentially methylated sites consisting of 8090 hypomethylated loci and 15,225 hypermethylated loci between the OSCC tumor and paracancerous samples were identified. As the hypermethylation of tumor suppressor gene, especially hypermethylation within the gene promoter region, could lead to the downregulation of the tumor suppressor gene and carcinogenesis [16], we focused on the intersection of the genes that hypermethylated loci were mapped to and the downregulated genes here. Candidate diagnostic hypomethylated sites that were distributed within the promoter region of the 131 downregulated genes and presented hypomethylation in normal samples ($\beta < 0.05$) were further screened. Finally, methylation sites cg02860732 and cg04342955, locating in gene *SHISA9*, were preliminarily identified as

Fig. 5 The logistic regression model presented relatively high diagnostic accuracy in OSCC. **a** The AUC for TCGA cohort. **b** The AUC for dataset GSE87053. **c** The AUC for dataset GSE123781

diagnostic methylation sites for OSCC. Then, their methylation level in GEO datasets GSE87053 and GSE123781 were further investigated. In consistence with the tendency in TCGA cohort, cg02860732 and cg04342955 showed higher methylation level in OSCC tumor samples than that in normal controls in GEO datasets, validating the accuracy of the two methylation sites in different datasets.

Both of the methylation sites cg02860732 and cg04342955 were located in gene *SHISA9*. *SHISA9*, also known as *CKAMP44*, encodes type-I transmembrane protein which is postsynaptically localized and recognized as a member of Shisa protein family [20]. As an AMPA receptor-related protein, SHISA9 was observed in the complexes of AMPA receptor-related proteins and was able to control the short-term plasticity and currents modulated by AMPA receptor [21, 22]. Moreover, a previous study indicated that *SHISA9* overexpression resulted in more rapid and stronger desensitization of AMPA receptor, as well as weaker recovery from desensitization [23]. While AMPA receptors were reported to be involved in cancer development, such as proliferation, migration, and survival of cancer cells, and the AMPA receptor antagonists were considered as promising anticancer agents via inhibiting cancer cell viability [24]. It was found that AMPA receptor activation could enhance the migration and invasion of cancer cells in pancreatic cancer through Kras-MAPK signaling activation [25]. Besides, AMPA receptors were proved to promote the invasion and development of glioma [26]. Herein, we inferred that *SHISA9* might play an inhibitory role in OSCC by mediating the desensitization of AMPA receptor, which alleviated the activity of AMPA receptor [27]. As a consequence, the downregulated expression of *SHISA9* was probably associated with the OSCC carcinogenesis. However, the specific role of *SHISA9* in OSCC and the underlying mechanism warrant further investigation. To our knowledge, this was the first report of *SHISA9* as a potential tumor suppressor gene in OSCC.

In addition, a logistic regression model for OSCC diagnosis was identified based on cg02860732 and cg04342955, and obtained high AUC values for both of the training set and validation set. These results suggested that this logistic regression model presented relatively high diagnostic accuracy in OSCC. In clinical practice, early screening and accurate diagnosis of OSCC cases without obvious symptoms using conventional approaches remain to be difficult. Several studies have focused on DNA methylation analysis, attempting to explore methylation-related diagnostic biomarkers in OSCC [28, 29]. Compared with these researches that involved multiple genes, only gene *SHISA9* was implicated in the diagnostic model we identified, which was assumed to be more feasible.

In conclusion, our study identified cg02860732 and cg04342955 as potential diagnostic methylation sites for OSCC and a logistic regression model based on them for the diagnosis of OSCC which achieved good performance. These results provided a promising monitoring tool for OSCC diagnosis and might be helpful for timely treatment and therapy guidance of the patients.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12957-021-02129-1.

Additional file 1.

Additional file 2: Table S1. Analysis results of differentially expressed genes.

References
1. Lauritano D, Lucchese A, Contaldo M, Serpico R, Muzio LL, Biolcati F, et al. Oral squamous cell carcinoma: diagnostic markers and prognostic indicators. J Biol Regul Homeost Agents. 2016;30:169–76.
2. Hema KN, Smitha T, Sheethal HS, Mirnalini SA. Epigenetics in oral squamous cell carcinoma. J Oral Maxillofac Pathol. 2017;21:252–9 https://doi.org/10.4103/jomfp.JOMFP_150_17.
3. Ravi KS, Naithani M, Kaur S, Reddy KS, Pasi R. Abnormal nuclear variations in response to radiotherapy- as a tool in treatment planning and assessment of prognosis. J Clin Diagn Res. 2016;10:AC08–12 https://doi.org/10.7860/JCDR/2016/19907.8338.
4. Vucicevic Boras V, Fucic A, Virag M, Gabric D, Blivajs I, Tomasovic-Loncaric C, et al. Significance of stroma in biology of oral squamous cell carcinoma. Tumori. 2018;104:9–14 https://doi.org/10.5301/tj.5000673.
5. Radhika T, Jeddy N, Nithya S, Muthumeenakshi RM. Salivary biomarkers in oral squamous cell carcinoma - an insight. J Oral Biol Craniofac Res. 2016;6: S51–S4 https://doi.org/10.1016/j.jobcr.2016.07.003.

6.   Toyoshima T, Vairaktaris E, Nkenke E, Schlegel KA, Neukam FW, Ries J. Cytokeratin 17 mRNA expression has potential for diagnostic marker of oral squamous cell carcinoma. J Cancer Res Clin Oncol. 2008;134:515–21 https://doi.org/10.1007/s00432-007-0308-8.

7.   Ishiwata J, Kasamatsu A, Sakuma M, Iyoda M, Yamatoji M, Usukura K, et al. State of heat shock factor 1 expression as a putative diagnostic marker for oral squamous cell carcinoma. Int J Oncol. 2012;40:47–52 https://doi.org/10.3892/ijo.2011.1178.

8.   Tang H, Wu Z, Zhang J, Su B. Salivary lncRNA as a potential marker for oral squamous cell carcinoma diagnosis. Mol Med Rep. 2013;7:761–6 https://doi.org/10.3892/mmr.2012.1254.

9.   Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum Mol Genet. 2001;10:687–92 https://doi.org/10.1093/hmg/10.7.687.

10. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010;28:1057–68 https://doi.org/10.1038/nbt.1685.

11. Shen S, Wang G, Shi Q, Zhang R, Zhao Y, Wei Y, et al. Seven-CpG-based prognostic signature coupled with gene expression predicts survival of oral squamous cell carcinoma. Clin Epigenetics. 2017;9:88 https://doi.org/10.1186/s13148-017-0392-9.

12. Foy JP, Pickering CR, Papadimitrakopoulou VA, Jelinek J, Lin SH, William WN Jr, et al. New DNA methylation markers and global DNA hypomethylation are associated with oral cancer development. Cancer Prev Res (Phila). 2015; 8:1027–35 https://doi.org/10.1158/1940-6207.CAPR-14-0179.

13. Langevin SM, Butler RA, Eliot M, Pawlita M, Maccani JZ, McClean MD, et al. Novel DNA methylation targets in oral rinse samples predict survival of patients with oral squamous cell carcinoma. Oral Oncol. 2014;50:1072–80 https://doi.org/10.1016/j.oraloncology.2014.08.015.

14. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell. 2010;17:510–22 https://doi.org/10.1016/j.ccr.2010.03.017.

15. Cancer Genome Atlas Research N, Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, et al. Comprehensive molecular characterization of papillary renal-cell carcinoma. N Engl J Med. 2016;374:135–45 https://doi.org/10.1056/NEJMoa1505917.

16. Cui J, Rohr LR, Swanson G, Speights VO, Maxwell T, Brothman AR. Hypermethylation of the caveolin-1 gene promoter in prostate cancer. Prostate. 2001;46:249–56 https://doi.org/10.1002/1097-0045(20010215)46:3<249::aid-pros1030>3.0.co;2-#.

17. Kirk J, et al. N Engl J Med. 2004;350:947–8 author reply -8. https://doi.org/10.1056/NEJM200402263500920.

18. Morgan AE, Davies TJ, Auley MTM. The role of DNA methylation in ageing and cancer. Proc Nutr Soc. 2018;77:412–22 https://doi.org/10.1017/S0029665118000150.

19. Dai Y, Lv Q, Qi T, Qu J, Ni H, Liao Y, et al. Identification of hub methylated-CpG sites and associated genes in oral squamous cell carcinoma. Cancer Med. 2020;9:3174–87 https://doi.org/10.1002/cam4.2969.

20. Karataeva AR, Klaassen RV, Stroder J, Ruiperez-Alonso M, Hjorth JJ, van Nierop P, et al. C-terminal interactors of the AMPA receptor auxiliary subunit Shisa9. PLoS One. 2014;9:e87360 https://doi.org/10.1371/journal.pone.0087360.

21. Yang B, Dormann C, Vogt MA, Sprengel R, Gass P, Inta D. Facilitated c-Fos induction in mice deficient for the AMPA receptor-associated protein Ckamp44. Cell Mol Neurobiol. 2016;36:1215–8 https://doi.org/10.1007/s10571-015-0307-2.

22. Kunde SA, Rademacher N, Zieger H, Shoichet SA. Protein kinase C regulates AMPA receptor auxiliary protein Shisa9/CKAMP44 through interactions with neuronal scaffold PICK1. FEBS Open Bio. 2017;7:1234–45 https://doi.org/10.1002/2211-5463.12261.

23. von Engelhardt J, Mack V, Sprengel R, Kavenstock N, Li KW, Stern-Bach Y, et al. CKAMP44: a brain-specific protein attenuating short-term synaptic plasticity in the dentate gyrus. Science. 2010;327:1518–22 https://doi.org/10.1126/science.1184178.

24. Ruiz DS, Luksch H, Sifringer M, Temme A, Staufner C, Rzeski W, et al. AMPA Receptor antagonist CFM-2 decreases survivin expression in cancer cells. Anticancer Agents Med Chem. 2018;18:591–6 https://doi.org/10.2174/1871520618666180228123406.

25. Herner A, Sauliunaite D, Michalski CW, Erkan M, De Oliveira T, Abiatari I, et al. Glutamate increases pancreatic cancer cell invasion and migration via AMPA receptor activation and Kras-MAPK signaling. Int J Cancer. 2011;129: 2349–59 https://doi.org/10.1002/ijc.25898.

26. Piao Y, Lu L, de Groot J. AMPA receptors promote perivascular glioma invasion via beta1 integrin-dependent adhesion to the extracellular matrix. Neuro Oncol. 2009;11:260–73 https://doi.org/10.1215/15228517-2008-094.

27. David JC, Yamada KA, Bagwe MR, Goldberg MP. AMPA receptor activation is rapidly toxic to cortical astrocytes when desensitization is blocked. J Neurosci. 1996;16:200–9.

28. Morandi L, Gissi D, Tarsitano A, Asioli S, Gabusi A, Marchetti C, et al. CpG location and methylation level are crucial factors for the early detection of oral squamous cell carcinoma in brushing samples using bisulfite sequencing of a 13-gene panel. Clin Epigenetics. 2017;9:85 https://doi.org/10.1186/s13148-017-0386-7.

29. Arantes LM, de Carvalho AC, Melendez ME, Centrone CC, Gois-Filho JF, Toporcov TN, et al. Validation of methylation markers for diagnosis of oral cavity cancer. Eur J Cancer. 2015;51:632–41 https://doi.org/10.1016/j.ejca.2015.01.060.

Publisher's Note