Introduction

Spectral cytopathology (SCP), a combined technique employing infrared micro-spectral imaging for data collection and analysis by unsupervised multivariate statistics, Principal Component Analysis (PCA) has been established as a sensitive detection methodology for disease diagnosis [1]. Unlike traditional cytopathology which relies on the visual inspection of stained cells to achieve a diagnosis, SCP is based on a physical (spectroscopic) measurement of the biochemical composition of a cell at the time of exfoliation. SCP identifies regions of cells that can be linked to disease status since they are directly related to changes in the biochemical composition of the cell [2]. These biochemical changes may not be reflected in the morphology of the cell, which is another advantage of SCP.

SCP is a sensitive, objective diagnostic tool that has been shown to screen for early stages of disease ultimately leading to improved patient prognosis. The goal of the research presented in this poster reflects the preliminary efforts of researchers from five different areas of the oral cavity in order to begin a pre-clinical trial ultimately leading to an automated technology for use in a hospital setting.

ORAL CANCER BACKGROUND

The most recent estimates for oral cancer, as published by the American Cancer Society, state that in the year 2009 over 25,000 people will be diagnosed and 7,000 people will die of oral cancer in the United States. Approximately half of the people who will be diagnosed will survive five years after their diagnosis [3]. Oral cancer is often unable to be detected until late in its development because there is no current screening protocol. Therefore, the morbidity from oral cancer is higher than in many other diseases, Hodgkin’s lymphoma, multiple myeloma, the testes, and endocrine system cancers such as thyroid, or skin cancer (malignant melanoma) [4].

Oral cancers tend to advance to more severe stages of the disease because they typically do not produce symptoms that can be easily diagnosed. Once the oral cancehrs have reached these more severe stages they often metastasize to other localities of the oral cavity, and sometimes to other parts of the body. A patient who has previously been diagnosed with oral cancer has a 20-fold higher risk of developing a second form of cancer even after treatment [4].

Experimental

Normal cellular samples from five different regions of the oral cavity: mouth floor, cheeks, gums, hard palate, and tongue, are collected from volunteers at the Laboratory for Spectra Diagnostics (LSD), and clinical samples are received from Tufts Medical Center (TMC) and Tufts University School of Dental Surgery (TUSD).

All samples are exfoliated from the oral cavity using a commercial cytobrush which is then immersed in various liquid-based medium (SurePath) which are removed from the cytobrush by vortexing. Filtered to remove blood and debris, samples are then centrifuged in Hanks Balanced Salt Solution (ATCC, Manassas, VA) and centrifuged in order to recover a cell pellet. Cytospin centrifugation (Cytospin, Thermo Shandon, Pittsburgh, PA) is used to prepare samples onto low pass density (SurePath) slides (Fisher, Inc., Chesterland, OH) that are permeable for reflectance infrared micro-spectroscopy. The slides are then washed in Millipore water to remove all salt crystals and placed in a desiccator.

A Perkin Elmer Spectrum One/Spotlight 400 (PE 400) imaging IR micro-spectrometer is used to collect data in imaging mode from a 4 mm x 4 mm area of the sample spot (reflectance measurement, 4 cm⁻¹ spectral resolution, 2 scans/pixel, 0.25 μm pixel size, 400-7000 cm⁻¹). Raw data sets from the PE 400 are then imported into a MATLAB based program which allows the data to be analyzed and processed with various machine learning algorithms. Samples are then subjected to a variety of machine learning algorithms including but not limited to: weights of features, Principal Component Analysis (PCA), and various unsupervised and supervised statistical methods. A model is used to analyze the data in an objective manner to mathematically differentiate between classes.

Results

The ongoing SCP study of oral cells will be expanded to include a larger sampling of patients at Northeastern University investigating three areas of the oral cavity: mouth floor, cheek, and tongue. These three areas show mathematically significant differentiation from each other and preliminary studies show other areas (tongue, lower palate, etc.) to cluster within these three groups. The student population will provide data of a broad variety of samples with a good mix of ethnicity, race, gender, nicotine and alcohol use. Also, this population will provide a good sampling of patients with surer prostate infections.

All potential participants involved in this study will be given a form stating that results from the study will remain confidential and collected samples will be used for research purposes and new methodology development only. A standard medical history form with questions such as ethnicity, race, health habits, etc. will be collected along with the samples. Volunteer samples will be collected by a trained individual of the Laboratory for Spectroscopy. All forms that will be used in this study have been approved by the Research Integrity Office at Northeastern University.

Along with the creation of a large dataset of oral cellular samples for the three specified regions, there will also be a focus on the commercialization and automation of the data acquisition and data processing techniques. The use of barcodes will allow for the information for each sample to be randomly assigned a name and allow confidentiality between user and volunteer. Also, a MATLAB based program will combine the separate features currently used for data processing into one user interface is being developed.

By collecting these expanded datasets, it is possible to begin determining the overall accuracy in terms of sensitivity and specificity of SCP. A benefit of these larger datasets will be the ability to eventually train diagnostic algorithms that will be able to detect abnormalities accurately and reliably in aid to the development of a screening method for oral cancer. Diagnostic algorithms have yet to be applied to the diagnosis of oral cytology since large datasets are necessary from dozens of patients with various diagnoses of abnormal oral cytology. Once trained, the diagnostic algorithm can perform the analysis of a very large data set with very little time to produce a reliable diagnostic platform.

The final step towards integrating SCP into a hospital setting after pre-clinical trials have been completed is the production of a faster instrument. This instrument will need to decrease data collection time from the current 6 hours to less than 10 minutes in order to be an ideal method for quick diagnoses in an operating room.

Discussion

Preliminary results show distinct differences between three areas of the oral cavity (mouth floor, cheek, and tongue) when compared to multiple areas; therefore, further investigation of these three regions will allow for the creation of predictive models to diagnose oral disease. Also, normal cheek cells were compared to viral infections as well as drug induced specimens. These changes are not seen morphologically, however spectral differences are present allowing these samples to be distinguished from normal samples.

Expansion of the “normal” dataset will permit the ongoing development of an automated screening workflow for early diagnosis of disease. This new protocol will ultimately allow oral cancer to be detected earlier, and as a result will reduce the morbidity from oral cancer since earlier preventative treatment can be initiated before the cancer is allowed to progress to advanced stages. This pre-clinical study will provide data to further SCP’s potential as a diagnostic tool and will open doors to future possibilities of SCP’s application.