Plasma cell-free DNA methylome profiling in pre- and post-surgery oral cavity squamous cell carcinoma

Krupal B Patel, MD1,*; Tapan A Padhya, MD2; Jinyong Huang, PhD3; Liang Wang, MD, PhD3; and Xuefeng Wang, PhD4,5*

1Department of Head and Neck-Endocrine Oncology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida, USA.
2Otolaryngology - Head and Neck Surgery, University of South Florida Morsani College of Medicine, Tampa, USA.
3Department of Tumor Biology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida, USA.
4Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33612, USA
5Moffitt Cancer Center Immuno-Oncology Program, Tampa, FL 33612, USA
*To whom correspondence should be addressed.

Corresponding Authors:
Krupal B Patel, MD
Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612
Tel: 813-745-8036
Email: Krupal.Patel@moffitt.org

Xuefeng Wang, PhD
Moffitt Cancer Center, 12902 Magnolia Drive, MRC, Tampa, FL 33612
Tel: 813-745-6710; Fax: 813-745-6107
Email: xuefeng.wang@moffitt.org

Funding support: This work has been supported in part by Moffitt Clinical Science Fund; a National Institute of Health grant R01DE030493; and Moffitt’s Biostatistics and Bioinformatics Shared Resources, Tissue Core, and Genomics Core Facilities at the H. Lee Moffitt Cancer Center & Research Institute, an NCI-designated Comprehensive Cancer Center (P30-CA076292).

Running head: Circulating cfDNA methylation biomarkers in OCSCC

Conflicts of interest disclosure: None declared
ABSTRACT

Purpose: Head and neck squamous cell carcinoma cancer (HNSCC), a highly heterogeneous disease that involves multiple anatomic sites, is a leading cause of cancer-related mortality worldwide. Although the utility of noninvasive biomarkers based on circulating cell-free DNA (cfDNA) methylation profiling has been widely recognized, limited studies have been reported so far regarding the dynamics of cfDNA methylome in oral cavity squamous cell carcinoma (OCSCC). It is hypothesized in this study that comparison of methylation profiles in pre- and post-surgery plasma samples will reveal OCSCC-specific prognostic and diagnostic biomarkers.

Materials and methods: Matched plasma samples from eight patients with OCSCC were collected at Moffitt Cancer Center before and after surgical resection. Plasma-derived cfDNA was analyzed by cfMBD-seq, which is a high-sensitive methylation profiling assay. Differential methylation analysis was then performed based on the matched samples profiled. As a strategy to further prioritize tumor-specific targets, top differential methylated regions (DMRs) were called by reanalyzing methylation data from paired tumor and normal tissue collected in the TCGA head and neck cancer cohort.

Results: In the top 200 HNSCC-specific DMRs detected based on the TCGA dataset, a total of 23 regions reached significance in the plasma-based DMR test. The top five validated DMR regions (ranked by the significance in the plasma study) are located in the promoter regions of genes PENK, NXPH1, ZIK1, TBXT and CDO1, respectively. The genome-wide cfDNA DMR analysis further highlighted candidate biomarkers located in genes SFRP4, SOX1, IRF4 and PCDH17. The prognostic relevance of candidate genes was confirmed by survival analysis using the TCGA data.

Conclusion: This study supports the utility of cfDNA-based methylome profiling as a promising noninvasive biomarker source for OCSCC and HNSCC.
INTRODUCTION

Head and neck squamous cell carcinoma cancer (HNSCC) is a highly heterogeneous disease that involves multiple anatomic sites including oral cavity, larynx, and pharynx. Outcomes of patients with HNSCC have not improved significantly over the past decade, with an overall five-year survival rate around 50%. There is a pressing need in the area to develop reliable prognostic and diagnostic biomarkers to enable better patient management, from early detection of disease to efficient monitoring of cancer recurrence following treatment. Despite their established role in cancer development, epigenetic biomarkers and DNA methylation (DNAme) remain understudied in HNSCC research. Currently, TCGA-HNSC is the only publicly available resource for mining DNAme patterns in head and neck cancer. Cell-free DNA (cfDNA) includes both genetic and epigenetic information and offers several advantages including monitoring tumor burden, and novel discovery of biomarkers for diagnosis and prognosis. cfDNA is thought to potentially incorporate metastatic sites thus addressing tumor heterogeneity. Aberrant DNA methylation changes are thought to occur early during tumorigenesis and enables tumor progression and thus may be a more specific and sensitive approach to identify minimal residual disease and prognosis. While genetic analysis of cfDNA can be challenging due to its low yield and being highly fragmented, plasma cfDNA next generation assays are starting to be utilized in routine clinical use for solid malignancies such as lung and colon cancers to make treatment decisions.

cfDNA has been reported to decrease to background level following surgery. Therefore, we hypothesized that comparing methylation profiles in pre- and post-surgery plasma samples will help validate HNSCC-specific prognostic and diagnostic biomarkers, and provides an opportunity for novel biomarker discovery. Here, we focus on single anatomic site and assess the feasibility of detecting cfDNA methylome in patients with locoregional oral cavity squamous cell carcinomas (OCSCC) and the methylome dynamics in post-operative setting. A high-sensitive cfDNA methylome profiling technique called cfMBD-seq was applied on collected plasma samples. Different from bisulfite conversion-based sequencing methods, cfMBD-seq capture and quantify methylated DNA by methyl-CpG binding protein (MBD). cfMBD-seq is able to generate high-quality sequencing read with ultra-low amount of input DNA (2-10 ng per ml), and has demonstrated better performance in terms of enrichment of CpG islands compared to similar protocols such as cfMeDIP-seq. To facilitate cfDNA methylation biomarker
prioritization with limited sample size, we first conducted a bioinformatics analysis to detect differentially methylated regions (DMRs) based on the matched tumor-normal tissue data collected from the TCGA-HNSC project. We hypothesized that the top cancer-specific DMRs detected in the TCGA-HNSC cohort will also exhibit differential methylation patterns between before- and after-surgery plasma samples. As an alternative strategy, we performed a genome-wide search of top DMRs based on the matched cfDNA methylation profiles. DMR analyses were conducted on different patient subgroups as a sensitivity analysis considering the presence of patient and sample heterogeneity. Once we identified top DMRs, we also examined their prognostic relevance in the TCGA patient data, as well as their performance in discriminating pre- and post-treatment plasma samples.

MATERIALS AND METHODS

Patient sample collection

To detect HNSC-specific cfDNA biomarkers, we studied the plasma samples collected from a cohort of head and neck cancer patients treated at Moffitt Cancer Center (Tampa, USA). In this pilot study, a total of 16 matched plasma samples were collected from 8 patients before and at least 4 weeks after surgery. The basic clinical characteristics of these patients are displayed in Table 1. The study was approved by Institutional Review Board at Moffitt. All patients were consented to the protocol and all samples are de-identified during the methylation profiling process and in the downstream analysis.

TCGA-HNSC analysis

To facilitate cfDNA methylation biomarker discovery, we first conducted a bioinformatics analysis to detect differentially methylated regions (DMRs) based on the tumor tissue methylation data collected from the TCGA-HNSC project. A total of 580 samples were profiled by the Illumina Infinium HumanMethylation450 BeadChip (450K array) in this cohort. Because the goal is to identify cancer-specific regions, our DMR analysis focuses on the 100 paired tumor and normal tissue methylation data collected from 50 patients. We downloaded level 3 DNA methylation data (beta value) from Broad Firehose web portal (gdac.broadinstitute.org) and all clinical data from GDC data portal (portal.gdc.cancer.gov). DMR analysis was performed using the bumphunter function implemented in the R package “minfi”, with the effect size
cutoff set at 0.3 and the resampling number at 1000. We restricted downstream analyses on regions > 5 bp in length, in which all regions contain at least two probes (L≥2). The detected regions were annotated against genome build UCSC hg19 by using the annotateDMRInfo function implemented in the “methylAnalysis” package.

Plasma cfDNA methylome profiling by cfMBD-seq

Cell-free DNA in plasma were profiled by cfMBD-seq, which is an enrichment-based ultra-low input cfDNA methylation profiling method recently developed by our team at Moffitt\textsuperscript{14}. Briefly, Maxwell RSC ccf DNA Plasma kit was used to extract the cfDNA from 1 ml of plasma. If one sample contains less than 5ng DNA, we extract DNA from another 1 ml plasma sample. We then combined the cfDNA from the first and second extraction (if needed for a patient sample) for the methylation enrichment and sequencing library preparation. Methylated DNA fragments were enriched and captured by using a MethylCap Kit. cfMBD libraries were prepared and quantified following the steps as previously described in the cfMBD-seq protocol, and sequenced by Illumina NextSeq 500/550 High Output Kit (75 cycles).

After fastq file merging and adapter trimming steps, sequence reads were aligned to hg19 assembly using BWA MEM (v0.7.10). Mapped reads were further sorted and filtered to remove low-quality and duplicated reads using samtools (v1.9) and picard tools (v1.82). R package “MEDIPS” was used to conduct coverage saturation analysis and downstream QC analysis. The “qsea” R package was used to calculate normalized methylation (beta) levels both at a genome-wide level and in targeted ROI regions (such as promoter regions). We applied fitNBglm function in “qsea” to perform the differential coverage analysis. The function fits a negative-binomial model for each genomic widow similar to the differential expression analysis function implemented in the package “edgeR”. To maximize the statistical power, the design matrix in the DMR analysis was formed in a paired DMR setting, in which an additive model formula is formed to include both treatment effect and subject effect. For statistical testing, a reduced model was fit by the function without the treatment term and the p-value is generated by comparing the likelihood ratio of the models against a Chi-square distribution.\textsuperscript{15,16} To remove potential noise regions with low coverage, we only consider regions (1kb window) with a minimum of 50 reads in all the DMR tests.
Prioritizing candidate cfDNA methylation biomarkers

Because the current study has limited sample size and HNSC samples exhibit high heterogeneity in nature, we propose two schemes to efficiently prioritize the most robust cfDNA biomarker panels while minimizing the false discovery markers or ones with weak clinical significance. As illustrated in the Fig 1A, the first biomarker discovery scheme only investigates DMRs that have been detected based on the analysis using the TCGA data. We hypothesize that the top cancer-specific DMRs detected based on the matched tumor-normal tissues will also exhibit differential methylation patterns between before- and after-surgery plasma samples. The stringent genome-wide multiple-testing correction is not required in this setting because it becomes a targeted biomarker validation analysis. In the second scheme, we perform genome-wide differential coverage analysis on cfDNA methylation data by only considering regions that are located in or nearby the promoter of known genes (defined as 5kb upstream and 2kb downstream of the TSS regions). Furthermore, as will be explained more in the next section, we considered different patient subgroups for the pre- and post-treatment DMR analysis. In each test, regions with an adjusted p-value less than 0.1 or unadjusted p-values less than $1 \times 10^{-6}$ will be reported. Similar to the TCGA analysis, the detected regions were annotated using the “methyAnalysis” package. In summary, we reason that both schemes are useful in identifying promising targets that could be further tested as diagnostic and prognostic markers in managing HNSC patients.

RESULTS

Candidate regions based on TCGA-HSNC analysis

The DMR analysis by comparing the TCGA-HNSC matched tumor and normal methylation profiles identified a total of 1468 significant regions (effect size cutoff of 0.3 and FWER adjusted p value<0.01) (Supplementary Table 1). As shown in Figure 1B, the majority (84.7%) of these top regions are hypermethylated DMRs; and more than half of regions are located in promoter (29.8%) or nearly (0-1kb) downstream regions of TSS (23.5%). When we narrow the list to the top 200 DMRs only, the proportion of
hypermethylated DMRs and DMRs in promoter region further increased to 97% and 52.5%, respectively. The average length of top 200 DMRs is 461 bp. The top five DMRs in the promoter region are located in genes *MARCHF11, ZNF154, ELMO1, ADCYAP1* and *PIEZO2*. A summary of top DMRs and their associated genes is provided in Table 2. Interestingly, we observed that many zinc-finger genes were enriched in the top DMR list, to only list those in the top 100 list: *ZNF154, ZNF582, ZNF135, ZNF136, ZNF577, ZNF781, ZNF529, ZNF132, ZNF85, ZNF583, ZNF471*, and *ZNF665*.

Targeted DMR validation on plasma cfDNA methylomes

In this section, we focus on validating the significant DMRs detected between tumor and normal tissues. We first performed a paired DMR screening test by comparing the matched pre- and post-treatment cfMBD-seq data, using a fixed window size at 1kb. In the top 200 DMRs detected in the TCGA dataset, we found a total of 23 overlapping gene regions reached significance (p<0.05) in the plasma DMR test (Supplementary Table 2), including the two regions in *ZNF154* and *ELMO1* (top five candidate DMRs from the TCGA analysis). Another two regions *ADCYAP1* and *PIEZO2* from the top five candidate DMRs also reached marginal significance (p-value ~0.06). The normalized methylation levels at these top regions across the 16 plasma samples are depicted in Figure 2 (ranked by plasma DMR p-values). The top five validated regions are located in the promoter regions of genes *PENK, NXPH1, ZIK1, TBXT* and *CDO1*. A clear pattern revealed by Figure 2 is that the TCGA-based DMRs showed the best discriminating power between pre- and post-treatment samples for patients P1 and P7, followed by P4 and P8. Overall, patients P1 and P7 showed the most drastic methylation changes between the pre- and post-treatment samples across most targeted DMRs, potentially due to the fact that both patients (both are male) and had T4 tumors. The pattern of methylation is heterogeneous between different patients as evidenced by results in Figure 2. For example, changes in methylation in patient P4 are more pronounced in DMRs in *NXPH1, ZIK1* and *CNTNAP2*, while changes in P2 (also with a T4 tumor) are more pronounced in *HOXD9, TMEM132C* and *SORCS3*. 
Genome-wide cfDNA DMR analysis on patient subgroups

Given the heterogeneity of head and neck cancer at the molecular level and the limited sample size, we reason that it is more powerful to perform DMR analyses based on subsets of patients based on their clinical characteristics. The patient subset information and the resulted top significant or suggestive DMRs (defined as a p-value < 0.1 level for adjusted p-value or at $10^{-7}$ level for unadjusted p-value) are listed in Table 3. It is interesting to note that the DMR test based on the four patients (P1, P4, P7 and P8) that had the most concordant patterns with TCGA data resulted in the highest number of significant DMRs, followed by the two patient subsets containing T4 tumors (P1, P2, P3 and P7). We excluded patient P5 in most tests (except for the tongue-site subgroup) because the genome-wide PCA analysis (Supplementary Figure S1) indicated that the pre-treatment cfDNA methylation profiles could be a potential outlier, which may explain why there was no significant DMR when all patients are included in the test. The multiple-contrast DMR tests can also be considered as sensitivity analyses that provide further confidence for overlapped findings. It was observed that the genomic regions in gene PENK, SFRP4 and SOX17 were selected in at least three tests, suggesting they might be further prioritized for biomarker validation. Figure 3A shows the detailed methylation levels in top regions that were identified based on the TCGA-concordant subgroup, including two top-ranked validated DMRs listed in Figure 2 (PENK and ZIK1). Similar to the pattern observed in Figure 2, patients P1 and P7 showed the most drastic changes between the pre- and post-treatment samples. Through assessing the gene expression levels of these top genes in TCGA-HNSC data, we observed that four genes (ZIK1, IRF4, PCDH17 and PENK) demonstrated significant or suggestive association with patient overall survival, as shown in Figure 3B-E. Collectively, these findings suggest that these four genes may have tumor suppressor functions and are often hypermethylated in HNSC tumor samples or pre-treatment plasma samples.
Clustering analysis of targeted plasma cfDNA methylation regions

Finally, we tested the performance of top DMRs (as well as the model saturation in terms of the number of biomarkers included) in discriminating pre- and post-treatment plasma samples. The two heatmaps in Figure 4 illustrate the unsupervised clustering results generated based on top 30 DMRs and top 200 DMRs (from the DMR test using all samples but P5), respectively. It shows that the top 30 regions (most of them are hypermethylated in pre-treatment samples) are already sufficient to separate pre- and post-treatment plasma samples except for the P5 pre-treatment sample. This was expected, because the global PCA analysis also indicated that this sample could be a potential outlier. But when top 200 regions were included, this sample, together with all other samples, can be correctly separated.

DISCUSSION

While the current study does not have sufficient sample size to comment on prognostic value of cfDNA methylation, in this study, we successfully demonstrate the feasibility of isolating cfDNA from plasma and a two-pronged approach in identifying top candidate biomarkers by first identifying DMRs from the TCGA dataset and then validating them in our cfDNA samples.

A biomarker for minimal residual disease is advantageous, especially in the setting of oral cavity squamous cell carcinoma patients, a population where 5-year survival rates are estimated between 40-60% for patients with advanced-stage disease, and the majority of the recurrence occurs in the first 2 years. Moreover, if the biomarkers can predict tumor immune response it would be even more preferable. DNAme has the potential to incorporate both of the above features in addition to it being highly tissue specific, and more responsive to genetic variations and environmental exposure. HNSCC is a heterogenous disease and thus focusing on cfDNA DNAme may be advantageous to increase specificity. Previous studies have focused on targeted panels of methylation in either serum/plasma of HNSCC patients. Recently, a study by Burgener et al. did demonstrate tumor-naïve detection of ctDNA by simultaneously profiling mutations and methylation.
In our study, we identified hypermethylation of zinc finger genes, such as ZNF154, in the TCGA tumor samples compared to normal tissue. This is consistent with previous studies where their expression has been shown to correlate with tumor suppressor activity.\textsuperscript{33} ZNF154 has been reported as diagnostic marker in liquid biopsies in multiple cancers.\textsuperscript{34}

Our study also suggests that top 30 DMRs are sufficient to differentiate between pre-treatment and post-treatment samples suggesting that a signature based on these 30 DMRs maybe sufficient to determine minimal residual disease. Many genes in the top DMR list have also been suggested as liquid biopsy methylation biomarkers in other cancer types, such as ZNF154 for multiple cancers,\textsuperscript{34} ELMO1 for gastric cancer\textsuperscript{35}, suggesting that they are reliable cancer-relevant epigenetic biomarkers. The promoter methylation level of IRF4 and PCDH17 have been previously reported as potential liquid biopsy biomarkers for colorectal cancer\textsuperscript{36,37} and bladder cancer.\textsuperscript{38}

The top five validated regions were located in the promoter regions of genes PENK, NXPH1, ZIK1, TBXT and CDO1. Through our analysis 4 candidate genes were identified that may have prognostic value in addition to their role in determining minimal residual disease – ZIK1, IRF4, PCDH17 and PENK. ZIK1 (ZNF762) is part of the Zinc Finger protein group with a KRAB-A domain and found on chromosome 19.\textsuperscript{39} KRAB box-A is a transcription repressor module\textsuperscript{39} and it is plausible that ZIK1 is epigenetically regulated tumor suppressor gene.\textsuperscript{33,40,41} Interferon regulatory factor 4 (IRF4) is a member of the Interferon family and is specifically expressed in lymphocytes\textsuperscript{42} regulating immune responses, immune cell proliferation and differentiation.\textsuperscript{43} While its role in hematologic malignancies has been described previously,\textsuperscript{44} IRF4 expression in lung adenocarcinoma has been associated with favorable prognosis.\textsuperscript{45} Protocadherin 17 (PCDH17) is part of the cadherin superfamily responsible for cell adhesion and possible tumor growth, migration and invasion.\textsuperscript{46,47} PCDH17 methylation has been noted in urological cancers including esophageal,\textsuperscript{48} gastric,\textsuperscript{49} colon,\textsuperscript{49} and bladder cancers.\textsuperscript{50} Proenkephalin (PENK) is expressed in nervous and neuroendocrine systems as part of the opioid pathway\textsuperscript{51}, but is also involved in cell cycle regulation and implicated in head and neck,\textsuperscript{52} gastric,\textsuperscript{53} colon,\textsuperscript{54} breast,\textsuperscript{55} pancreatic,\textsuperscript{56} osteosarcoma,\textsuperscript{57} and bladder\textsuperscript{58} cancers. NXPH1 is primarily expressed in nervous system and is a secreted glycoprotein that forms complexes with alpha neurexins – a group of protein that promote adhesion between dendrites and
axons. In breast cancer samples, NXPH1 methylation levels were lower compared to normal tissues and was more likely to be methylated in low-grade dysplasia than in high-grade dysplasia. In prostate cancers with Gleason score ≥ 7, NXPH1 expression level was upregulated and was incorporated in a 10 gene signature that predicted biochemical recurrence. However, a negative correlation was noted in patients with pancreatic cancer with regards to lymph node metastasis. NXPH1 methylation has also been implicated in neuroblastoma and was incorporated in a 5 gene prognostic signature where it was down regulated suggestive of playing a tumor suppressive role. This suggests that tissue specific changes maybe at play. T-box transcription factor T (TBXT) expression is implicated in mesodermal specification during vertebrate development and is epigenetically silenced in human fetus development at 12 weeks. TBXT expression has been reported in a number of solid malignancies including head and neck, lung, breast, colon, prostate and chordoma – with hypothesis that it promotes epithelial-mesenchymal transition and targeting it may help in cancer control. Promoter methylation CDO1 has also been previously identified as diagnostic biomarkers in lung cancer.

Limitations of this study include limited samples size to draw definitive prognostic conclusion. cfDNA has been correlated with overall stage and subsite. Our study included primarily advanced stage disease and mainly oral cavity squamous cell carcinomas. It is well established that advanced stage cancers and different subsites will have different methylation pattern and lymphatic drainage, thus it is plausible that similar results may not be evident in lower stage disease. In summary, we identified multiple candidate DMRs that allowed distinction between pre-treatment and post-treatment samples suggesting its utility for minimal residual disease and potential as prognostic biomarker.
REFERENCES


8. Vrba L, Futscher BW: A suite of DNA methylation markers that can detect most common human cancers. Epigenetics 13:61-72, 2018


FIGURE LEGENDS

Figure 1. Overview of the proposed DMR analysis on OCSCC plasma samples.
A. The experimental design and overall analytical workflow for cfDNA methylation profiling on pre- and post-treatment OCSCC patient samples.
B. Pie charts showing the distribution of methylation status and genomic locations in the top detected DMRs.

Figure 2. The normalized methylation levels of top DMRs across the matched plasma samples from 8 patients.

Figure 3. Prioritizing cfDNA DMRs based on the four-patient subgroup.
A. Methylation levels in top regions that were identified based on the TCGA-concordant patient subgroup (P1, P4, P7 and P8)
B. (C,D,E) Kaplan-Meier plots validating the prognostic significance of four genes that include detected DMRs (based on the gene expression and survival data from the TCGA-HNSC data).

Figure 4. The unsupervised clustering results generated based on top 30 DMRs (A) and top 200 DMRs (B) showing the separation of pre- and post-treatment samples.
Table 1. Baseline clinical characteristics and treatment of patients included in the plasma cfMBD-seq analysis.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sex</th>
<th>Age range</th>
<th>Site</th>
<th>pT</th>
<th>pN</th>
<th>M</th>
<th>Tumor size (cm)</th>
<th>DOI (mm)</th>
<th>PNI</th>
<th>LVI</th>
<th>LN deposit (cm)</th>
<th>ENE</th>
<th>Previous treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Male</td>
<td>56-60</td>
<td>Mandible</td>
<td>4a</td>
<td>3b</td>
<td>0</td>
<td>5.3</td>
<td>30</td>
<td>Y</td>
<td>N</td>
<td>2.0</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>P2</td>
<td>Male</td>
<td>51-55</td>
<td>Mandible</td>
<td>4a</td>
<td>3b</td>
<td>0</td>
<td>4</td>
<td>28</td>
<td>N</td>
<td>N</td>
<td>2.5</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>P3</td>
<td>Male</td>
<td>56-60</td>
<td>Tongue</td>
<td>4a</td>
<td>3b</td>
<td>0</td>
<td>6.5</td>
<td>35</td>
<td>Y</td>
<td>Y</td>
<td>1.9</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>P4</td>
<td>Female</td>
<td>65-70</td>
<td>Maxilla</td>
<td>1</td>
<td>X</td>
<td>0</td>
<td>0.3</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>N</td>
</tr>
<tr>
<td>P5</td>
<td>Male</td>
<td>76-80</td>
<td>Tongue</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3.6</td>
<td>14</td>
<td>Y</td>
<td>N</td>
<td>0.9</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>P6</td>
<td>Female</td>
<td>41-45</td>
<td>Tongue</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.2</td>
<td>7</td>
<td>Y</td>
<td>Y</td>
<td>N/A</td>
<td>N/A</td>
<td>N</td>
</tr>
<tr>
<td>P7</td>
<td>Male</td>
<td>76-80</td>
<td>Tongue</td>
<td>4a</td>
<td>2a</td>
<td>0</td>
<td>5.2</td>
<td>41</td>
<td>Y</td>
<td>Y</td>
<td>1.7</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>P8</td>
<td>Female</td>
<td>76-80</td>
<td>Buccal Mucosa</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>0.6</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

**pN – pathologic nodal classification; pT – pathologic tumor classification; M – Metastasis; PNI – Perineural invasion; LVI – Lymphovascular invasion; LN – Lymph Node; ENE – Extranodal extension**
Table 2. Table of top genes containing significant DMRs based on TCGA-HNSC matched tumor and normal tissue methylation profiles.

<table>
<thead>
<tr>
<th>Rank methods</th>
<th>Top genes*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes contain top 100 DMRs (promoter regions)</td>
<td>MARCHF11, ZNF154, ELMO1, ADCYAP1, PIEZO2, CCDC181, KCNA3, OTX2-AS1, PCDHC4, DPP10-AS1, MIR129-2, ZNF582, WT1, DDR5, CBLN4, DPP6, SORCS3, ZNF135, GALR1, BOLL, ZSCAN18, BARHL2, ZNF577, CLVS2, ABCC9, INSC, GRM6, ZNF781, ZNF529, TBX5, IGA8, GCSAML, CDO1, ZNF132, IRF8, NXX2-6, ZC4H2, ZSCAN1, NKKPL, NPY, PENK, ZNF85, ADAMTS16, EVX2, NETO1, ZNF583, VAX1, HOXD9, KLHL34, CFTR, PCSK1, ZNF471, ZIK1, SHISA3, SIX6, ZNF665</td>
</tr>
<tr>
<td>Genes contain top 100 DMRs (1kb downstream of TSS)</td>
<td>HOXD9, TRH, GRIA4, ZNF542P, PAX6, GSC, NTN, PPFIA3, PABPC5, SFTA3, NID2, HOXA11-AS, ASCL1, MAGI2-AS3, LOC100289656</td>
</tr>
<tr>
<td>Genes contain top 100 DMRs (other regions)</td>
<td>EN1, SOX17, LHX1, PANTR1, NXX1, ZNF136, LINC00461, PAX2, SEPTIN9, VAX1, ZIC4, HOTAIR, UBD, GBX2, ZNF559-ZNF177, NXX2-2, SOX1, SOX11, GSC, LINC01304, LINC01623, PITX2, PCDHA13, RMCX5-GPRASP2</td>
</tr>
<tr>
<td>Top genes ranked by number of DMRs (&gt;3 DMRs; No. of DMRs indicated in parentheses; Genes containing top 100 DMRs are highlighted in bold)</td>
<td>ZIC4 (11), ZIC1 (9), EN1(8), SOX1(8), TBX5-AS1(8), SOX17(7), BARHL2(6), ZBED9(6), EVX2(5), HOXB3(5), HOXD10(5), LINC00461(5), OTX2(5), PAX2(5), SOX3(5), VAX1(5), HOTAIR(4), HOXA3(4), HOXC4(4), HOXC6(4), LHX8(4), MIR124-2HG(4), MKI67(4), NXPH1(4), ONECUT2(4), PAX6(6), PAX6-AS1(4), PDH1(4), PITX2(4), PTF1A(4), PTPRN2(4), TBX15(4), TLX1(4), TLX3(4), UBD(4)</td>
</tr>
</tbody>
</table>

*Full list of DMRs and their locations are listed in Supplementary Table 1.
Table 3. Table of top cfDNA DMRs identified based on comparing pre- and post-treatment cfDNA methylation profiles by considering different patient subgroups.

<table>
<thead>
<tr>
<th>Patient group/subgroup</th>
<th>Subset rationale</th>
<th>Top cfDNA DMRs (hg19) T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired samples from P1, P4, P7, and P8</td>
<td>Patterns most consistent with TCGA results</td>
<td>chr7:37956001-37957000 (SFRP4) &lt;br&gt; chr8:67873001-67874000 (TF724) &lt;br&gt; chr2:45171001-45172000 (SIX3*) &lt;br&gt; chr11:32458001-32459000 (WT1-AS) &lt;br&gt; chr8:53478001-53479000 (ALKAL1) &lt;br&gt; chr8:57358001-57359000 (PENK) &lt;br&gt; chr13:93879001-93880000 (GPC6) &lt;br&gt; chr8:55370001-55371000 (SOX17) &lt;br&gt; chr4:183062001-183063000 (TENM3-AS1*) &lt;br&gt; chr1:119535001-119536000 (TBX15*) &lt;br&gt; chr19:58095001-58096000 (ZIK1)</td>
</tr>
<tr>
<td>Paired samples from P1, P2, P3, and P7</td>
<td>Patients with T4 tumors (male patients)</td>
<td>chr7:37956001-37957000 (SFRP4) &lt;br&gt; chr1:47696001-47697000 (TAL1) &lt;br&gt; chr5:1594001-1595000 (SDHAP3) &lt;br&gt; chr7:87257001-87258000 (ABC81/RUNDC3B) &lt;br&gt; chr8:57358001-57359000 (PENK)</td>
</tr>
<tr>
<td>Paired samples from P3, P5, P6, and P7</td>
<td>Tongue site only</td>
<td>chr11:43602001-43603000 (MIR129-2)</td>
</tr>
<tr>
<td>P4, P6, and P8</td>
<td>Female patients only</td>
<td>No region significant at 10^-7 level</td>
</tr>
<tr>
<td>Paired samples from P1, P2, P3, P4 and P7</td>
<td>All T4 patients+ P4 (signature patient)</td>
<td>chr7:37956001-37957000 (SFRP4) &lt;br&gt; chr8:57358001-57359000 (PENK) &lt;br&gt; chr8:55370001-55371000 (SOX17) &lt;br&gt; chr3:129693001-129694000 (TRH) &lt;br&gt; chr14:48143001-48144000 (MDGA2*) &lt;br&gt; chr14:48145001-48146000 (MDGA2)</td>
</tr>
<tr>
<td>Paired samples from P1, P2, P3, P4, P6, P7, and P8</td>
<td>All patients except P5 (potential outlier)</td>
<td>chr8:57358001-57359000 (PENK) &lt;br&gt; chr7:37956001-37957000 (SFRP4) &lt;br&gt; chr8:55370001-55371000 (SOX17)</td>
</tr>
</tbody>
</table>

†Unless otherwise specific, DMRs are significant at 0.1 level for adjusted p-value or at 10^-7 level for unadjusted p-value. DMRs not in the promoter region of the gene are indicated by **. 
A

TCGA-HNSC cohort

Matched Tumor and normal tissue methylation (100 samples, 50 pairs)

Significant regions based on DMR analysis (bumphunter)

Validation based on plasma cfDNA: pre- and post-surgery treatment (16 samples, 8 pairs)

Differential coverage analyses of plasma cfDNA methylome (multiple contrasts)

B

All Top DMRs

Hypomethylation

>1kb downstream

>1.5kb upstream of TSS

Promoter

Hypermethylation

0–1kb downstream of TSS

Top 200 DMRs

Hypomethylation

>1kb downstream

>1.5kb upstream of TSS

Promoter

Hypermethylation

0–1kb downstream of TSS
Fig. 2

All rights reserved. No reuse allowed without permission.
chr8:67873001-67874000 (TCF24)
chr7:37956001-37957000 (SFRP4)
chr2:45171001-45172000 (SIX3)
chr11:32458001-32459000 (WT1-AS)
chr8:57358001-57359000 (PENK*)
chr8:53478001-53479000 (ALKAL1)
chr8:55370001-55371000 (SOX17)
chr13:58203001-58204000 (PCDH17*)
chr6:392001-393000 (IRF4*)
chr4:81952001-81953000 (BMP3)
chr13:93879001-93880000 (GPC6)
chr19:58095001-58096000 (ZIK1*)

Fig. 3