**Overview and Gap**

The SEER Cancer Reports estimated oral cancer (OC) incidence in the US at 11.3%/100,000 (n=53,000 new cases) in 2019, representing 3% of all new cancers. Approximately 10,860 deaths are directly attributable to OC. Because the oral cavity is highly innervated, vascularized and associated with rich lymphatics facilitating metastasis, late-stage presentation of OC requires aggressive treatment and is associated with high mortality. While preventable and treatable if diagnosed early, OC survival rate decreases with advancing stage, with stage I and II survival rate >80%, decreasing to 60% for stage III and <40% for stage IV with systemic metastasis2. The estimated cost associated with OC in the US in 2017 exceeded $4 billion4. Highest oral cancer risk has been associated with behavioral factors including smoking, alcohol consumption, and positive status for human papilloma virus (HPV) for certain oral cancers, including oropharyngeal and tonsillar tumors.

While estimated overall 5 year survival (2009-2015) has been increasing, achieving 65.3% in 20151, trends between 1999-2015 documented increased incidence of certain HPV 16-associated OCs (2.7% in men and 0.8% in women)2. In 2015, nearly 19,000 HPV-associated OCs were diagnosed, establishing it as the most frequently-diagnosed HPV-associated cancer2. Moreover, between 1999 and 2016, a notable increase in tongue cancer (2.5 to 3.8 per 100,000) was observed with >17,000 new cases predicted in 20191. An estimated 70% of oropharyngeal cancers are linked to HPV 16, a sexually transmitted virus, while HPV 16+ status is associated with 3% of oral cavity cancers. National SEER data documented a rise in both number (burden) and rates (risk) of 28% and 5%, respectively for OCs between 2007 and 20163. Whereas HPV immunization is predicted to reduce the rates of HPV-associated OC over time, current rates of compliance with full vaccination among teens is estimated at only 51%4. The upward trends in OCs are concerning to healthcare professionals, as no biomarkers to detect OC emergence are currently available.

To address this gap, the current study proposes to pilot whether cell free DNA (cfDNA) arising from OC is detectable in saliva. Christiano et al (2019) recently described a cell free DNA (cfDNA) fragmentation approach involving creation of libraries that are subjected to genome-wide (GW) analysis, termed ‘DNA evaluation of fragments for early interception’ (DELFI)5. The investigators found that tumor-derived cfDNA has higher variability in fragment length compared to healthy cells (p<0.001 for variance ratio test). While conducting GW profiling on cfDNA libraries of patients with various cancers, they observed GW regional differences across various cancer types and were able to both detect cfDNA of tumors and profile tissue of origin. The DELFI approach achieved an overall AUC of 0.94 for detection of patients with cancer. Moreover, combining DELFI with mutational profiling resulted in detection of mutations in 66% of patients with cancer. Since OC was not profiled by Christiano et al5, we propose to address this gap. Further, epigenetic analysis applying DELFI methylation (meDELFI) analysis is proposed to examine profiles noted in OC across advancing cancer stage.

**Specific Aims and Measurable Objectives**

***Hypothesis:*** The current pilot study proposes to test the following hypotheses:

1. cfDNA of OCs are detectable in saliva applying cfDNA fragmentation analysis; and

b) GW DELFI and meDELFI will reveal distinct genetic and epigenetic profiles for OCs, mutations, and epigenetic changes associated with advancing tumor stage.

***Specific aims*** to test these hypotheses include:

**Specific Aim 1:** Pilot DNA fragmentation and WG-DELFI/mutational analysis of cfDNA libraries in prospectively collected saliva to compare profiles of healthy individuals and those of patients with incident OCs.

**Specific Aim 2:** Pilot array-based meDELFI analysis of cfDNA in prospectively collected saliva of patients with OCs and healthy cancer-free subjects

**Secondary Aim:** Create a biobank for additional collection of blood and storage of residual saliva specimens from participants enrolled for specific aim 1&2 for a future study to scale cfDNA profiling in saliva applying cfDNA fragmentation analysis, DELFI/mutational analysis and meDELFI established in this pilot study to validate findings.

**Significance** Currently, reliable biomarkers for translation into clinical practice that could advance precision medicine (PM) approaches for detecting OC are lacking despite advances in liquid biopsy surrounding OCs and other head and neck cancers6. Ideally, non-invasive tests to evaluate cfDNA in high-risk individuals for detection of early-stage disease would enhance potential for optimal clinical outcomes. Mutational analysis of these heterogeneous cancers would further expand insight into genetic contribution to OC etiology, including definition of genetic regions contributing to pathogenesis. Future studies may explore detailed mapping to define candidate loci and potential molecular targets. Tracking epigenetic changes by meDELFI with advancing tumor stage is expected to align with specific genetic regions potentially contributing to causation that may be more densely mapped in the future.

**Clinical Impact**

Aside from microscopic examination of biopsies to determine diagnosis and staging, no screening tests are currently available to identify risk OC. Availability of a non-invasive screening test with potential to monitor for malignant transformation or define mutational profiles longitudinally across the temporal window of disease progression offers potential for establishing patterns associated with cancer stage. Moreover, screening and early detection of these cancers would support more timely interventions at stages when best survival rates and clinical outcomes are achievable. Further mutational analysis of OCs, a heterogeneous group of cancers, may benefit patients whose genetic profile includes molecular targets where therapeutic interventions are currently available or may identify novel therapeutic targets that inform development of new interventions. For example Hu et al (2015) observed that HPV16 activates the epidermal growth factor receptor (EGFR) to regulate β catenin translocation7  . This observation identified HPV 16+ OCs as candidates for therapy with cetuximab, a monoclonal antibody which specificity for EGFR. Notably, in a clinical trial comparing outcomes of radiotherapy in patients randomized to treatment with and without cetuximab for patients with HPV16 positive cancers, overall and five-year survival nearly doubled for patients randomized to cetuximab compared to the arm receiving radiotherapy-only 8.

**Innovation**

While potential value of the approach in cancer screening was only recently described by Christiano et al, (2019)5, application and performance of DELFI/mutational analysis and meDELFI of cfDNA in saliva of patients with OC remains to be explored. We posit that because these cancers occur in the oral cavity, cfDNA concentrations in saliva may be more abundant than in circulation, especially during early-stage of disease when cfDNA is less easily detected. This pilot study proposes to establish concurrent DELFI/mutational analysis and meDELFI testing thereby enhancing its utility in characterizing malignant transformation across the stages of OC. Moreover, because good profile resolution was already observed at 1-2X genome coverage by Christiano et al in profiling other cancers5, the need to only conduct a single WG analysis promotes inherent utility as a cost-effective screening approach with potential applicability in the clinical setting.

**Approach: Study Design:** A prospective design will be applied for the pilot study to collect saliva and blood from subjects with newly diagnosed OC and normal individuals. Patients will be enrolled at both MCHS and UW Madison. The flow diagram in below provides a high-level overview of the study design.

**Preliminary data and sample size estimates:**

Preliminary data to support enrollment of a convenience sample: The present study proposes to enroll up to 40 patients as a convenience sample to test feasibility of establishing the assay and obtain preliminary profiles of OCs originating at various sites within the greater oral cavity. This sample size mirrors that of the Christiano et al (n=38) first used in establishing the assay. Since the study proposes prospective enrollment, patients consenting to enroll who meet enrollment criteria will be enrolled sequentially at time of clinical diagnosis. Table 1 summarizes the number of cases based on ICD10 code (applying rule of two diagnostic codes) abstracted from the Marshfield Clinic Health Systems (MCHS) data warehouse diagnosed annually at MCHS across the past 4 years. Enrollment of up to 10 cases and matched controls each across sites is proposed. Incidence rates support capacity to achieve the proposed sample size. As a pilot study whose main objective proposes assay development and descriptive analyses of findings, sample size estimates supporting detection of statistically significant differences between cases and controls are not projected.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Oral Cavity or Head/Neck Cancer Case Counts 01/01/2016-11/12/2019** |  | | | | | |  |  |  |  |
| Table 1 | 2016 | 2017 | 2018 | 2019 |  |
| DX\_CATEGORY: malignant neoplasms of | COUNT | COUNT | COUNT | COUNT | ALL |
| floor of mouth | 11 | 4 | 2 | 2 | 19 |
| gum | 9 | 3 | 2 | 1 | 15 |
| hypopharynx | 11 | 1 | 1 | 3 | 16 |
| lip |  | 5 | 2 | 5 | 12 |
| nasopharynx | 15 | 2 | 3 | 1 | 21 |
| oropharynx | 27 | 7 | 4 | 7 | 45 |
| other and ill-defined sites in the lip, oral cavity and pharynx | 3 | 3 | 1 | 0 | 7 |
| other and unspecified major salivary glands | 10 | 4 | 1 | 4 | 19 |
| other and unspecified parts of mouth | 39 | 9 | 12 | 6 | 66 |
| other and unspecified parts of tongue | 71 | 9 | 18 | 14 | 112 |
| palate | 9 | 3 | 1 | 4 | 17 |
| tonsil | 72 | 13 | 10 | 10 | 105 |
| ALL | 277 | 63 | 57 | 57 | 454 |

**Methods**

**Patient identification and enrollment process:**

***Cases (n=10/institution)***: The study will undergo central IRB review with MCHS as the IRB of record. After approval, MCHS and UW Carbone Cancer Center patients receiving a new diagnosis of OC will be approached for interest in study enrollment if they meet the following eligibility criteria:

* no prior history of cancer
* no prior or current exposure to radiotherapy, chemotherapeutic approaches, or cancer therapy that could impact on quality of the DNA or epigenetic profiles
* greater than 21 years of age

Potentially-eligible cases will be identified by oncologist co-investigators at either site in the course of clinical practice. Research coordinators will be alerted to approach the patient for obtainment of informed consent and enrollment.

***Controls (n=10/institution):*** At MCHS, potential controls will be identified in the primary care setting. Daily appointment rosters of patients scheduled for well checks will be scanned and patients who represent potential matches for enrolled cases based on: no current or past history of cancer diagnoses, sex, age range (+/- 5 years) and smoking status (as defined for cases). Research coordinators will approach the patients for interest in study enrollment. Informed consent will be obtained from patients indicating interest in participation following verification that the potential subject has no prior history of cancer or cancer treatment. Alcohol exposure history (frequency and rate of consumption) will also be documented for controls. Controls enrolled at UW Madison will be sought among family members or friends of the index case with matching criteria and data collection applied as described for MCHS subjects.

***Data (cases and controls)****:* Smoking status (current, former, non smoker), alcohol use: (frequency (daily/occasional) and quantity/exposure), HPV immunization status, age, sex, diagnostic code, stage, tumor site, tumor characteristics documented by a pathologist, and confirmation of no prior history of cancer or treatment in association with newly diagnosed cancer.

***Specimen collection for case and controls****:*  Cases and controls will be requested to donate 2 ml of saliva collected using Pure-Sal Saliva Purification Kit for Liquid Biopsy collection (Oasis Diagnostics) compatible with collection of salivary specimens for cfDNA analysis. Saliva is collected as per manufacturer instructions and passed through a built in filter in the collection cartridge for capture in Eppendorf tubes prior to storage at -80˚C. Specimens will be batched for testing in month 4 and 8. Patients will be requested to donate 2ml of blood collected into EDTA. Briefly, blood is stored at 4˚C and processed within two hours by centrifugation at 800xg for 10 minutes at 4˚C. Plasma is separated and re-centrifuged at 18,000 x g at 4˚C for an additional 15 minutes prior to storage at -80˚C to ensure clearance of cellular debris. Approximately 10 cases and controls will have both saliva and blood tested in month 8 to compare whether cfDNA is detectable in both specimens at early and late stage of OC. Patients will be offered $25 incentive to compensate them for their time. Residual salivary and plasma specimens not tested during the study will be banked for future studies at MCHS.

**Laboratory-based analysis of saliva: assay overview**

Briefly, the rationale supporting this assay assumes that detection of GW alterations in association with tumor DNA has greater sensitivity for predicting presence of cancer than detection of a single alteration. For the proposed analysis, cfDNA is subjected to fragmentation in order to profile GW alterations in individuals with cancer compared to profiles from healthy individuals. Fragments of cfDNA originating from tumors were generally shown to have more variability in length, with short to long cfDNA fragment profiles typically favoring shorter fragments on average compared to those from healthy subjects where fragment length was shown to be more consistent. When GW mapping of cf fragments in 5Mb windows is performed, the genomic libraries align in a position-specific manner across the genome resulting in cancer-specific whole genome profiles that feature approximately 20,000 reads at 1-2x coverage. Because of the shorter fragment length of cfDNA originating from tumor cells compared to those from health individuals, whose fragment lengths were shown to exhibit inter-individual consistency, distinct profiles exhibiting genomic differences in fragment length are detectable in a locus-specific manner. Similarly, methylation patterns surrounding profiled regions associated with smaller cfDNA of tumor origin can be compared to methylation patterns of healthy subjects in these genomic regions.

Analysis of saliva samples will be performed on batched specimens and analyzed twice at 4-month intervals during the study period. Establishment of the assay will be undertaken at the UW Madison Technology Center Service Laboratory, Madison Wisconsin with oversight by Joshua Hyman, PhD, Core Director applying analysis of GW DELFI and meDELFI analysis as previously described by Christiano et al5  following cfDNA fragmentation as described by Phallen10 et al (2017). The gDNA library preparation and application of GW sequencing will be carried out at UW Biotechnology center and cfDNA will be prepped for methylation analysis applying MagMeDIP-seq (Diagenode, Denville NJ). Genome wide sequencing and methylation analysis will be performed using NovaSeq SP flowcells (Illumina, Inc. San Diego, CA).

**Data Analysis**

***Raw Data processing:*** Low-pass GW raw data processing will be done as our previous publication11 [1]. All paired-end (PE) or single-end (SE) fastq files will be trimmed using trim-galore version 0.6.5 to remove low-quality bases and biased read positions. Next, the reads are encoded to map to a three-letter genome via conversion of all cytosines to thymidines or all guanines to adenines for the reads that appear to originate from the reverse-complement strand. Reads will be mapped using Bowtie 2 (version [2.3.5.1](https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.3.5.1)) and Bismark (version 0.22.3), with the default set to both the Watson-strand- and Crick-strand-converted genomes. Alignments with mapping-quality scores of < five will be discarded, retaining only those with a higher best-mapping-quality score in either the Watson or Crick strand. Thereafter encoded read sequences will be replaced by the original read sequences in the final BAM files. Overlapping paired-end reads will also be clipped with the bamUtils clipOverlap function.

***Evaluation of meDELFI:*** MeDelfi analysis will be carried out as described by Christiano et al5. Locally- weighted scatterplot smoothing (LOWESS) regression analysis on both long and short fragments will be carried out with a span setting of 0.75 to the scatterplot of average fragment GC versus coverage calculated for each 100-kb bin to minimize bias attributable to differences in GC. Saliva fragment length due to differences in length due to GC will be subtracted from the LOESS model. For each sample, remaining short and long fragments lengths not impacted by GC will be restored to their original scale by adding back the GW median short and long estimates of coverage. To reduce feature space and noise further, the total GC-adjusted coverage in 5-Mb bins will be calculated. To compare the variability of fragment lengths between cases and control, standard deviation (SD) of the short to long fragmentation profiles per individual will be calculated and median SD compared across groups by a Wilcoxon rank-sum test.

***Mutation Analysis***

***Machine learning***: Validated ML analytical approaches will be applied in identifying detection of malignant transformation associated with OC by GW DELFI, meDELFI and mutational analysis. Potential informatics approaches to conduct analysis includes key supervised ML approaches including Naïve Bayes, Logistic Regression, Random Forest, Multi-layer perceptron-Artificial Neural Network among others to see which algorithms perform optimally, or categorize applying support vector machines. ML algorithm will be developed and executed by using ML libraries (for example WEKA, Matlab or R). ML models will be trained, tested and evaluated using 10-fold cross validation. Risk prediction will be treated as a classification problem, where the datasets will be sorted into two categories based on patient diagnosis of various OCs. Performance of various ML algorithms will be evaluated and compared using the following metrics: total accuracy, recall, sensitivity and specificity. To select the best performing model(s), Receiver Operating Characteristics (ROC) curve, area under ROC (AUC) and Recall-Precision curves for each ML algorithm will be plotted.

Should we propose mutational analysis?

Limitations: The current study is exploratory and the study team cannot anticipate the cancer type, stage of site. As such, the study may reveal variability in profiles with tissue types. Thus not all oral cancer types may be observed among cases and low numbers per oral cancer site and stage may preclude determination of whether profiles vary across patients with cancers in the same tissue, but observation of variability would also be an important finding and could be further explored in a future study in larger sample sizes. Collection of tissue samples as a point of comparison with profiles of various OCs is not currently proposed, but several publically available databases with data on normal oral tissue specimens are available for comparison, albeit not with specimens from the same individual. Such analyses may be feasible in future studies but are not supportable in the current study.