Article:
Abram, T.J., Floriano, P.N., Christodoulides, N. et al. (2016) 'Cytology-on-a-chip' based sensors for monitoring of potentially malignant oral lesions. Oral Oncology, 60. pp. 103-111. ISSN: 1368-8375

https://doi.org/10.1016/j.oraloncology.2016.07.002

'Cytology-on-a-Chip' Based Sensors for Monitoring of Potentially Malignant Oral Lesions

Timothy J. Abram[a, @], Pierre N. Floriano[b, @], Nicolaos Christodoulides[a], Robert James[c], A. Ross Kerr[d], Martin H. Thornhill[e], Spencer W. Redding[f], Nadarajah Vigneswaran[g], Paul M. Speight[h], Julie Vick[c], Craig Murdoch[e], Christine Freeman[e], Anne M. Hegarty[i], Katy D'Apice[j], Joan A. Phelan[d], Patricia Corby[j], Ismael Khouly[j], Jerry Bouquot[g], Nagi M. Demian[k], Y. Etan Weinstock[l], Stephanie Rowan[f], Chih-Ko Yeh[f,m], H. Stan McGuff[n], Frank R. Miller[o], Surabhi Gaur[a], Kailash Karthikeyan[a], Leander Taylor[a], Cathy Le[a], Michael Nguyen[a], Humberto Talavera[a], Rameez Raja[a], Jorge Wong[a], John T. McDevitt[a,p,q **]

[a]Rice University, Department of Bioengineering, Houston, TX, USA
[b]NeoTherma Oncology, Houston, TX, USA
[c]Rho Inc., Chapel Hill, NC, USA
[d]New York University College of Dentistry, Department of Oral and Maxillofacial Pathology, Radiology & Medicine, New York, NY, USA
[e]Academic Unit of Oral & Maxillofacial Medicine & Surgery, University of Sheffield School of Clinical Dentistry, Sheffield, UK.
[f]The University of Texas Health Science Center at San Antonio, Department of Comprehensive Dentistry and Cancer Therapy and Research Center, San Antonio, TX, USA
[g]The University of Texas Health Science Center at Houston, Department of Diagnostic and Biomedical Sciences, Houston, TX, USA
[h]Academic Unit of Oral & Maxillofacial Pathology, University of Sheffield School of Clinical Dentistry, Sheffield, UK.
[i]Unit of Oral Medicine, Charles Clifford Dental Hospital, Sheffield Teaching Hospitals National Health Service Foundation Trust, Sheffield, UK
[j]New York University College of Dentistry, Bluestone Center for Clinical Research, New York, NY, USA
[k]The University of Texas Health Science Center at Houston, Department of Oral and Maxillofacial Surgery, Houston, TX, USA
[l]The University of Texas Health Science Center at Houston, Department of Otolaryngology-Head and Neck Surgery, Houston, TX, USA
[m]South Texas Veterans Health Care System, Geriatric Research, Education, and Clinical Center, San Antonio, TX, USA
[n]The University of Texas Health Science Center at San Antonio, Department of Pathology, San Antonio, TX, USA
[o]The University of Texas Health Science Center at San Antonio, Department of Otolaryngology-Head and Neck Surgery and Cancer Therapy and Research Center, San Antonio, TX, USA
[p]Rice University, Department of Chemistry, Houston, TX, USA
[q]New York University, Department of Biomaterials, New York, NY, USA

** Corresponding Author: John T. McDevitt, Ph.D., Chair, Department of Biomaterials, Bioengineering Institute, New York University, 433 First Avenue, Room 820, New York, NY 10010-4086, USA, Email: mcdevitt@nyu.edu, Phone: 212-998-9204.

@ Authors contributed equally to the work.

Abstract word count: 247 (Limit =250); Manuscript word count: 3,490 (Limit = 3,500); Number of Tables: 1; Number of Figures: 4; Supplementary Materials: 6 References: 37

ABSTRACT

Despite significant advances in surgical procedures and treatment, long-term prognosis for patients with oral cancer remains poor, with survival rates among the lowest of major cancers. Better methods are desperately needed to identify potential malignancies early when treatments are more effective.

Objective: To develop robust classification models from cytology-on-a-chip measurements that mirror diagnostic performance of gold standard approach involving tissue biopsy.

Materials and Methods: Measurements were recorded from 714 prospectively recruited patients with suspicious lesions across 6 diagnostic categories (each confirmed by tissue biopsy -histopathology) using a powerful new ‘cytology-on-a-chip’ approach capable of executing high content analysis at a single cell level. Over 200 cellular features related to biomarker expression, nuclear parameters and cellular morphology were recorded per cell. By cataloging an average of 2,000 cells per patient, these efforts resulted in nearly 13 million indexed objects.

Results: Binary “low-risk”/ “high-risk” models yielded AUC values of 0.88 and 0.84 for training and validation models, respectively, with an accompanying difference in sensitivity + specificity of 6.2%. In terms of accuracy, this model accurately predicted the correct diagnosis approximately 70% of the time, compared to the 69% initial agreement rate of the pool of expert pathologists. Key parameters identified in these models included cell circularity, Ki67 and EGFR expression, nuclear-cytoplasmic ratio, nuclear area, and cell area.

Conclusions: This chip-based approach yields objective data that can be leveraged for diagnosis and management of patients with PMOL as well as uncovering new molecular-level insights behind cytological differences across the OED spectrum.

Highlights

- Cytology-on-chip approach permits rapid molecular and morphometric analysis
- Stable, robust predictive models created from single-cell data
- Prognostic cytology features identified including cell circularity, Ki67 expression
- Unique combination of parameters required for different diagnostic splits

Keywords

- **Cytology**
- **Oral cancer**
- **Oral epithelial dysplasia**
- **Microfluidic**
- **High content analysis**
- **Machine Learning**
- **Random Forest**
- **LASSO**

Abbreviations

- **PMOL – Potentially Malignant Oral Lesion(s)**
- **OED – Oral Epithelial Dysplasia**
- **OSCC - Oral Squamous Cell Carcinoma**
- **HCA – High Content Analysis**
- **AUC – Area Under the (receiver-operator characteristic) Curve**
- **NC ratio – Nuclear-Cytoplasmic area Ratio**

DISCLOSURES

Principal Investigator, John T. McDevitt, has an equity interest in SensoDX, LLC. and also serves on their Scientific Advisory Board. The terms of this arrangement have been reviewed and approved by New York University in accordance with its conflict of interest policies.

INTRODUCTION

It is estimated that 1-2% of adults in the United States present with a worrisome white or red patch or other potentially malignant oral lesion (PMOL) during a routine oral examination [1]. However, the vast majority of these lesions are benign, and only 1-2% will undergo progression into oral squamous cell carcinoma (OSCC) [2,3]. Oral healthcare providers are often the first line of defense in the early detection of oral cancer and are faced with the challenge of recognizing PMOL and deciding which patients to refer for tissue biopsy. This often difficult decision is becoming increasingly burdensome, confounded by the desire to reduce unwarranted biopsies and patient discomfort with the changing landscape of litigation directed at dentists for failing to refer patients [4]. To make matters worse, in PMOL, the histopatholgical diagnosis of oral epithelial dysplasia (OED) is not necessarily predictive of future malignant transformation, creating a demand for more sophisticated early risk assessment tools [5].

Though decades of studies aimed at developing non-invasive, adjunctive aids for monitoring oral lesions have not garnered widespread adoption [6], a new era of rapid, quantitative, and automated tools are beginning to pave the way towards data-driven clinical decision making. Recent advances in a diverse consortium of fields from automated sample processing to statistical machine learning, microfluidic-based single-cell analysis [7–9], and high content analysis/screening [10–14] have fueled a renewed interest in quantitative oral cytology. While offering strong potential for enhanced clinical insight relative to early disease detection, the “-omics” data derived from these new capabilities has a tendency to yield putative clinical models that do not perform as well

in later validation studies. A recent review of 28 studies involving molecular classifiers by Castaldi et al. (2010) [15] found that the majority selected cross-validation practices that overestimated model performance [by ~17% (median) in terms of specificity].

To address these challenges, our team of bioengineers, oral medicine clinicians, oral and maxillofacial pathologists, and cancer biologists, designed and executed a prospective, international clinical study with the ultimate goal of equipping dental practitioners with simple, automated, quantitative risk assessment tools to assist in making difficult biopsy referral decisions. Here we describe this single-cell cytology-on-a-chip approach in the context of developing a multi-parameter image-based clinical decision tool.

The general method for collection and processing of cells within a microfluidic structure was demonstrated previously in the context of a small pilot study involving 52 patients using the single biomarker, EGFR, in order to differentiate between normal mucosa and OSCC [16]. The pilot yielded preliminary logistic regression models with sensitivity and specificity of 97% and 93% respectively, alongside area under the receiver operating characteristic (ROC) curve (AUC) equal to 0.94. These promising results paved the way for this more comprehensive follow-up with a Phase 2/3 clinical characterization-association study.

Many previous quantitative cytology studies have confirmed that measureable differences exist between the extreme phenotypes of normal mucosa/benign lesions and malignant lesions [17–19] such as the increased proportion of small, highly circular cells that resemble more primitive stem cells. Though these differences can be surmised by visual examination by experts and non-experts alike, they miss the more

subtle spectrum of changes seen in PMOL representing the different grades of OED described by histopathologists [20]. An attempt to leverage the subtle measureable differences among OED cytology samples in order to develop objective classification models for PMOL has not existed until now.

MATERIALS AND METHODS

Study Population

This study was approved by the Institutional Review Boards of all participating institutions. Informed consent was obtained for all participants in the trial after the possible consequences of the procedures were explained. The study design and clinical protocol for this study have been reported previously in detail [21]. Briefly, lesion samples from a total of 714 patients were measured, of which 85 were previously diagnosed malignant cases. The slight enrichment of the malignant cases allowed for a more substantial model development process with more equivalent class sizes. All other patients were prospectively recruited based on their exhibiting PMOL for which scalpel biopsy was a necessary part of the standard clinical practice. Histopathological assessment of biopsy specimens was used to place lesions in one of six categories of oral epithelial dysplasia (OED). These comprised 348 benign, 49 mild dysplasia, 18 moderate dysplasia, 12 severe dysplasia, 2 carcinoma in situ (CIS), and 135 malignant lesions in addition to 150 healthy controls. To obtain greater confidence in the gold-standard pathological diagnoses, which have been notoriously unreliable [22,23], a 3-stage adjudication and consensus review process was performed which achieved 100%

consensus agreement from an initial 69.9% agreement rate between any two pathologists [21].

A summary of the major variables analyzed is provided in the Supplementary Methods and Supplementary Tables 1-3. The molecular biomarkers EGFR, αvβ6, CD147, β-catenin, MCM2, and Ki67 were selected based on their capacity, through prior immunohistochemistry studies, to distinguish stages of disease progression towards OSCC for patients with PMOL.

Cytology-on-a-chip Sample Processing

Specific details for cytology-on-a-chip sample processing can be found in the Supplementary Methods and are adapted from the indirect-immunoassay protocol described in Weigum, et. al. (2010) [16]. In summary, sample processing comprised the following steps: 1) the microfluidic device was primed with PBS at a flow rate of 735μL/min for 2 minutes, 2) a cell suspension in 20% glycerol/0.1% PBSA was delivered at 1.5mL/min for 2 minutes whereby single cells were captured on an embedded nanoporous membrane, 3) cells were washed with PBS at 1mL/min for 2.5min, 4) a primary antibody solution was delivered through a 0.2μm PVDF syringe filter at 250μL/min for 2.5min, 5) a wash step similar to step 3 was performed, 6) a secondary antibody solution was delivered under the same conditions as step 4, 7) a final wash step was performed, and 8) fluorescence images were automatically captured.

Data Processing

Unlike most biomarker classification models that are developed with a single measure for each biomarker per patient, this study cataloged an average of 2,000 cells

per patient, resulting in nearly 13 million indexed objects, each with over 200 unique measurements to digitize biomarker expression profiles, and cytoplasm and nuclear morphology. To reduce the magnitude of this database to a computationally appropriate size, statistical measures were used to represent the magnitude and distribution of cellular features for individual patients. These included: mean, median, variance, standard deviation, coefficient of variation (CV), skewness, kurtosis, $10^{th}/25^{th}/75^{th}/90^{th}$-percentiles, and z-scores of 0.5, 2.0, and 3.0 referring to the proportion of cells with biomarker values greater than 0.5/2.0/3.0 standard deviations from the mean. A more detailed description of the complete set of variables used in this study can be found in Supplementary Tables 1, 2, and 3. In order to develop predictive models, the data was split into optimization, training, and testing sets with 250, 311, and 153 patients, respectively.

RESULTS

We have previously reported on the ability to isolate and interrogate single-cells within microfluidic structures for immunophenotyping [24–26], bacterial spore detection [27], and oral exfoliative cytology [28,16,29]. In addition to cell capture, the microfluidic device also serves as a delivery system for efficient transport of fluorescently-labeled antibodies and wash buffer. This cytology-on-a-chip methodology permits concurrent analysis of molecular biomarker expression and cellular/nuclear morphology using over 200 fluorescence intensity and shape parameters for each identified cell extracted from multi-spectral fluorescence images (Fig. 1, Panel II and Supplementary Figure 1). The time to complete this chip-based image analysis is approximately 20 minutes following

sample preparation vs. about 1-3 days to complete a typical gold standard pathology exam.

Model Development

The completion of this clinical trial resulted in the formation of several different classification models used to correlate the biomarker signatures of individual patients to a pre-defined disease category. These categories were derived by dichotomizing the clinical spectrum of diagnoses at several cut-points (Fig. 2a) according to the 2005 WHO 5-point histopathological grading system of OED [20] and the binary “low/high”-risk grading system of OED proposed by Kujan et al [30].

Statistical machine learning techniques that minimize effect size inflation and selection bias were used to develop classification models, namely Random Forests using “out of bag” validation and the L1-regularized logistic regression (LASSO) methodology. In addition, different subsets of parameters were evaluated consisting of biomarkers (molecular + morphometric), clinical lesion characteristics, and demographic risk factors (Supplementary Tables 1, 2). The goal of this approach was to obtain a sparse model to prevent data over-fitting and yield generalizable models with high stability. The classification models were fit to 2/3 of the data (training dataset) while blinded to the remaining 1/3 of the data (test dataset). Importantly, assays were completed in a blind fashion for all samples in the validation phase.

In addition to creating classification models, random forests can be used to automatically probe variable importance in high dimensional datasets. Variable importance, quantified by the Gini-index decrease, can be interpreted as the relative ability of a specific parameter to discriminate between “case” and “non-case” for the

diagnostic split in question, and is therefore an estimate of the informative value of a parameter. These results are summarized as a heatmap (Fig. 2*b, c* and Supplementary Figure 2).

The key information here obtained reveals that molecular biomarkers are better suited for distinguishing benign lesions from earlier stage dysplastic + malignant lesions than for the higher diagnostic splits (cumulative normalized Gini-index ranges: 0.204 – 0.810 (2|3), 0.137 – 0.554 (3|4), 0.139 – 0.540 (H|L), 0.111 – 0.432 (4|5)). Conversely, the morphometric parameters cell area, cell circularity, and NC ratio along with the proliferation biomarker Ki67 demonstrate high importance across all 4 dichotomous splits.

To confirm the stability of the cytology dataset as it relates to model development, Gini-indices across 5 percentile summary measures (Fig. 2*c*) are compared to box plots of key variables (Fig. 2*d*). Variable importance is consistent with the observed biomarker trends for patients in different diagnostic categories. This key finding supports the development of stable models that translate the underlying biological phenomena into measures of variable importance.

LASSO Automated Variable Selection

The use of LASSO methodology helps prevent over-fitting a model by iteratively shrinking parameter effect sizes. The popularity of this method stems from its ability to automatically select the most influential variables and eliminate redundant model parameters that can be a major factor in high-dimensional data sets. A variable association visualization for the LASSO methodology is presented as a chord diagram (Fig. 3) where the chord width corresponds to the relative contribution of each

parameter to model performance across the 4 diagnostic splits, in terms of its standardized odds-ratio.

Key parameters identified by the LASSO methodology include proliferation biomarker Ki67, cell area, cell circularity, nuclear area, and NC-ratio. Consistent with the random forest variable importance analysis, the odds-ratios from molecular biomarkers are found to play a more significant role in differentiating benign lesions from dysplastic + malignant lesions than in models at other diagnostic splits (combined odds-ratios of 9.558 (2|3), 3.337 (3|4), 4.087 (H|L), 3.253 (4|5)). Of the molecular biomarkers, Ki67 provides the most discriminatory information with odds-ratios of 1.162 (median, log-scale), 1.354 (10^{th} percentile, log-scale), 1.184 (10^{th} percentile, log-scale), and 1.307 (10^{th} percentile, log-scale) ((2|3), (3|4), (H|L), (4|5)). Several parameters demonstrate consistent model effects across all 4 diagnostic splits including cell area (average odds-ratios of 0.835, 0.935, 0.969, 0.904), nuclear area (1.111, 1.023, 1.116, 1.028), and lesion size (1.139, 1.072, 1.124, 1.183), ((2|3), (3|4), (H|L), (4|5)). Furthermore, LASSO results indicated that each of the four diagnostic targets requires a unique combination of variables to achieve its highest performance.

Cellular Phenotype Identified by Morphometric Parameters

These key indicators for detecting dysplastic or malignant changes can also be leveraged along with the cytology image database created in this study to identify unique cellular phenotypes. In addition to routinely identified cell phenotypes such as binucleated cells and cells with micronuclei, 4 additional "categories" are created to describe unique cell sub-populations. These categories are defined based on significant differences across the key parameters corresponding to smoothness of cytoplasmic

borders, cell circularity, cell area, nuclear area, NC-ratio, DAPI intensity, and Phalloidin intensity. The distribution of these custom phenotype labels for 300 randomized cells across 4 different patient diagnostic groups is compared (Fig. 4). Significantly higher proportions of cells that fit definitions for type B, C, D, and F in patients with OSCC are identified, referring to cells with higher than average circularity, higher NC-ratio, smaller overall cytoplasm area, enlarged nuclei, and those described as polynucleated. Morphometry-based phenotype identification also enables white blood cell (WBC) enumeration in cytology samples, an indicator with potential diagnostic utility that will be discussed in future publications. The ability to translate quantitative cytology metrics into visualizable cellular features has the potential to assist cytopathologists in discovering novel feature sets based on objective information.

Model Performance

A single primary diagnostic model and numerous secondary models are selected by the study team through a rigorous process involving participation of three independent statisticians. This blinding process allows for this trial to complete both model development and model validation. The latter is accomplished in a fully blinded manner with the oversight of a contract research organization (Rho Inc., Chapel Hill, NC) using external data that is not employed for model development. A summary of the performance values obtained for the LASSO and random forest classification models is provided (Table 1).

The study team selected the LASSO logistic regression model trained with the molecular biomarker and lesion characteristic datasets for differentiating “Low-risk” and “High-risk” lesions as the primary classification model in this study based on its superior

performance of the LASSO model in the development stage (based on averaged high sensitivity, specificity, and AUC). These efforts yield AUC values of 0.88 and 0.84 for the training and validation models, respectively.

Across the training and validation datasets, the LASSO models that includes lesion characteristics outperforms those trained on biomarkers alone by boosting specificity an average of 6.56% (SD = 3.77%) (Sens: avg = -0.05%, SD = 4.31%; AUC: avg = 2.69%, SD = 1.28%). Interestingly, the addition of lesion characteristics did not appear to affect random forest models significantly (Sens: avg = -1.78%, SD = 2.19%; Spec: 3.34%, SD = 4.22%).

On average, validation performance for random forests displays a 2.58% drop in sensitivity (SD= 3.50%) and an increase of 3.01% for specificity (SD=2.31%), compared to a 5.88% sensitivity drop for LASSO (SD = 5.07%) and a corresponding increase of 3.79% in specificity (SD = 2.96%). Additionally, LASSO model AUCs decreases an average of 4.0% (SD = 2.41%) across the different diagnostic splits between the validation and training datasets.

Parameters with significant information in discriminating between “Low-risk” and “High-risk” lesions selected by the primary LASSO model includes cell circularity (90^{th} percentile), nuclear Ki67 intensity (10^{th} percentile and coefficient of variation (cv)), cell-surface EGFR intensity (standard deviation), NC-ratio (median), nuclear area (skewness and cv), cell area (25^{th} percentile), and the lesion characteristics lichen planus and lesion size (long axis). Box plots illustrating the stable, monotonic trends of these variables across different diagnostic categories can be found in Supplementary Figure 3.

DISCUSSION

In developing a chip-based approach to obtain a quantitative risk assessment for monitoring PMOL, this study sought to address three main questions: 1) Can a high content analysis (HCA) workflow be applied to primary patient cells? 2) How does model performance and composition change as a function of the diagnostic split position? 3) Can quantitative cytology tests produce diagnostic accuracy that rivals the gold standard pathology tests? Each of these key knowledge gaps is discussed below.

Can HCA be applied to primary patient cells?

Though HCA has been predominantly driven by screening applications involving cell lines grown in multi-well plates and advances in high-throughput laboratory automation, this study exemplifies how HCA can be applied to primary, patient-derived samples for personalized cellular scoring. This cytology-on-a-chip approach is unique in its ability to analyze hundreds of morphometric and biomarker expression parameters automatically across an entire cytology sample agnostic to clinical judgment or clinical history. In some instances, parameters that contributed to model performance and stability, such as NC-ratio, nuclear area, and Ki67 expression, confirmed previous reports of their roles in cyto- and histo-pathological grading [16],[17],[18],[31],[32].

This study also identifies biomarkers that present new insights into how the quantification of cell and nuclear morphometry can provide rich data streams that track with the different grades of OED. In the primary LASSO model, two nuclear area measures (skewness and cv) are identified as important variables in discriminating between “low-risk” and “high-risk” lesions. As summary measures of the distribution of nuclear area, both provide insight into the level of ansionucleosis (variation in size and

shape of nuclei [33]) for each patient sample, which is a prominent feature of dysplasia and malignancy [34],[33], but not often quantified.

Cell circularity, a measure of the "roundness" of the cytoplasmic membrane, is frequently identified in several models to be a highly performing parameter. An increase in cell circularity may be attributed to the decrease in the amount of cytoplasm and reduction in the degree of cellular cohesion which have both been reported to occur with increased dysplasia grade [32]. The dominant summary statistic for cell circularity in these models is its 90th percentile measure, unlike the range of summary measures for cell area and NC-ratio, indicating cell circularity is skewed towards a fraction of cells with circularity measures in the upper 10% of the entire cell sample (Fig. 4 - cell types "B" and "C").

How does model performance change with diagnostic classification?

One consistent finding from this study is that, at the higher end of the clinical spectrum (moderate-severe dysplasia cut-off), LASSO models (validation sens/spec/AUC = 85.7%/80.0%/0.883) and random forest models (validation sens/spec = 78.8%/68.3%) outperform models distinguishing low-grade lesions in terms of sensitivity, specificity, and AUC (Table 1). Longitudinal studies would be necessary to validate the malignant potential of these identified lesions. Histopathological grading has been shown to behave similarly, where higher accuracy and inter-observer agreement have been reported for high grade dysplastic or malignant lesions than for low grade dysplastic lesions, where considerable overlap from inflammatory and reactive changes can confound grading systems [35]. However, longitudinal studies are necessary to validate the malignant potential of these identified lesions.

Can quantitative cytology rival gold standard diagnostic accuracy?

Though some drop in performance is expected between training and validation, similar values for these measures can indicate low risk of model over-fitting and greater generalizability. When comparing model performance between training and validation datasets across all 4 diagnostic splits in this study, random forest models displayed the greatest stability, with a mean difference of 2.58% for sensitivity (SD: 3.50%) and 3.01% for specificity (SD: 2.31%). The LASSO models demonstrated fair stability with a mean difference of 5.88% for sensitivity (SD: 5.07%) and 3.79% for specificity (SD: 2.96%). Indeed, robust classification models are developed that accurately mirrored the performance of histopathological assessment by expert pathologists. In terms of accuracy, the “low/high-risk” model accurately predicted correct diagnoses approximately 70% of the time, compared to the 69% initial agreement rate of the pool of expert pathologists. Furthermore, the level of agreement between model development and model validation far exceeds typical studies as reported by Castaldi et al. (2010) [15]. Furthermore, the design of the model development process presented here championed an additional level of rigor to ensure future generalizability and the absence of human bias which included oversight from a contract research organization, three layers of blinding to the true classification of patient lesions, two sets of independent biostatisticians, and utilization of statistical methods that performed automatic feature selection.

In addition to the objective information that can be obtained with quantitative cytology, the ‘cytology-on-a-chip’ technique presented here is amenable to complete process automation due to its foundations in microfluidic reagent handling, image

acquisition, and computational algorithms. Future portable analyzers, similar to those being developed in the McDevitt lab [37], will have the potential to bring sophisticated tools for monitoring PMOL in patients with suspicious lesions to regions with limited access to expert pathologist review.

CONCLUSIONS

We have demonstrated the utility of a new cytology-on-chip that is capable of high-content, single-cell analysis across cellular and nuclear morphometric and molecular biomarker expression measurements. This new clinical decision tool has been developed and validated in the context of a major clinical study and has resulted in a rich database that has been exploited to develop new routines that provide insights into cytology characteristics associated with PMOL. These efforts demonstrate robust clinical performance and stable parameters for LASSO models with sensitivity and specificity values of ~85% and 70% in the model development phase and similar values (within 7%) in the validation phase.

Additionally, superior model performance is found to be associated with heterogeneous data sources. Models trained on data from biomarker expression, morphometric features, clinical impressions, and patient risk factors achieves superior performance compared to models trained on restricted data. This essential finding has implications for future diagnostic applications and adjunctive aid development by emphasizing the need to integrate several different sources of information into a successful risk assessment rather than relying on the expression of a single biomarker or feature.

ACKNOWLEDGMENTS

The authors would like to thank Rho, Inc. (Chapel Hill, North Carolina) for providing assistance with patient data management and statistical analysis. The study team would especially like to thank Vanessa Hearnden, BSc, PhD of the University of Sheffield for technical and administrative support and Deborah Holt, MBBS, BDS, PhD of Sheffield Teaching Hospitals NHS Foundation Trust for performing biopsies.

The authors would also like to thank the University of Texas Health Science Center at San Antonio (UTHSCSA), University of Texas Health Science Center at Houston (UTHSCH), New York University / Bluestone Center for Clinical Research, Sheffield Teaching Hospitals NHS Foundation Trust and the University of Sheffield for assistance in obtaining clinical samples.

GRANT AND PHILANTHROPIC SUPPORT

Funding for this work was provided by the National Institutes of Health (NIH) through the National Institute of Dental and Craniofacial Research (Award Number 1RC2DE020785-01, 5RC2DE020785-02, 3RC2DE020785-02S1, 3RC2DE020785-02S2). The content of this paper is solely the responsibility of the authors and does not necessarily represent or reflect the official views of the NIH or the US government. Segments of this work are supported by Renaissance Health Service Corporation and Delta Dental of Michigan.

REFERENCES

[1] Shulman JD, Beach MM, Rivera-Hidalgo F. The prevalence of oral mucosal lesions in U.S. adults: Data from the Third National Health and Nutrition Examination Survey, 1988-1994. J Am Dent Assoc 2004;135:1279–86.

[2] Lee JJ, Hong WK, Hittelman WN, Mao L, Lotan R, Shin DM, et al. Predicting cancer development in oral leukoplakia: Ten years of translational research. Clin Cancer Res 2000;6:1702–10.

[3] Napier SS, Speight PM. Natural history of potentially malignant oral lesions and conditions: An overview of the literature. J Oral Pathol Med 2008;37:1–10. doi:10.1111/j.1600-0714.2007.00579.x.

[4] Lydiatt DD. Cancer of the oral cavity and medical malpractice. Laryngoscope 2002;112:816–9.

[5] Holmstrup P, Vedtofte P, Reibel J, Stoltze K. Long-term treatment outcome of oral premalignant lesions. Oral Oncol 2006;42:461–74. doi:10.1016/j.oraloncology.2005.08.011.

[6] Macey R, Walsh T, Brocklehurst P, Kerr AR, Liu JL, Lingen MW, et al. Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions. Cochrane Database Syst. Rev., John Wiley & Sons, Ltd; 2015.

[7] Pushkarsky I, Liu Y, Weaver W, Su T-W, Mudanyali O, Ozcan A, et al. Automated single-cell motility analysis on a chip using lensfree microscopy. Sci Rep 2014;4:4717.

[8] Weaver WM, Tseng P, Kunze A, Masaeli M, Chung AJ, Dudani JS, et al. Advances in high-throughput single-cell microtechnologies. Curr Opin Biotechnol 2014;25:114–23. doi:10.1016/j.copbio.2013.09.005.

[9] Tan AP, Dudani JS, Arshi A, Lee RJ, Tse HTK, Gossett DR, et al. Continuous-flow cytomorphological staining and analysis. Lab Chip - Miniaturisation Chem Biol 2014;14:522–31. doi:10.1039/c3lc50870f.

[10] Ljosa V, Sokolnicki KL, Carpenter AE. Annotated high-throughput microscopy image sets for validation. Nat Methods 2012;9:637–637. doi:10.1038/nmeth.2083.

[11] Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 2006;7:R100. doi:10.1186/gb-2006-7-10-r100.

[12] Kamentsky L, Jones TR, Fraser A, Bray M, Logan DJ, Madden KL, et al. Improved structure, function and compatibility for cellprofiler: Modular high-throughput image analysis software. Bioinformatics 2011;27:1179–80. doi:10.1093/bioinformatics/btr095.

[13] Singh S, Carpenter AE, Genovesio A. Increasing the Content of High-Content Screening An Overview. J Biomol Screen 2014:1087057114528537. doi:10.1177/1087057114528537.

[14] Bray M-A, Fraser AN, Hasaka TP, Carpenter AE. Workflow and metrics for image quality control in large-scale high-content screens. J Biomol Screen 2012;17:266–74. doi:10.1177/1087057111420292.

[15] Castaldi PJ, Dahabreh IJ, Ioannidis JPA. An empirical assessment of validation practices for molecular classifiers. Brief Bioinform 2011;12:189–202. doi:10.1093/bib/bbq073.

[16] Weigum SE, Floriano PN, Redding SW, Yeh C-K, Westbrook SD, McGuff HS, et al. Nano-Bio-Chip sensor platform for examination of oral exfoliative cytology. Cancer Prev Res (Phila Pa) 2010;3:518–28. doi:10.1158/1940-6207.CAPR-09-0139.

[17] Johnston DG. Cytoplasmic : nuclear ratios in the cytological diagnosis of cancer. Cancer 1952;5:945–9.

[18] Goldsby JW, Staats OJ. Nuclear changes of intraoral exfoliated cells of six patients with sickle-cell disease. Oral Surg Oral Med Oral Pathol 1963;16:1042–8.

[19] Ogden GR, Cowpe JG, Wight AJ. Oral exfoliative cytology: review of methods of assessment. J Oral Pathol Med 1997;26:201–5. doi:10.1111/j.1600-0714.1997.tb01224.x.
[20] Barnes L, Eveson JW, Reichart P, Sidransky D. World Health Organization classification of tumors: pathology and genetics of head and neck tumors. Lyon: IARC Press; 2005.
[21] Speight PM, Abram TJ, Floriano PN, James R, Vick J, Thornhill MH, et al. Inter-Observer Agreement in Dysplasia Grading: Towards an Enhanced Gold Standard for Clinical Pathology Trials. Oral Surg Oral Med Oral Pathol Oral Radiol n.d. doi:10.1016/j.oooo.2015.05.023.
[22] Bosman FT. Dysplasia classification: Pathology in disgrace? J Pathol 2001;194:143–4. doi:10.1002/1096-9896(200106)194:2<143::AID-PATH883>3.0.CO;2-Z.
[23] Warnakulasuriya S, Reibel J, Bouquot J, Dabelsteen E. Oral epithelial dysplasia classification systems: Predictive value, utility, weaknesses and scope for improvement. J Oral Pathol Med 2008;37:127–33. doi:10.1111/j.1600-0714.2007.00584.x.
[24] Rodriguez WR, Christodoulides N, Floriano PN, Graham S, Mohanty S, Dixon M, et al. A microchip CD4 counting method for HIV monitoring in resource-poor settings. PLoS Med 2005;2:0663–72. doi:10.1371/journal.pmed.0020182.
[25] Floriano PN, Acosta S, Christodoulides N, Weigum S, McDevitt JT. Microchip-based enumeration of human white blood cells. Methods Mol Biol Clifton NJ 2007;385:53–64.
[26] Jokerst JV, Floriano PN, Christodoulides N, Simmons GW, McDevitt JT. Integration of semiconductor quantum dots into nano-bio-chip systems for enumeration of CD4+ T cell counts at the point-of-need. Lab Chip - Miniaturisation Chem Biol 2008;8:2079–90. doi:10.1039/b817116e.
[27] Floriano PN, Christodoulides N, Romanovicz D, Bernard B, Simmons GW, Cavell M, et al. Membrane-based on-line optical analysis system for rapid detection of bacteria and spores. Biosens Bioelectron 2005;20:2079–88. doi:10.1016/j.bios.2004.08.046.
[28] Weigum SE, Floriano PN, Christodoulides N, McDevitt JT. Cell-based sensor for analysis of EGFR biomarker expression in oral cancer. Lab Chip - Miniaturisation Chem Biol 2007;7:995–1003. doi:10.1039/b703918b.
[29] Floriano PN, Abram T, Taylor L, Le C, Talavera H, Nguyen M, et al. Programmable bio-nanochip-based cytologic testing of oral potentially malignant disorders in Fanconi anemia. Oral Dis 2015;21:593–601. doi:10.1111/odi.12321.
[30] Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. Oral Oncol 2006;42:987–93. doi:10.1016/j.oraloncology.2005.12.014.
[31] Yu Y-H, Morales J, Feng L, Jack Lee J, El-Naggar AK, Vigneswaran N. CD147 and Ki-67 overexpression confers poor prognosis in squamous cell carcinoma of oral tongue: A tissue microarray study 2014.
[32] Babshet M, Nandimath K, Pervatikar S, Naikmasur V. Efficacy of oral brush cytology in the evaluation of the oral premalignant and malignant lesions. J Cytol Indian Acad Cytol 2011;28:165–72. doi:10.4103/0970-9371.86342.
[33] Mohan H, Mohan S. Essential Pathology For Dental Students. JP Medical Ltd; 2011.
[34] Gray W, Kocjan G. Diagnostic Cytopathology. Elsevier Health Sciences; 2010.
[35] Montgomery E. Is there a way for pathologists to decrease interobserver variability in the diagnosis of dysplasia? Arch Pathol Lab Med 2005;129:174–6.
[36] Fuller C, Camilon R, Nguyen S, Jennings J, Day T, Gillespie MB. Adjunctive diagnostic techniques for oral lesions of unknown malignant potential: Systematic review with meta-analysis. Head Neck 2015;37:755–62. doi:10.1002/hed.23667.
[37] McRae MP, Simmons GW, Wong J, Shadfan B, Gopalkrishnan S, Christodoulides N, et al. Programmable bio-nano-chip system: a flexible point-of-care platform for bioscience and clinical measurements. Lab Chip 2015;15:4020–31. doi:10.1039/C5LC00636H.

Figure Captions

Figure 1 - Diagram of cytology-on-a-chip processing and sample images. Panel I.) Representative histopathological (H&E staining) images (A-B) and immunofluorescence-cytology images (C-D) for 4 different patients. (A, C) are derived from Benign (Fig. 1.II.A = lichen planus diagnosis) and (B, D) from OSCC diagnoses as confirmed from independent agreement between two reviewing pathologists. Scale bars for A, B, C, and D = 100 μm. Panel II.) Diagram of "cytology-on-chip" sample processing in which a brush cytology sample is collected (A), processed in a suspension, and delivered through the microfluidic platform (B) to a cell-capture, nano-porous membrane (C). Multi-spectral fluorescence images are recorded (D) and analyzed with automated software to identify single cells (E) and extract these regions for measurement (F).

Figure 2 - **Variable importance from Random Forest models**. A) Visual representation of diagnostic spectrum and the 4 diagnostic splits used in this trial to dichotomize diagnoses into either "Case" or "Non-case". B) Univariate heat map of Gini values resulting from Random Forest modeling to demonstrate variable importance across all 4 diagnostic splits (y axis). Gini values from each model were scaled between 0 and 1 to generalize relative variable importance across all models. A value of 1 implies the variable is better able to discriminate between "case" and "non-case" than a variable with a value closer to 0. Groups of parameters are labeled by their corresponding marker; single boxes represent specific summary measures (see Supplementary Figure **2** for more detailed labeling). Heatmaps should not be interpreted as "expression", but rather as the information content associated with each parameter in its ability to differentiate between "case" and "non-case". C) Parameter subset from (B) to focus on summary percentile measurements (p10, p25, p50, p75, p90 = 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentile values). D) Box-and-whisker plots showing the distribution of median values for Circularity (unit-less value between 0 and 1), Ki67 (units = arbitrary fluorescence units (afu), nuclear-to-cytoplasmic (NC) ratio (unit-less ratio), and Cell Area (units = px^2), respectively. The box bottom and top represent the 25^{th} and 75^{th} percentiles, respectively. Median values are connected between boxes, and whiskers down/up to 1.5 interquartile range. (Ben = "benign", Mild = "mild dysplasia", Mod+ = "moderate/severe/CIS dysplasia", Mal = "malignant").

Figure 3 - **Chord Diagram of LASSO model parameter odds-ratios.** Chord width refers to the relative contribution of a particular variable, based on standardized odds-ratios, calculated by exponentiating individual parameter coefficients from the logistic regression models. Odds-ratios of single parameters represent the odds that a model will predict the "Case" diagnosis for an increase of one standard deviation for the standardized (unit-less) parameter while holding all other parameters constant. Model splits are identified on the right side and their corresponding variables on the left side. Parameters are further color-coordinated by categorical grouping: Lesion characteristics (L. Size = Lesion Size, L. Color = Lesion Color, LP = presence of the clinical features of lichen planus), Nuclear parameters (NC = NC-ratio, Nuc Area = nuclear area), Biomarkers (αvβ6, CD147, EGFR, Ki67, MCM2), and Cytomorphometric parameters (circularity, cell area). Summary statistic measures include A: coefficient of variation, B: variance, C: median, D: 10^{th} percentile, E: 25^{th} percentile, F: 75^{th} percentile, G: 90^{th} percentile, H: skewness, I: standard deviation, J: >0.5 Z-Score, K: >2.0 Z-Score, L: short-axis, M: long-axis, *: Log-scale, ^{2}:squared

Figure 4 - Cellular Phenotype Identified by Morphometric Parameters. Panel I) Scatterplot and density histograms for two morphometric parameters (Maximum Feret diameter and mean Phalloidin intensity) used to distinguish sub-populations of cells. Panel II) Bar plot of cell counts for each of the phenotypes identified by Panel III for 300 randomly selected cells from patients with final adjudicated diagnoses in categories "Normal", "Benign", "Dysplastic" (including mild, moderate, severe dysplasia and CIS), and "OSCC". These plots are visualized as a continuous line where peaks refer to the number of cells identified in each case in order to illustrate a "phenotype fingerprint" of the disease categories. Panel III) Representative images of unique cellular phenotypes identified by significant differences in key morphometric parameters. Each thumbnail is cropped to the same dimensions of 120 μm x120 μm. Phenotypic categories included A) Cells with smooth cytoplasmic border and high circularity, but low NC-ratio, B) Cells with high circularity, high NC-ratio, and medium cytoplasm area, C) Cells with high circularity, high N-C ratio, and small cytoplasm area, D) large cells with enlarged nuclei, E) Binucleated cells, F) Polynucleated cells, G) Cells with micronuclei, and H) Normal appearing squamous cells.

Table 1. Final model performance values for LASSO and random forest models. ("LC = parameter set with of lesion characteristics; Sens = sensitivity, Spec = specificity, AUC = area under ROC curve).

			Training			Validation		
	Dichotomous split	**Parameters**	**Sens %**	**Spec %**	**AUC**	**Sens%**	**Spec%**	**AUC**
Lasso	Low || High	Biomarkers Only	90.8	58.7	0.871	85.0	59.2	0.802
		Biomarkers + LC	90.7	64.5	0.884	78.6	70.4	0.836
	Benign || Mild	Biomarkers Only	90.1	45.0	0.814	92.7	53.1	0.800
		Biomarkers + LC	89.9	52.0	0.844	89.1	55.1	0.846
	Mild || Moderate	Biomarkers Only	89.7	60.1	0.869	82.9	60.5	0.809
		Biomarkers + LC	89.6	67.2	0.880	82.9	72.3	0.839
	Moderate || Severe	Biomarkers Only	89.4	72.6	0.903	77.1	73.6	0.846
		Biomarkers + LC	90.8	73.8	0.917	85.7	80.0	0.883
Random Forest	Low || High	Biomarkers Only	89.5	51.1	-	86.8	53.0	-
		Biomarkers + LC	89.3	51.3	-	86.8	51.3	-
	Benign || Mild	Biomarkers Only	91.0	42.0	-	90.9	50.0	-
		Biomarkers + LC	89.0	48.0	-	90.9	52.0	-
	Mild || Moderate	Biomarkers Only	89.7	49.4	-	87.2	51.8	-
		Biomarkers + LC	87.0	50.4	-	87.2	52.6	-
	Moderate || Severe	Biomarkers Only	90.9	56.3	-	84.9	59.2	-
		Biomarkers + LC	87.7	65.6	-	78.8	68.3	-

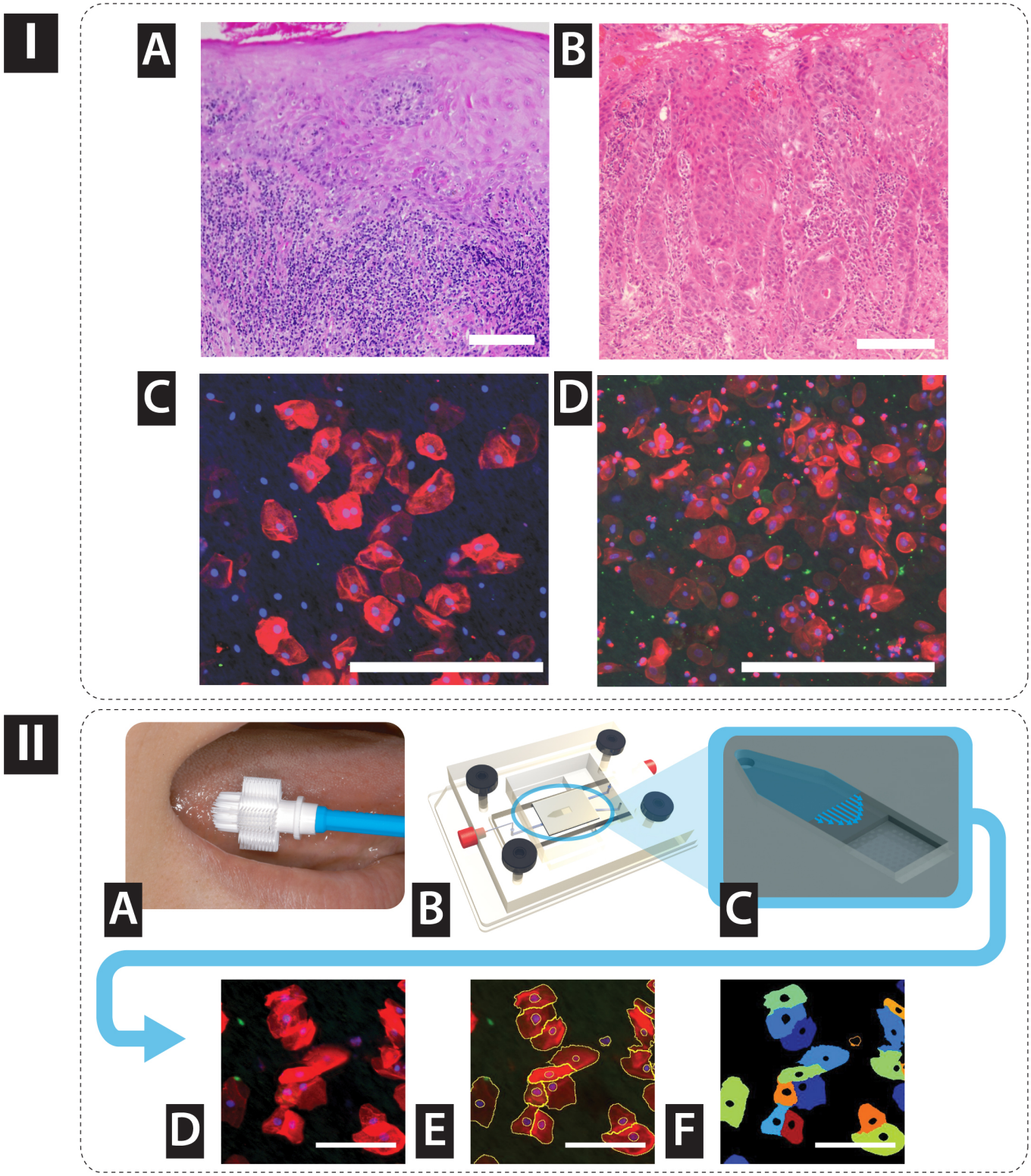


Figure 1.

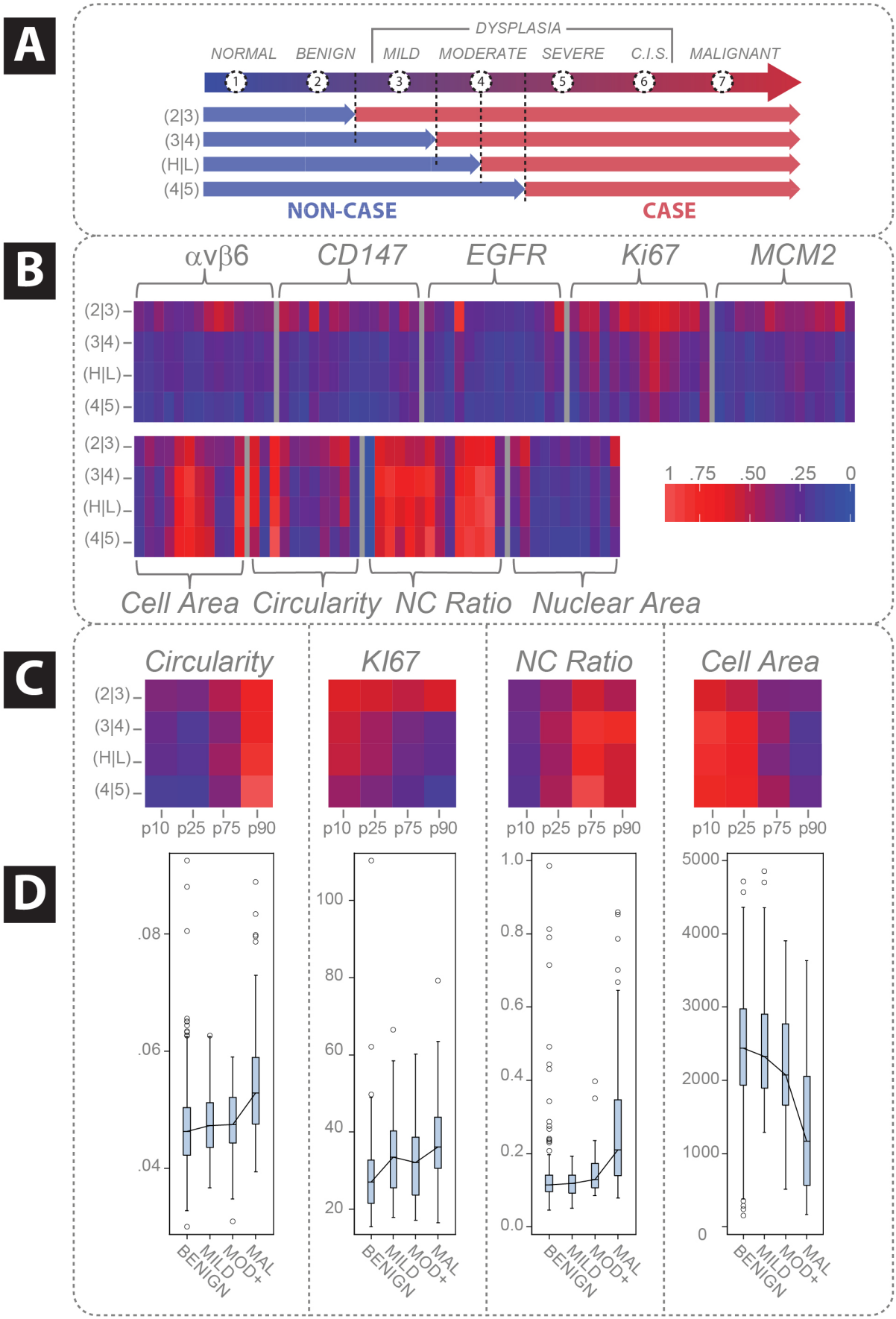

A
DYSPLASIA
NORMAL
BENIGN
MILD
MODERATE
SEVERE
C.I.S.
MALIGNANT
1
2
3
4
5
6
7
(2|3)
(3|4)
(H|L)
(4|5)
NON-CASE
CASE
B
αvβ6
CD147
EGFR
Ki67
MCM2
(2|3)
(3|4)
(H|L)
(4|5)
1 .75 .50 .25 0
Cell Area
Circularity
NC Ratio
Nuclear Area
C
Circularity
KI67
NC Ratio
Cell Area
(2|3)
(3|4)
(H|L)
(4|5)
p10 p25 p75 p90
D
.08
.06
.04
100
80
60
40
20
1.0
0.8
0.6
0.4
0.2
0.0
5000
4000
3000
2000
1000
0
BENIGN
MILD
MOD+
MAL

Figure 2.

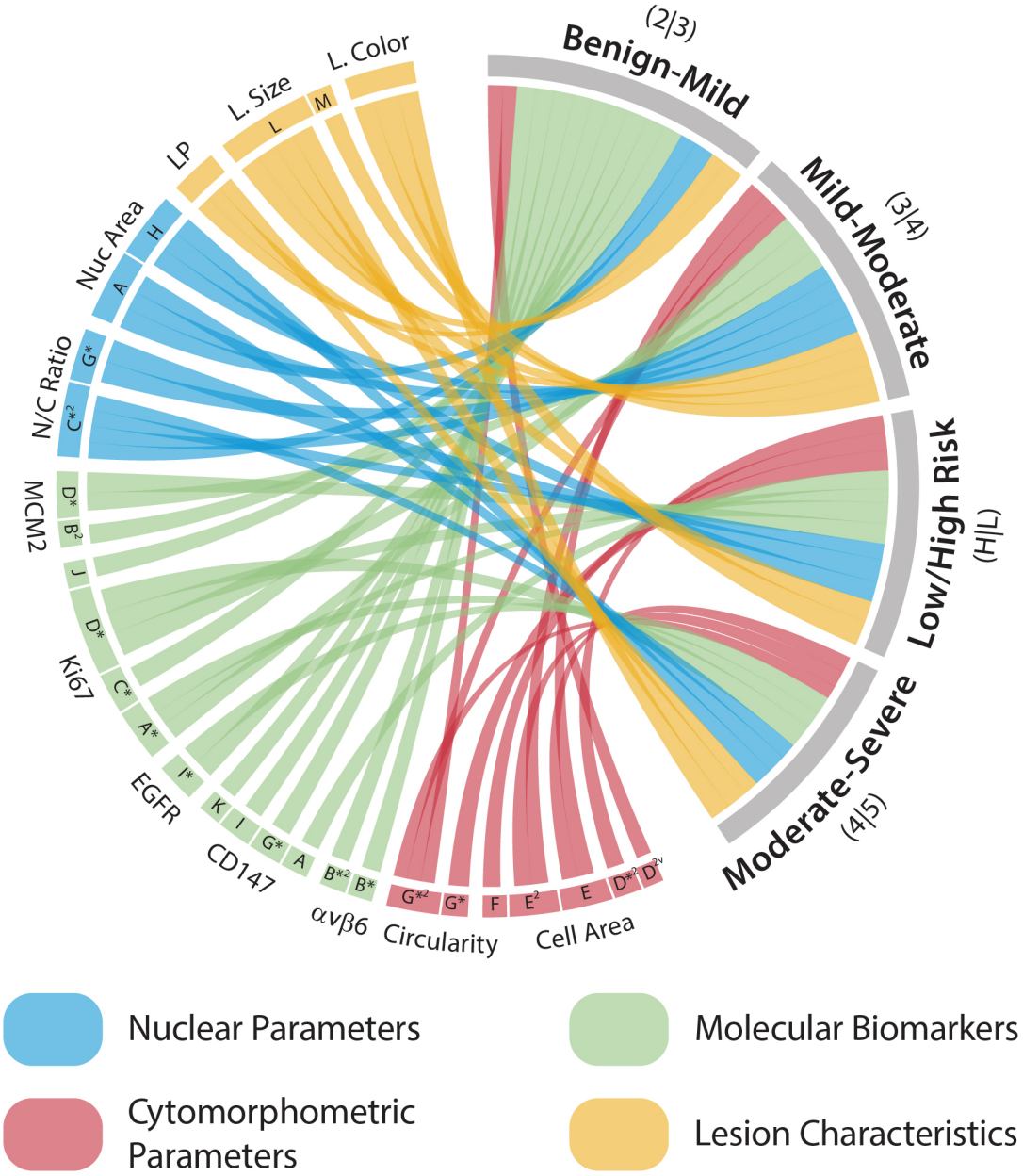

Benign-Mild
(2|3)
Mild-Moderate
(3|4)
Low/High Risk
(H|L)
Moderate-Severe
(4|5)
L. Color
L. Size
LP
Nuc Area
N/C Ratio
MCM2
Ki67
EGFR
CD147
αvβ6
Circularity
Cell Area
Nuclear Parameters
Molecular Biomarkers
Cytomorphometric Parameters
Lesion Characteristics

Figure 3.

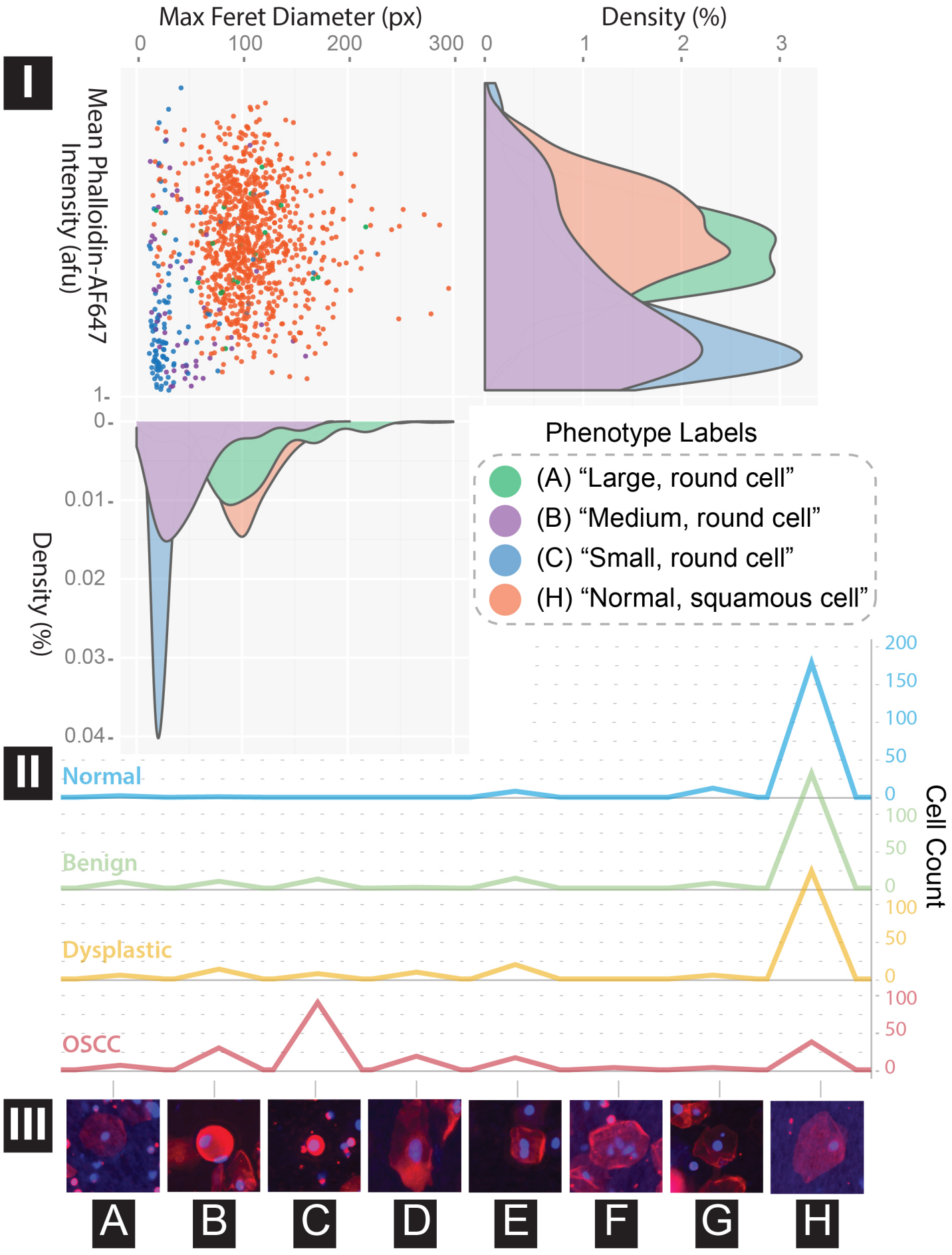

I
Max Feret Diameter (px)
0
100
200
300
Density (%)
0
1
2
3
Mean Phalloidin-AF647 Intensity (afu)
1
0
Density (%)
0.01
0.02
0.03
0.04
Phenotype Labels
(A) “Large, round cell”
(B) “Medium, round cell”
(C) “Small, round cell”
(H) “Normal, squamous cell”
II
Normal
Benign
Dysplastic
OSCC
200
150
100
50
0
Cell Count
III
A
B
C
D
E
F
G
H

Figure 4.

SUPPLEMENTARY MATERIALS

- **Supplementary Methods**
- **Supplementary Figure 1**: “Cytology-on-a-chip” device evolution
- **Supplementary Figure 2**: Heatmap of random forest variable importance based on Gini index
- **Supplementary Figure 3**: Box and whisker plots of selected variables identified in primary LASSO model
- **Supplementary Table 1**: Detailed description of the different morphometric and biomarker measurements collected in this trial
- **Supplementary Table 2**: Lesion characteristics and Demographic risk factor parameters and definitions
- **Supplementary Table 3**: Abbreviations and Full names of biomarker labels from Supplementary Figure 2

Supplementary Methods

Rationale on Biomarker Selection

In addition to EGFR (epidermal growth factor receptor) which was studied in our previous trial[1], the biomarker selection was expanded to 5 additional molecular biomarkers (αvβ6, CD147, β-catenin, MCM2 (Minichromosome Maintenance Complex Component 2), and Ki67). These markers fall predominantly into two groups: 1) cell surface markers: **αvβ6** (an integrin receptor undetectable in normal oral epithelium, but highly expressed in dysplasia and OSCC[2,3]); **CD147** (EMMPRIN) (a multifaceted molecule that facilitates tumor progression by several mechanisms[4]); **β-catenin** (a transcription factor in the Wnt pathway, which promotes transcription of genes involved in cellular proliferation and apoptosis inhibition[5]); **EGFR** (a transmembrane glycoprotein whose overexpression may contribute to tumor progression[5]); and 2) nuclear markers of proliferation/cell cycle regulation: **MCM2** (an essential component for DNA replication associated with deregulated expression in dysplastic and malignant epithelial cells[6,7]); **Ki67** (a marker of proliferation that is overexpressed at initial stages of oral carcinogenesis[5]).

Cytology-on-a-Chip Protocol

Immediately after brush cytology samples were collected, cells were harvested by vortexing the brush head in minimum essential medium (MEM) culture media, followed by a PBS wash, re-suspension in FBS containing 10% of the cryo-preservative dimethyl-sulfoxide (DMSO), frozen, and stored in a -80^{o}C freezer at the clinical site until samples were shipped in batches to the McDevitt Laboratory at Rice University.

Prototype laminate microfluidic devices were assembled in-house by manually aligning separate component layers. The base layer was laser machined from 3/8" acrylic (McMaster Carr, Elmhurst, IL, # 8774K42) to accommodate fluidic ports and a sub-membrane cavity. A stainless steel frit was embedded in the acrylic base as a membrane support. Directly over the frit, a polycarbonate track-etched membrane with 0.4μm pores was anchored by the fluid delivery adhesive channel assembly. These microfluidic channels were cut from double-sided adhesive film (DSA), vinyl, and thermoplastic film using a precision plotter cutter with sub-25-μm resolution (Summa D75, Summa Inc., Seattle, WA). Dimensions for the

device were as follows: channel height = 125μm, channel width = 1.2mm, and exposed membrane area = 20mm^2.

Prior to processing on the microfluidic device, patient samples were thawed rapidly in a 37^{0}C water bath, washed with PBS, and fixed for one hour in 0.5% formaldehyde prepared fresh from a 16% stock solution (Polysciences, Warrington, PA, #18814-20). After fixation, cells were washed twice in PBS, re-suspended in 150μL 0.1%PBS with 0.1% BSA (PBSA), and stored at 4^{0}C until ready to process. Before sample delivery, the cell suspension was diluted in a 20% glycerol/0.1% PBSA solution to improve cell distribution across the membrane and to reduce cell clumping.

Using a custom built manifold connecting external fluidic tubing to the inlet and outlet ports of the microfluidic device, the assembly was positioned on a robotically controlled microscope stage (ProScan II, Prior Scientific, Cambridge, UK) and connected to a peristaltic pump (SciQ 400, Watson Marlow, Wilmington, MA) and manually controlled 6-position injector valve (Vici, Valco Instruments, Houston, TX). Antibody stock solutions were vortexed for 30 seconds and centrifuged at 14,000rpm for 5min before preparing working dilutions to avoid precipitates.

All microfluidic cytology-on-a-chip assays contained Phalloidin and DAPI in the secondary antibody cocktail, but each was specific for a single molecular biomarker primary-secondary antibody pair. Working dilutions of antibodies were prepared in 0.1%PBSA with 0.1% Tween-20 (EMD Millipore, Billerica, MA, # 655206). Primary monoclonal antibodies were raised from either mouse (EGFR (Life Technologies, Carlsbad, CA, #MS-378-P, 10μg/mL) and β-catenin (BD Transduction Labs, San Jose, CA, #610154, 20μg/mL)), rabbit (avb6 (Abcam, Cambridge, MA, #Ab124968, 6μg/mL), Ki67 (Abcam #Ab15580, 29μg/mL), and MCM2 (Abcam #Ab108935, 10μg/mL)), or goat (CD-147 (EMMPRIN) (R&D Systems, Minneapolis, MN, #AF972, 20μg/mL). AlexaFluor-488 conjugated secondary antibodies were specific for F (ab')$_2$ fragments of mouse IgG (Life Technologies #A11017, 20μg/mL for EFGR and β-catenin), rabbit IgG (Life Technologies #A11070, 50μg/mL for avb6, 64μg/mL for Ki67, and 23.5μg/mL for MCM2), or goat IgG (Life Technologies #A11078, 40μg/mL for CD147). A working concentration of 0.33μM was used for Phalloidin-AlexaFluor-647 (Life Technologies #A22287) and 5μM for DAPI (Life Technologies #D3571).

In summary, the lab-on-a-chip sample processing was comprised of the following steps: 1) the device was primed with PBS at a flow rate of 735μL/min for 2 minutes, 2) the cell suspension in 20% glycerol/0.1% PBSA was delivered at 1.5mL/min for 2 minutes, 3) cells were washed with PBS at 1mL/min for 2.5min, 4) the primary antibody solution was delivered through a 0.2μm PVDF syringe filter at 250μL/min for 2.5min, 5) a wash step similar to step 3 was performed, 6) the secondary antibody solution was delivered under the same conditions as step 4, 7) a final wash step was performed, and 8) automated image capture was performed.

Sample Digitization

Images were recorded with a motorized reflected fluorescence microscope (Olympus BX-RFAA) equipped with a CCD camera (Hamamatsu ORCA-03G) through a 10x objective (10x/0.30NA UPlanFI, Olympus). A total of 25 unique fields of view (FOVs) repeated for 3 different z-focal planes were automatically captured across a 20mm^2 area using a robotic x-y-z microscope stage. Due to the complex three-dimensional morphology of oral squamous cells, multiple z-focal planes were captured and subsequently combined into a single, enhanced depth-of-field image to simplify the multi-spectral detection of the three fluorescent labels using the “stack focuser” macro built into ImageJ (http://rsb.info.nih.gov/ij/plugins/stack-focuser.html).

Combinations of custom macros and the open-source image analysis tools ImageJ [30] and Cell Profiler [11] were developed to automatically detect individual cells and define their nuclear and cytoplasmic boundaries as individual regions of interest (ROI). These ROIs were used to obtain intensity measurements associated with the three spectral channels and were used to define morphometric parameters. The DAPI and Phalloidin molecular labels served primarily to assist in the automated segmentation of individual nuclei and cytoplasm, respectively.

Random Forests

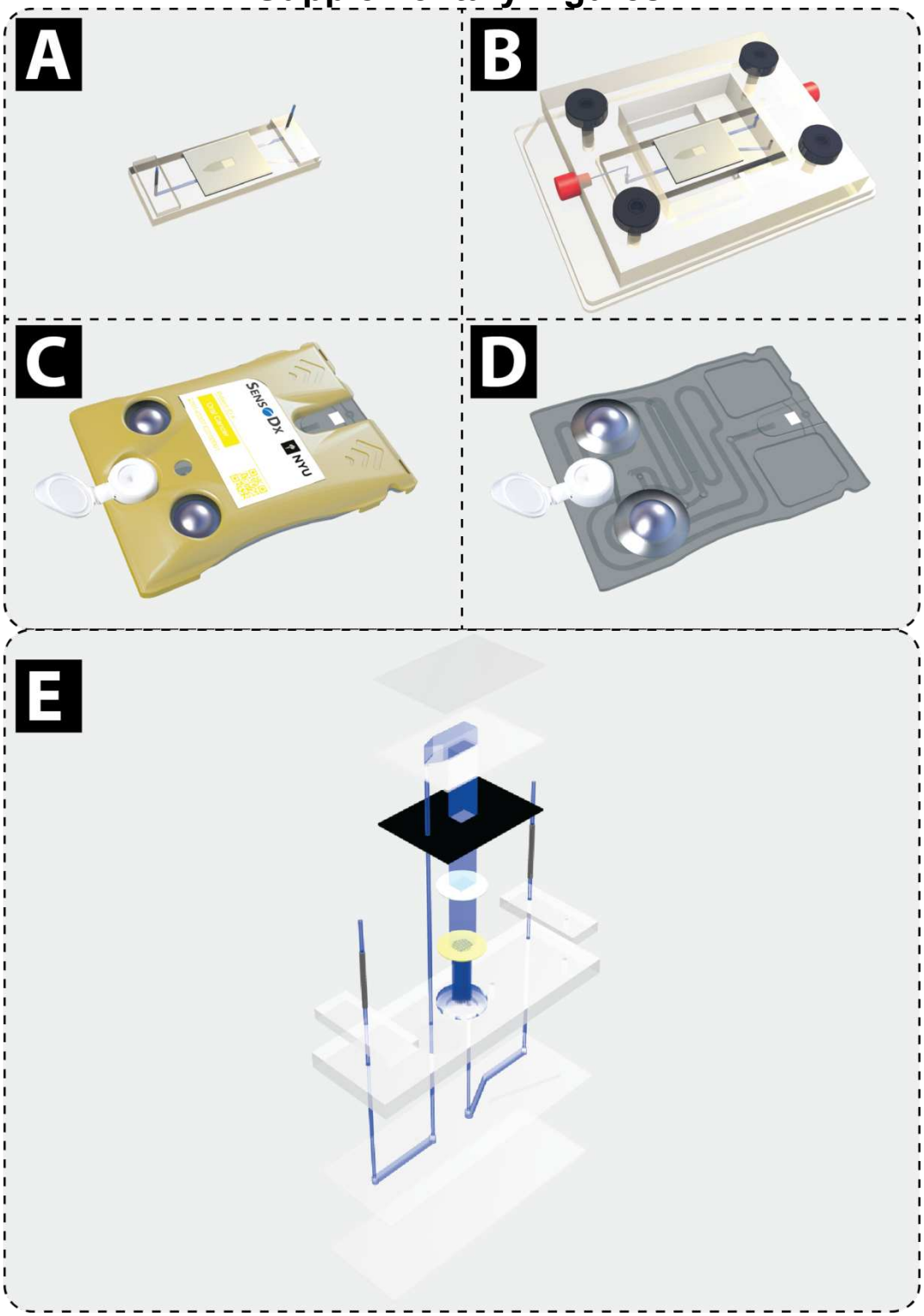
Random forests are an ensemble modeling approach composed of a collection of single decision trees and have been shown to be extremely resistant to outliers [31]. During model development, 1,000 randomly generated classification trees were fit to bootstrapped samples, with replacement. Each node of

the generated decision trees was fit to a randomly generated subset of 15 parameters. Initial random forest analyses used equal weights for “non-case” and “case” with cutoff points of 0.5 and 0.5, respectively. Additional weights of (0.6, 0.4), (0.65, 0.35), and (0.7, 0.3) were used to adjust the decision weighing between “non-case” and “case”, respectively. The sample weighting and choice of cutoff were manipulated in order to achieve approximately 90% and 85% sensitivities in the training models. The random models require that there are no missing independent variables, therefore two patients were excluded.

LASSO L1-Regularized-Logistic Regression

The LASSO methodology is a shrinkage and subset selection method that iteratively shrinks parameter effect sizes in order to help prevent against over-fitting. Model parameters identified by the LASSO methodology are the remaining parameters for each model after all other model parameter effects shrink to zero. Prior to LASSO model fitting, all parameters were standardized as z-scores. Model decisions were weighted to (2,1) for “non-case” and “case” respectively for the benign-mild dysplasia split and (3,1) for all others.

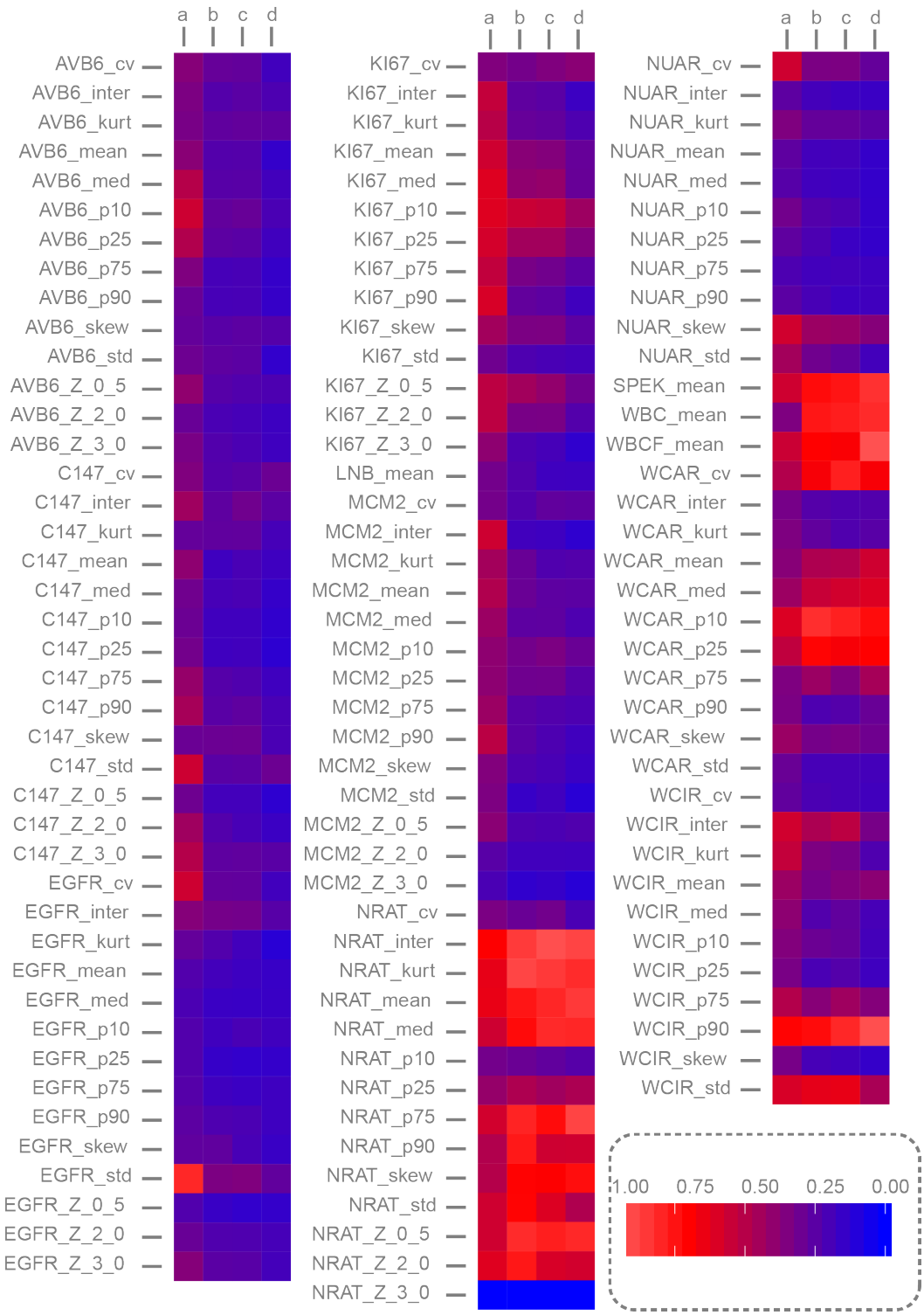
Supplementary Figures



Supplementary Figure 1. "Cytology-on-a-chip" device evolution.

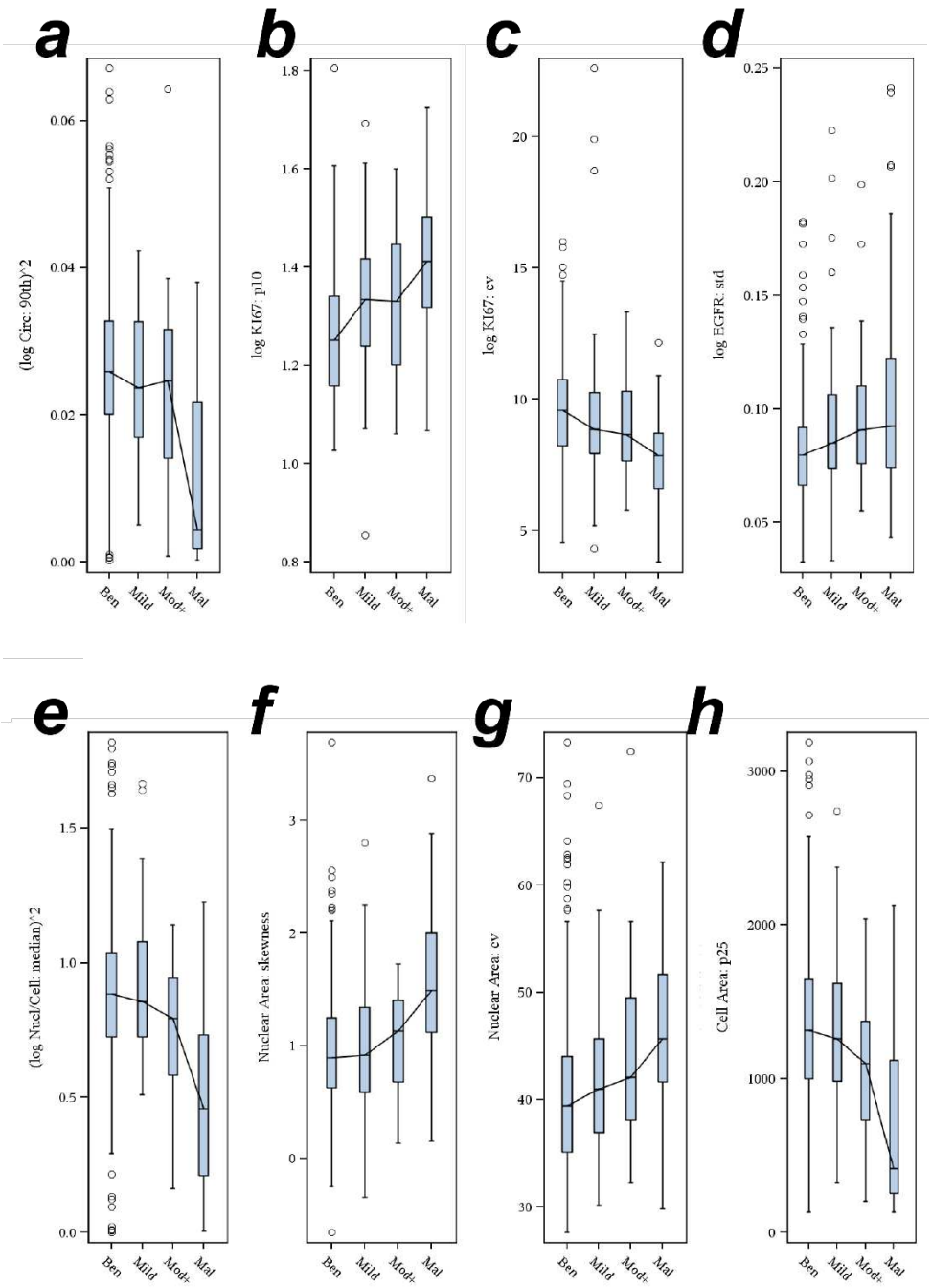
A) Rendering of original "flow cell" device used for cytology-on-a-chip measurements throughout the trial. The prototype device interfaced to external pumping and fluid handling equipment via two stainless-steel inlet and outlet ports. B) Rendering of flow cell

assembly with interchangeable fluid manifold. While not used for measurement collection in this trial, this improved chip-to-world interface has made subsequent assays more successful by reducing the occurrence of failures via inlet leaks. C) Rendering of form-factor “cartridge” device for next-generation cytology-on-a-chip sample processing (shown with protective shell). D) Rendering of same cartridge without shell to expose underlying network of microfluidic channels. While the ability to isolate and interrogate single cells across a micro-porous membrane is the same as the original flow cell (A), all assay operations following sample preparation (such as pumping, reagent filtering, and waste storage) are now contained within an integrated device. E) Exploded diagram of original flow cell (A), enlarged to show detail of various layers of the assembly. The fluid path from inlet, across the membrane, and through the outlet is highlighted in blue.



Supplementary Figure 2. Heatmap of random forest variable importance based on Gini index.

(Repetition of **Fig. 2*b***, but enlarged to show individual labels. Explanation of labels can be found in **Supplementary Table 3**).a = benign/mild-dysplasia, b = mild/moderate-dysplasia, c = low-risk/high-risk, d=moderate/severe-dysplasia.



Supplementary Figure 3. Box and whisker plots of selected variables identified in primary LASSO model.

These variables include a) cell circularity, 90th percentile (log-squared); b) Ki67, 10th percentile; c) Ki67, coefficient of variation (cv) (log-scale); d) EGFR expression, standard deviation (log-scale), e) NC ratio, median (log-squared); f) nuclear area, skewness; g) nuclear area, cv; and h) cell area, 25th percentile. Ben: benign, Mild: mild dysplasia, Mod+: moderate/severe/CIS dysplasia, Mal: malignant. The box bottom and top represent the 25th and 75th percentiles, respectively. Median values are connected between boxes, and whiskers down/up to 1.5 interquartile range.

Supplementary Table 1. Detailed description of the different morphometric and biomarker measurements collected in this trial.

	Parameter	Whole Cell	Nucleus
Image Intensity (Whole cell, Nucleus; R/G/B channels)	Mean intensity value	WCMean[red], [green]	NuMean[green], [blue] *Average value within the whole cell or nucleus selection. This is the sum of the intensity values of all the pixels in the selection divided by the number of pixels. [red] has QA/QC value and [blue] has limited descriptive value, whereas [green] is the most important for surface markers. For intracellular markers, the NuMean[green] is most descriptive.*
	Standard deviation of intensity	WCStdDev[red], [green]	NuStdDev[green], [blue] *Standard deviation of the intensity values used to generate the mean intensity value. [red] useful for Phalloidin, QA/QC and descriptive, [green] for surface markers.*
	Modal Value of intensity	WCMode[red], [green]	NuMode[green], [blue] *Most frequently occurring value within the selection. Corresponds to the highest peak in the histogram. Similar to Mean in terms of value.*
	Minimum and Maximum Intensity value	WCMin and WCMax[r/g/b]	NuMin and NuMax[r/g/b] *Minimum and maximum intensity values within the selection. Limited descriptive value, may be used for QA/QC.*
	Integrated density	WCIntDen[r/g/b]	NuIntDen[green], [blue] *Calculates and displays "IntDen" (the product of Area and Mean Gray Value) – Dependent values.*
	Median intensity value	WCMedian[red], [green]	NuMedian[green], [blue] *The median value of the pixels in the image or selection. This again is similar to Mean and Mode in terms of utility.*
Morphometric Parameters	Area	WCArea[red]	NuArea[blue] *Area of whole cell or nucleus selection in square pixels determined in red from Phalloidin stain.*
	Perimeter	WCPerim[red]	NuPerim[blue] *Length of profiled selection: whole cell or nucleus. Has value in QA/QC for evaluating profile.*
	Circularity	WCCirc[red]	NuCirc[blue] *A value of 1.0 indicates a perfect circle. As the value approaches 0.0, it indicates an increasingly elongated shape. Values may not be valid for very small particles. Calculated as 4π*area/perimeter^2*
	Feret Diameter	WCFeret[red]	WCFeret[blue] *The longest distance between any two points along the selection boundary, also known as maximum caliper.*
	Nuclear to Cytoplasmic (NC)-ratio	N_to_C_ratio *Critical value, calculated as ratio of NuArea / WCArea.*	

Supplementary Table 1 (Continued). Detailed description of the different morphometric and biomarker measurements collected in this trial.

Parameter	Whole Cell	Nucleus
Background Signal	Overall Background	Background[red], [green], [blue] *The mean intensity of the entire image – may inform about non-specific binding of antibodies on membrane, or presence of large artifacts*
	Signal to Background	StoBWCMed[r/g/b] StoBNucMed[r/g/b] *Measure of the signal to background ratio, with the signal measured as WCMed[r/g/b] or NuMed[r/g/b], and calculated as (signal – background)/background.*
		StoBWCAve[r/g/b] StoBNucAve[r/g/b] *Measure of the signal to background ratio, with the signal measured as WCMean[r/g/b] or NuMean[r/g/b], and calculated as (signal – background)/background.*

Supplementary Table 2. Lesion characteristics and Demographic risk factor parameters and definitions.

Lesion Characteristics	
Lesion with Diffuse Borders	
Lesion size (diameter along the long axis)	
Lesion size (diameter along short axis, i.e. perpendicular to long axis)	
Lesion area	
Lesion color red, or red/white, excluding lichen planus	
Lichen planus: binary measure completed by clinician at time of brush cytology sample collection where "Positive" indicated the presence of the clinical features of lichen planus, and "negative" indicated the absence of such features	
Demographic Risk Factors	
Short Reference	*Description*
Age	Age in years
Sex	Sex (Male:1; Female:0)
Alcohol: 12 drinks in life	Alcohol: 12 drinks in life
Alcohol: 12 drinks in past year	Alcohol: 12 drinks in past year
Alcohol days per year	Average days of alcohol use per year
Drinks per day	Average number of drinks on days drinking
Alcohol drink years	Alcohol drink years (average drinks per day times number of years drinking
>=100 Cigarettes in lifetime	At least 100 cigarettes in lifetime
Current cigarette smoker	Smoke Cigarettes now
Years smoking cigarettes	How many years smoked cigarettes
Cigarettes per day	Average cigarettes smoked per day
Tobacco Pack Years	Combined Cigarette, Cigar/Pipe Pack Years. Five cigars were considered equal 1 cigarette pack; and, 10 pipes were considered equal 1 cigarette pack. However since only combined pipe and cigar use was recorded: 7.5 cigars/pipes were set equal 1 cigarette pack (mean of cigar and pipe cigarette pack equivalents.
20 Cigars/ Pipes in life	At least 20 cigars or pipes in lifetime
Current cigar/pipe use	Smoke Pipes or Cigars Now
Years of Cigar/Pipe use	Years Smoke Pipes or Cigars
Cigars/Pipes per day	Average Smoke Pipes or Cigars per day
Cigar/Pipe years	Pipes or Cigars per day times years
Chew tobacco use in life	Ever use chewing tobacco or snuff
Current chew tobacco use	Use chewing tobacco or snuff currently
Years chewing tobacco or snuff	Years chewing tobacco or snuff
Containers chewing or snuff per week	Containers chewing or snuff per week
Chew Tobacco/snuff Container years	Container Years chewing/snuff per day (average containers per day time years

Supplementary Table 2 (Continued). Lesion characteristics and Demographic risk factor parameters and definitions.

Demographic Risk Factors	
Both high risk tobacco and alcohol use	Combined high risk alcohol and tobacco use (> 20 equivalent cigarette/cigar/pipe pack years and/or, if male: an average of 2 or more drinks per day, or if female: an average of 1 or more drinks per day).
>20 tobacco pack years	More than 20 cigarette pack years (or cigar/pipe pack year equivalents).

Supplementary Table 3. Abbreviations and Full names of biomarker labels from Supplementary Figure 2

Summary Statistic Measures	
Abbreviation	**Full name**
cv	Coefficient of variation
inter	Variance
kurt	Kurtosis
mean	Mean
med	Median
p10	10^{th} percentile
p25	25^{th} percentile
p50	50^{th} percentile
p75	75^{th} percentile
p90	90^{th} percentile
skew	Skewness
std	Standard deviation
Z_0_5	>0.5 Z-Score (percent of cells with biomarker values greater than 0.5 standard deviations away from healthy cells)
Z_2_0	>2.0 Z-Score (percent of cells with biomarker values greater than 2.0 standard deviations away from healthy cells)
Z_3_0	>3.0 Z-Score (percent of cells with biomarker values greater than 3.0 standard deviations away from healthy cells)
Biomarker Names	
Abbreviation	**Full name**
AVB6	αvβ6
C147	CD-147 (EMMPRIN)
EGFR	EGFR (epithelial growth factor receptor)
KI67	Ki-67
LNB	Number of lone nuclei
MCM2	MCM2
NRAT	Nuclear-to-Cytoplasmic area ratio
NUAR	Nuclear Area
SPEK	“Speckled cell” – unidentified intermediate cells smaller than WBCs
WBC	White blood cells (count)
WBCF	White blood cells (fraction of total cells)
WCAR	Whole cell area
WCIR	Whole cell circularity

SUPPLEMENTARY MATERIAL REFERENCES

1. Weigum SE, Floriano PN, Redding SW, et al. Nano-Bio-Chip sensor platform for examination of oral exfoliative cytology. *Cancer Prev Res (Phila Pa)*. 2010;3(4):518-528. doi:10.1158/1940-6207.CAPR-09-0139.

2. Li H-X, Zheng J-H, Fan H-X, Li H-P, Gao Z-X, Chen D. Expression of αvβ6 integrin and collagen fibre in oral squamous cell carcinoma: Association with clinical outcomes and prognostic implications. *J Oral Pathol Med*. 2013;42(7):547-556. doi:10.1111/jop.12044.

3. Ylipalosaari M, Thomas GJ, Nystrom M, et al. Alpha v beta 6 integrin down-regulates the MMP-13 expression in oral squamous cell carcinoma cells. *Exp Cell Res*. 2005;309(2):273-283. doi:10.1016/j.yexcr.2005.06.008.

4. Yu Y-H, Morales J, Feng L, Jack Lee J, El-Naggar AK, Vigneswaran N. CD147 and Ki-67 overexpression confers poor prognosis in squamous cell carcinoma of oral tongue: A tissue microarray study. 2014. http://www.scopus.com/inward/record.url?eid=2-s2.0-84924184522&partnerID=40&md5=bbef6ed26def20b8b6d49656651fce02.

5. Daniel F, Fava M, Hoffmann R, Campos M, Yurgel L. Main molecular markers of oral squamous cell carcinoma. *Appl Cancer Res*. 2010;30(3):279-288.

6. Williams GH, Romanowski P, Morris L, et al. Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. *Proc Natl Acad Sci U S A*. 1998;95(25):14932-14937. doi:10.1073/pnas.95.25.14932.

7. Scott IS, Odell E, Chatrath P, et al. A minimally invasive immunocytochemical approach to early detection of oral squamous cell carcinoma and dysplasia. *Br J Cancer*. 2006;94(8):1170-1175. doi:10.1038/sj.bjc.6603066.