Oral Oncology 92 (2019) 6–11

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Development of a cytology-based multivariate analytical risk index for oral cancer

Timothy J. Abram[a,1], Pierre N. Floriano[a,b], Robert James[c], A. Ross Kerr[d], Martin H. Thornhill[e], Spencer W. Redding[f], Nadarajah Vigneswaran[g], Rameez Raja[a], Michael P. McRae[a,h], John T. McDevitt[a,h,*]

[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] Department of Oral & Maxillofacial Medicine, Surgery and Pathology, School of Clinical Dentistry, University of Sheffield, Sheffield, UK
[f] The University of Texas Health Science Center at San Antonio, Department of Comprehensive Dentistry and Mays Cancer 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] Department of Biomaterials, Bioengineering Institute, New York University, 433 First Avenue, Room 820, New York, NY 10010-4086, USA

ARTICLE INFO

Keywords:
Oral cancer
Risk assessment
Multi-class classification
Cytology
Model ensembles

ABSTRACT

Objectives: The diagnosis and management of oral cavity cancers are often complicated by the uncertainty of which patients will undergo malignant transformation, obligating close surveillance over time. However, serial biopsies are undesirable, highly invasive, and subject to inherent issues with poor inter-pathologist agreement and unpredictability as a surrogate for malignant transformation and clinical outcomes. The goal of this study was to develop and evaluate a **M**ultivariate **A**nalytical **R**isk **I**ndex for **O**ral Cancer (MARIO) with potential to provide non-invasive, sensitive, and quantitative risk assessments for monitoring lesion progression.

Materials and methods: A series of predictive models were developed and validated using previously recorded single-cell data from oral cytology samples resulting in a "continuous risk score". Model development consisted of: (1) training base classification models for each diagnostic class pair, (2) pairwise coupling to obtain diagnostic class probabilities, and (3) a weighted aggregation resulting in a continuous MARIO.

Results and conclusions: Diagnostic accuracy based on optimized cut-points for the test dataset ranged from 76.0% for Benign, to 82.4% for Dysplastic, 89.6% for Malignant, and 97.6% for Normal controls for an overall MARIO accuracy of 72.8%. Furthermore, a strong positive relationship with diagnostic severity was demonstrated (Pearson's coefficient = 0.805 for test dataset) as well as the ability of the MARIO to respond to subtle changes in cell composition. The development of a continuous MARIO for PMOL is presented, resulting in a sensitive, accurate, and non-invasive method with potential for enabling monitoring disease progression, recurrence, and the need for therapeutic intervention of these lesions.

Introduction

While the incidence and mortality of many types of cancer have declined over the last 50 years, rates for oral and pharyngeal cancers (OPC), have seen remarkably little improvement during this same time. In the United States, the 5-year survival rate of patients diagnosed with OPC, is currently estimated at 63% [1]. Oral cavity cancers are predominantly epithelial-derived oral squamous cell carcinomas (OSCC), and the majority are preceded by "precancerous" lesions, though only a relatively small proportion (approximately 5%) will undergo malignant transformation. The term "potentially malignant oral lesion" (PMOL) describes an epithelial lesion encountered by a clinician, for which

Abbreviations: PMOL, Potentially Malignant Oral Lesion(s); OED, Oral Epithelial Dysplasia; OSCC, Oral Squamous Cell Carcinoma; AUC, Area Under the ROC (receiver-operator characteristic) Curve; FA, Fanconi Anemia; MSE, Mean Squared Error; MDLP, minimum description length principle

* Corresponding author at: Department of Biomaterials, Bioengineering Institute, New York University, 433 First Avenue, Room 820, New York, NY 10010-4086, USA.
E-mail addresses: robert_james@rhoworld.com (R. James), mcdevitt@nyu.edu (J.T. McDevitt).
[1] Current address: Velox Biosystems, Irvine, CA, USA.

https://doi.org/10.1016/j.oraloncology.2019.02.011
Received 12 July 2018; Received in revised form 18 December 2018; Accepted 11 February 2019
Available online 13 March 2019
1368-8375/

there is no obvious benign explanation, and which sufficiently elevates an index of suspicion to indicate biopsy to rule out malignancy or epithelial dysplasia. The spectrum of histopathology for PMOLs may be variable, from a carcinoma at one end of the spectrum, to carcinoma-in-situ (CIS), to various grades of oral epithelial dysplasia (OED - severe, moderate or mild), to benign diagnoses. In general, patients with PMOLs are referred to experienced clinicians for management in secondary care settings, such as head and neck surgeons, oral and maxillofacial surgeons, and oral medicine specialists.

The uncertainty of which patients or lesions will transform obligates such clinicians to serially monitor patients with PMOLs [2] and to biopsy lesions that have clinically evolved during these surveillance intervals. However, serial biopsies are undesirable, highly invasive, and histopathology suffers from inherent issues with poor inter-pathologist agreement and unpredictability as a surrogate for malignant transformation and clinical outcomes [3,4].

New approaches and new diagnostic/adjunctive tools that can provide non-invasive, sensitive, and quantitative risk assessments for monitoring the progression of patients with a history of oral cancer and/or PMOLs are urgently needed. Recent breakthroughs in artificial intelligence, machine vision, and deep learning and their broad application to fields involving human uncertainty and decision-making represent a frontier in diagnostic testing [5,6]. A continuous risk score derived from non-invasive cytopathological measurements has the potential to aid clinicians in PMOL surveillance by providing objective information that could be used for a multitude of applications including monitoring disease progression, indicating the need for re-biopsy, lesion mapping, and evaluating a response to chemoprevention or other therapeutics.

Historically, an attempt to create an index for oral dysplasia was made in 1969 by Smith and Pindborg by aggregating scores recorded for cellular and molecular features in histopathology slides on a 1–10 scale resulting in a final range from 0 to 75, known as the "epithelial dysplasia index" (EDI) [7]. When compared to the 5-stage Brothwell categorical grading system [8], the EDI was associated with a 47% inter-observer agreement rate compared to 62% for the Brothwell system [9]. The failed adoption of the EDI demonstrates the subjectivity among pathologists in diagnosing oral epithelial dysplasia (OED) [9] and emphasizes the need for risk metrics that are based on quantitative measurements free from human bias.

Using single-cell morphometric, nuclear and biomolecular measurements extracted from non-invasively collected PMOL cytology samples, this paper explores the development of a **M**ultivariate **A**nalytical **R**isk **I**ndex for **O**ral Cancer (MARIO) that is based on gold-standard histopathology with an added rigorous pathology adjudication process and can provide a non-invasive assessment of PMOLs in a surveillance setting.

Materials and methods

The data used in this study were obtained from a large "cytology-on-chip" database from patients with PMOLs paired with adjudicated gold-standard pathology diagnoses [10]. Single-cell data from a total of 506 patients were used to develop the MARIO, comprised of 125 Normal/Non-neoplastic, 253 Benign, 40 mild-Dysplastic, 13 moderate-Dysplastic, 10 severe-Dysplastic, and 65 Malignant lesions, according to the 2005 WHO definitions [10]. Because the present analysis included additional morphometric parameters that were not part of the original previous study, only subjects with complete data sets were enrolled in the present study (506 of 714). Details for the construction of the cytology database along with patient recruitment information can be found in Abram et al., 2016 [10]. Approximately 300 parameters describing the morphometric appearance and biomarker staining intensity profiles of each individual cell, recorded using the open source software, Cell Profiler [11], were stored in a MySQL database (Amazon Web Services RDS). All data analysis and model development was carried out in the statistical computing environment, R (v 3.2.1) [12]. While the original study was completed with two sets of statisticians in an infrastructure that supported completely blinded transfer and data analysis, the infrastructure was no longer available for this current study. However, in the spirit of avoiding data overfitting, the original raw data was randomized and a hold-out dataset was reserved using stratified random sampling to achieve equal class proportions between the test (25%) and the training (75%) datasets following data imputation, scaling, and centering to normalize the effect of different parameters.

Due to the low number of dysplastic cases commonly observed with prospective PMOL recruitment, "mild", "moderate", and "severe dysplasia" cases were grouped into one "Dysplastic" category resulting in four diagnostic classes with at least 50 patients each (Normal, Benign, Dysplastic, Malignant). The development of the MARIO involved the following steps: (1) training multiple base classification models for each class pair, (2) pairwise coupling to obtain class probabilities, and (3) a weighted aggregation to obtain a final risk score on a continuous scale. The overall methodology flow chart which summarizes the various contributions to the equations is provided Supplementary Fig. 1.

Due to the ability of the L1-regularized logistic regression (Lasso) methodology to prevent model over-fitting in high-dimensional data sets by iteratively shrinking parameter effect sizes, the Lasso technique was utilized in this study to build base classification models. Between the four diagnostic classes, a total of six base classification models were developed: Normal/Benign, Normal/Dysplastic, Normal/Malignant, Benign/Dysplastic, Benign/Malignant, and Dysplastic/Malignant. Overall error rate and mean square error (MSE) representing misclassifications for the six base classification models are shown in Supplementary Table 1.

The pairwise coupling method described by Price et al. (1995) [13] was used to convert base classification model scores into class probabilities for Normal, Benign, Dysplastic, and Malignant classes. Since probabilities calculated by this method do not necessarily sum to 1.0, the final class scores were normalized by dividing by each observation's sum of scores. This process resulted in a total of four probabilities per patient corresponding to the different class definitions (e.g. $P_{Normal} = 0.1$, $P_{Benign} = 0.8$, $P_{Dysplastic} = 0.1$, and $P_{Malignant} = 0.0$).

Finally, the MARIO was computed as a weighted average of the individual class probabilities. Class weights were implemented to account for the increase in severity in the ordered diagnoses from Normal through Benign, Dysplastic, and Malignant classes and were determined through an iterative process to maximize the performance of the MARIO. Final performance of the MARIO was evaluated both by its ability to correctly order cases along the diagnostic spectrum from low to high risk, and its ability to correctly predict diagnostic categories.

In order to add additional context to the MARIO, the frequency of five defined cell phenotypes was evaluated across three ranges of the MARIO score. The selected cell phenotypes included: (A) large, normal-appearing squamous cells; (B) medium-sized cells with enlarged nuclei (higher nuclear-cytoplasmic ratios); (C) small, highly circular cells; (D) lymphocytes (reflecting the inflammatory microenvironment); and (E) lone nuclei, a sign of high grade disease. Using a training set of these manually-identified cell phenotypes, "similar" cells were extracted from the entire single-cell database by iteratively computing a similarity matrix based on the Euclidean distance between objects in the single-cell database and the training set (*dist* function in the *proxy* [14] package in R). The cellular features used to compute the similarity matrix were restricted to 42 parameters that covered cellular and nuclear morphometry and staining intensity. A manual cutoff for the "similarity range" was tuned to confirm the integrity of returned cells by only selecting objects with an arbitrary distance less than 50 to the target cell. Cell phenotypes were aggregated into frequency plots for three ranges of the MARIO covering low (0–30), medium (30–60), and high (60–100) risk scores.

Furthermore, the ability of the MARIO to respond to subtle changes

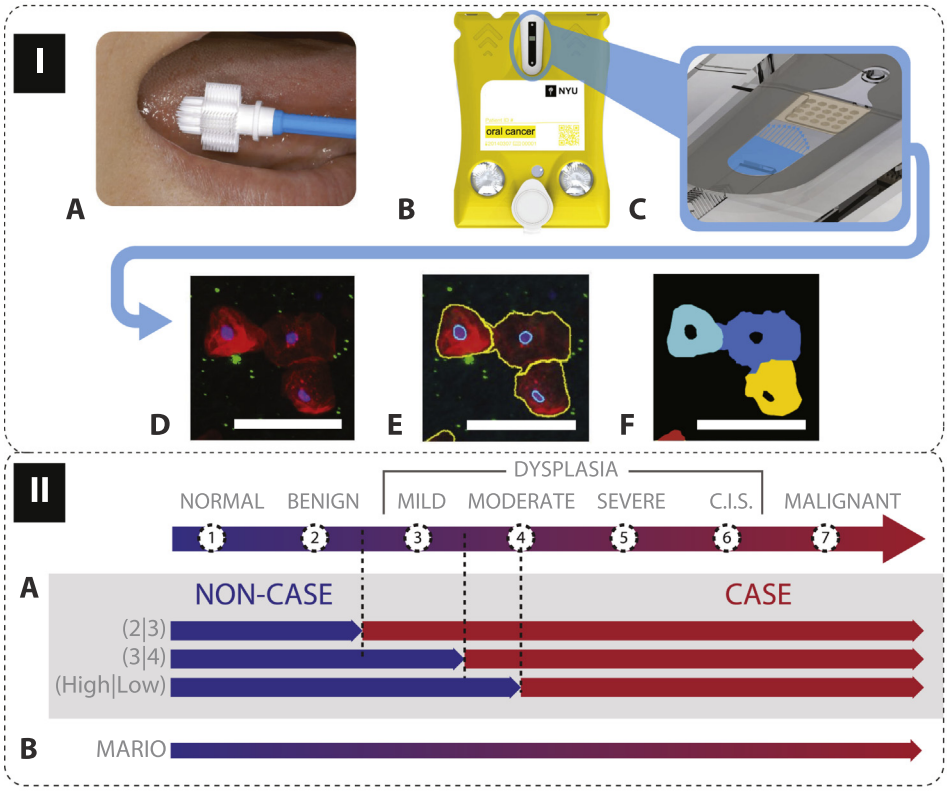

Fig. 1. Risk stratification diagram and "Cytology-on-chip" workflow. Panel I: (A) Suspicious lesion is sampled via "brush sample" technique, (B/C) single cells are captured on a nano-porous membrane embedded within a microfluidic channel, (D) multispectral fluorescence images are recorded across a raster-scan of the membrane, (E) algorithms identify cellular boundaries based on signal contrast, and (F) regions of interest (ROIs) are extracted for quantification (scale bar = 100 μm). Panel II: The 7-stage diagnostic spectrum proposed by the 2005 WHO guidelines displayed as a continuous number line. (A) Binary risk assessment in the primary clinical setting scenario, where the main goal is to refer suspicious lesions for biopsy [10], (B) Continuous score of the MARIO.

Table 1
MARIO performance of the training and test datasets.

	Training Set (75% of data)				Test Set (25% of data)			
	Max Score Vote (Accuracy, %)			79.80%	79.02%			
	Class Binning (Accuracy, %)			74.54%	72.80%			
	Vote Order (Spearman's rho)			0.876	0.886			
	Class Order (Spearman's rho)			0.827	0.805			
	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Normal	97.37	100.00	96.52	0.983	97.60	100.00	96.81	0.984
Benign	78.74	59.47	97.91	0.787	76.00	55.56	96.78	0.762
Dysplastic	82.68	72.92	84.01	0.785	82.40	73.33	83.64	0.785
Malignant	90.29	85.71	90.96	0.883	89.60	87.50	89.91	0.887

at the single cell level was evaluated by simulating changes in cytology sample composition. In this simulation, 10 randomly selected normal patients with no clinical lesions had a percentage of archived cells from their cytology samples randomly exchanged with cells from a pooled set of 25 OSCC patient cytology samples, ranging from 1% to 90%. Median MARIO values were recorded across 10 replicates for each percentage per patient.

Results

Using data acquired from a large "cytology-on-chip" database, a simple, continuous **M**ultivariate **A**nalytical **R**isk **I**ndex for **O**ral Cancer was developed. In this previous study, data was collected according to the following summary: Patients presenting with one or more suspicious lesions received a non-invasive "brush sample" of the lesion surface (Fig. 1.Ia); layers of cells collected from the brush were vortexed and processed through a microfluidic cartridge (Fig. 1.Ib) containing an embedded nano-porous membrane (Fig. 1.Ic) which captured and isolated single cells for fluorescent biomarker labeling; single-cell data with approximately 300 cytomorphometric parameters were extracted from multispectral fluorescence images of the membrane surface (Fig. 1.Id), after automated image analysis algorithms identified cellular boundaries (Fig. 1.Ie) and their corresponding regions (Fig. 1.If). While this previous study focused on the parameter selection of various binary models (Fig. 1.IIa), the MARIO developed in this study exists on a continuous-scale similar to the theoretical diagnostic spectrum.

Of the six base classification models, the Lasso models differentiating classes with greater separation along the diagnostic spectrum outperformed those discerning classes with less separation (e.g. Benign / Malignant MSE = 0.2049 (± 0.86), Dysplastic/ Malignant MSE = 1.8852 (± 4.75), Supplementary Table 1). Following pairwise coupling, within the model training dataset, the overall classification accuracy based on majority class probabilities was 75%. Of the 25% misclassified observations, 71% achieved the correct classification by their second greatest class probability. Therefore, 93% of the true classification labels can be explained with the information contained in the two largest class probabilities. Final performance values for the MARIO are shown in Table 1. Performance of the continuous MARIO was evaluated by its ability to correctly rank class labels in order of increasing risk on the diagnostic spectrum and its ability to correctly predict diagnostic class categories.

The ability of the MARIO to correctly order all observations was evaluated by Spearman's rank correlation coefficient on a scale from

−1 to 1 where 1 implies perfect ordering in a monotonic, but not necessarily linear, increasing relationship. Spearman's coefficient value for the rank-order of class labels was 0.827 for the training dataset and 0.805 for the test dataset, demonstrating a strong positive relationship between the MARIO and position along the diagnostic spectrum. The ability of the MARIO to correctly predict diagnostic class categories was evaluated by discretizing the continuous score back into four distinct categories. Optimal cut-points were determined by minimizing class entropy by the minimum description length principle (MDLP) [15]. The process of aggregating discrete class probabilities into a continuous variable was associated with a minor decrease in overall accuracy of 6.4% in the test dataset (from 79.2% to 72.8%). In terms of diagnostic performance, lesions at the low and high end of the disease spectrum were more likely to be correctly classified by the MARIO than intermediate dysplastic cases. For example, Normal and Malignant cases resulted in MARIO sensitivity/specificity values of 100%/96.8% and 87.5%/89.9% in the test dataset, respectively while Benign and Dysplastic cases resulted in sensitivity/specificity values of 55.6%/96.8% and 73.3%/83.6% for the test dataset, respectively.

After performance values for the training and test datasets were recorded, the MARIO model was applied to all patients and plotted against their gold-standard diagnoses (Fig. 2). Boxplots of patient risk scores demonstrated a strong positive relationship against the original patient diagnoses based on the 2005 WHO criteria [16]. Average risk scores for confirmed normal (n = 125) and OSCC (n = 65) lesions were significantly different from all other categories (unpaired *t*-test,

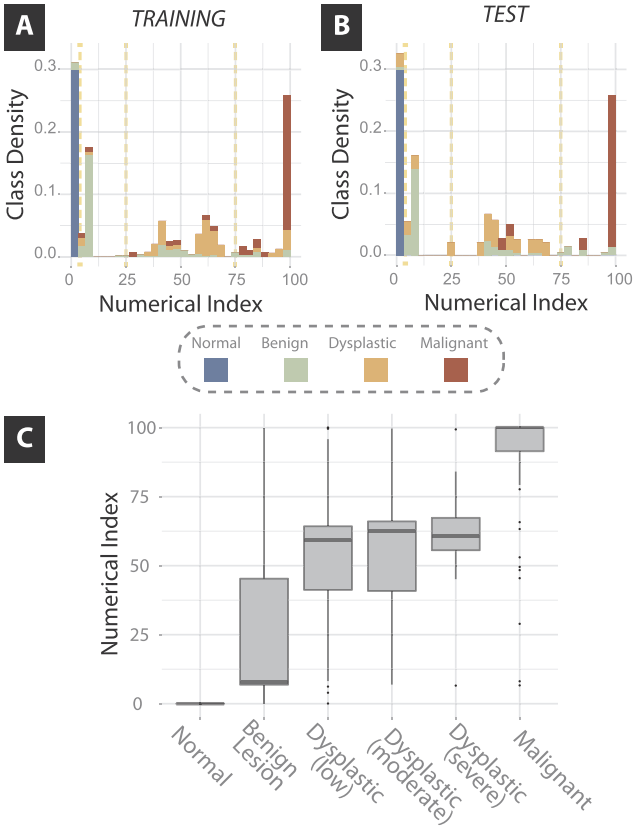

Fig. 2. MARIO Performance. Barplots scaled to class densities (y-axis) across the MARIO (x-axis) for (A) Training and (B) Test datasets. Color-coding represents the true patient class according to histopathology. Vertical dashed lines represent optimal cut-points for discretizing the continuous score into disease class domains based on minimizing class entropy in the training dataset. (C) Box-plot of MARIO values (y-axis) across the 6 different diagnostic categories (center line = median value, top/bottom box = inter-quartile range). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

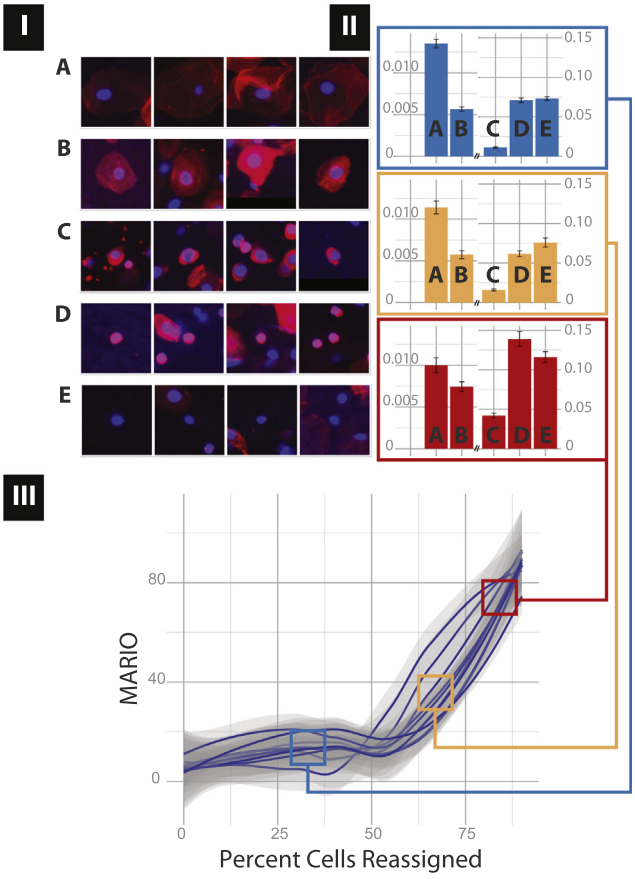

Fig. 3. Results from single-cell calibration exercise and phenotype query. Panel I) Representative images of cell phenotypes for frequency tables in Panel II: (A) Medium-sized rounded cell with enlarged nuclei, (B) Large, normal-appearing squamous cell, (C) Small, highly-circular cell, (D) Leukocyte, (E) lone nuclei. Panel II) Distribution of phenotype frequencies for patients with the identified range of risk scores (Blue: 0–25, Orange: 30–60, Red: 75–100). Left axis = A, B; Right axis = C, D, E. Error bars = standard deviation of phenotype frequency per patient. Panel III: Results from cell reassignment simulation where solid lines represent median MARIO values for each of 10 randomly selected healthy volunteer samples across the increasing percentage of their cells exchanged for cells from a corpus of OSCC patient cells (x-axis). Gray boundaries surrounding each line represents ± standard error across 10 replicates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$p < 0.0001$). Average scores for patients with benign (n = 253), mild (n = 40), moderate (n = 13), and severe dysplasia (n = 10) were not significantly discernable from each other (unpaired t-tests, p = 0.0758–0.5619). At the 4-class level, where dysplasia cases were grouped into a single category, risk score means for each group were significantly different except between patients in Benign (n = 253) and Dysplastic (n = 63) classes (unpaired *t*-test, 306 df, p = 0.2297). Cut-points for the MARIO on a scale from 0 to 100 were defined at positions 4, 25, and 75, resulting in classification intervals of 0–4 for Normal, 4–25 for Benign, 25–75 for Dysplastic, and 75–100 for Malignant. The large score range for the dysplastic category agrees well with conventional OED grading [16] given the aggregation of low, moderate, and severe/CIS dysplastic cases into a single category which in reality represent a range of disease severity and an accumulation of genetic alterations.

Frequencies of different cellular phenotypes across three regions of the MARIO spectrum are shown in Fig. 3 panel II. Progression of the MARIO from 0 to 100 was associated with a change in the distribution of the five cell phenotype categories. Phenotypes B (Medium-sized rounded cells with enlarged nuclei), C (Small, highly-circular cells), and

D (lymphocytes) displayed the greatest increase in frequency for increasing MARIO values. Notably, the greatest increase in phenotype frequency occurred between the medium and high-risk score ranges, exemplified by the 4x increase in frequency of phenotype C and the almost 2x increase of phenotype D. Lymphocyte frequency (phenotype D) in patient samples for medium range risk scores was lower than both low and high range scores. While most of the five phenotypes increased in frequency at higher risk score values, the large, normal-appearing squamous cells (phenotype B) dropped in frequency approximately 30%. This ability to summarize the frequency of various cell phenotypes on an individual patient level has the potential to enhance molecular-level insights behind carcinogenesis and the malignant transformation of PMOLs.

In addition to phenotype frequency, a simulation was conducted to evaluate the response of the MARIO to subtle changes at the single cell level within patient cytology samples (Fig. 3, panel III). Each of the 10 lesion-free patients randomly selected in this exercise are represented as a line where the median calculated MARIO across 10 iterations is plotted against the percentage of reassigned cells. All patients exhibited a marked increase in MARIO score shortly after more than half of the patients' cells were replaced with cells from pooled OSCC samples. This simulation also indicates that the calculation of the MARIO is sensitive enough to track changes in a cytology sample when only a small percentage of cells display a morphometric or biomarker expression alteration.

Discussion

When evaluating patients with PMOLs, pathologists face the challenge of imposing artificial categories on a continuous biological process which can result in potential bias and missed opportunities to signify cases that may not conform to a particular classification. Though these categories were required to train this MARIO, we hypothesize that multiple binary classification models along the diagnostic spectrum can be combined into a single continuous measure to enable clinicians to monitor PMOLs in more detail. Unlike existing adjuncts such as vital dyes, chemiluminescence, and tissue loss of fluorescence, a continuous risk score for PMOLs computed from non-invasive cellular measurements could be leveraged for a multitude of applications including monitoring patient response to treatment, determining when to refer a patient for initial biopsy or re-biopsy, mapping the oral mucosa for signs of recurrence, evaluating novel therapeutics and devices in clinical trials, or determining when to initiate more aggressive surgical intervention.

While achieving an overall accuracy of 72.8% that rivals expert pathologist accuracy of 69% in classifying PMOLs [17], the MARIO developed here demonstrated variable performance depending on the region of the diagnostic spectrum. As desired, the ability to accurately distinguish normal mucosa from reactive benign conditions and various grades of epithelial dysplasia resulted in excellent performance (sensitivity = 96.5%, specificity = 96.8%, test dataset). At the other end of the diagnostic spectrum, the MARIO was accurately able to distinguish malignant lesions from non-malignant lesions in 89.6% of the cases in the test dataset.

The lower sensitivity of Benign classification compared to other categories is due in part to the challenge of distinguishing particular benign inflammatory and ulcerative lesions with reactive epithelial atypia inflammatory conditions from chronic conditions and the tumor microenvironment seen in dysplastic or malignant lesions. Additionally, the subtle cellular changes that underscore the grading criteria for distinguishing benign lesions from mild dysplastic lesions have confounded pathologists in addition to computer algorithms, where lower inter-observer agreement and accuracy have been reported for low grade dysplastic lesions compared to high-grade disease [18]. One limitation of this study is the low number of patients in the intermediate range of the dysplastic spectrum; future studies will seek to enrich the recruitment of patients within this range to develop models that can more accurately differentiate the subtle grades of mild, moderate, and severe dysplasia. Furthermore, additional molecular biomarkers and morphometric features may be incorporated to achieve finer molecular-level detail that reflects the altered genetic state of these low-grade lesions.

As a supplement to the visual interpretation of the raw fluorescence cytology images, the frequency of five different cell phenotypes was automatically measured across low, medium, and high-risk score values. The resulting cell frequency trends agree well with conventional assessment of oral cytology and histology across the diagnostic spectrum. The increase in frequency of small, highly circular cells and cells with enlarged nuclei has been observed previously in the literature and has been attributed to the decrease in the amount of cytoplasm and reduction in the degree of cellular cohesion that occurs with increased dysplasia grade [19]. The high frequency of lymphocytes in the low risk score range is likely due to the presence of inflammatory conditions and ulcerative lesions among patients with benign conditions such as ulcerative and erosive lichen planus. By interactively selecting target cells and tuning the "similarity range", future studies could employ this novel cell query approach to investigate or label unique cell phenotypes in an effort to uncover molecular-level insights into PMOL progression.

The ability to condense the complex biological information contained in a PMOL cytology sample into a single MARIO has the potential to lead to unique insights surrounding the molecular phenomena associated with lesion progression. While prospectively recruited PMOL samples better reflect the true heterogeneity of the real-world patient population, they do not offer information with regard to individual lesion progression over time. However, the response of the MARIO to cytology sample alterations at the single-cell level can be simulated by randomly exchanging cells between patients with different initial diagnoses. By iterating this simulation for a sample of healthy volunteers and OSCC patients, we have observed that the MARIO is capable of tracking changes in a cytology sample when only a small percentage of cells display a morphometric or biomarker expression alteration.

The MARIO presented here was recently used in a retrospective study involving cytology-on-chip measurements from a high-risk population of patients with Fanconi Anemia (FA), a rare inherited chromosomal instability disorder associated with an 800-times higher risk of developing oral cancer than the general population [20,21]. Cytology-on-chip samples from 37 FA patients over two recruitment cohorts in the 2014 and 2016 Meeting for Adults with FA were used to compute their MARIO scores, including six patients who presented at both meetings, representing a monitoring period of two years. Of these six patients, MARIO scores increased significantly ($+ > 40\%$) for two patients, suggesting progression; decreased ($- > 40\%$) for two patients, suggesting non-progression or regression; and remained unchanged for the remaining two patients ($\pm 5\%$), indicating non-progression or stable disease [22]. One of the first two patients demonstrated a 27-fold increase in their MARIO, self-reported a recent pharyngeal cancer diagnosis within the 2-year period, and was notified to seek additional guidance from the patient's head and neck surgeon. This small pilot, proof-of concept study demonstrated that the MARIO score has potential to predict accurately the malignancy risk of PMOL in two patients as confirmed by follow-up scalpel biopsy, though significant further evaluation is needed to validate the MARIO for surveillance applications.

Conclusion

We have demonstrated that a multi-class classifier can be adapted to achieve a single, continuous "severity score" while achieving a performance that rivals expert pathologists (72.8% overall accuracy, individual class prediction accuracy from 76.0% to 97.6%) to provide clinicians with a metric to track PMOL progression at the patient level using non-invasive sampling. Because the MARIO doesn't require the

interpretation of an expert user, one possible application could involve the rapid identification and risk screening of lesions in an outpatient setting, such as a dentist office. While the application to oral medicine and the assessment of PMOL have been emphasized, this methodology could be generalized to other medical applications. Future efforts will focus on evaluating this MARIO for monitoring patients over time. Based on robust performance agreement between the model development and test datasets (± 1.7% in terms of accuracy), these methods are anticipated to generalize well with future collected data. Further optimization of these models is possible with future studies involving a larger number of patients with dysplastic lesions. An intriguing, bold visualization of a single metric representing thousands of single-cell morphometric and molecular attributes has been demonstrated with potential to efficiently assess the degree of progression or, even regression of PMOL under surveillance.

Disclosures

Principal Investigator, John T. McDevitt, has a financial 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.

Acknowledgements

The authors would like to thank the University of Texas Health Science Center at San Antonio (UTHSCSA) (Stephanie Rowan, Chih-Ko Yeh, Stan McGuff, Frank Miller), University of Texas Health Science Center at Houston (UTHSCH) (Jerry Bouquot, Nagi Demian, Etan Weinstock, Nancy Bass), New York University/Bluestone Center for Clinical Research (Joan Phelan, Patricia Corby, Ismael Khouly), Sheffield Teaching Hospitals NHS Foundation Trust and the University of Sheffield (Paul Speight, Craig Murdoch, Christine Freeman, Anne Hegarty, Katy D'Apice) for assistance in obtaining clinical samples. The authors would also like to thank Rho, Inc. (Chapel Hill, North Carolina) (Julie Vick) for providing assistance with patient data management and statistical analysis.

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. The work performed on Fanconi Anemia patient samples was supported by the Cancer Prevention Research Institute of Texas (Award Numbers RP140024-P1 (JTM) and RP101382 (JTM/NV)) as well as the Fanconi Anemia Research Fund.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oraloncology.2019.02.011.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin 2017;67:7–30.
[2] Al-Dakkak I. Oral dysplasia and risk of progression to cancer. Evid Based Dent 2010;11:91–2.
[3] Bosman FT. Dysplasia classification: pathology in disgrace? J Pathol 2001;194:143–4.
[4] 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.
[5] Granter Scott R, Beck Andrew H, Papke Jr. David J. AlphaGo, deep learning, and the future of the human microscopist. Arch Pathol Lab Med 2017;141(5):619–21.
[6] Bejnordi BE, Linz J, Glass B, Mullooly M, Gierach GL, Sherman ME, et al. Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images. ArXiv Prepr 2017. ArXiv170205803.
[7] Katz HC, Shear M, Altini M. A critical evaluation of epithelial dysplasia in oral mucosal lesions using the Smith-Pindborg method of standardization. J Oral Pathol 1985;14:476–82.
[8] Brothwell DJ, Lewis DW, Bradley G, Leong I, Jordan RCK, Mock D, et al. Observer agreement in the grading of oral epithelial dysplasia. Community Dent Oral Epidemiol 2003;31:300–5.
[9] Manchanda A, Shetty D-C. Reproducibility of grading systems in oral epithelial dysplasia. Med Oral Patol Oral Cir Bucal 2012;17:e935–42.
[10] Abram TJ, Floriano PN, Christodoulides N, James R, Kerr AR, Thornhill MH, et al. "Cytology-on-a-chip" based sensors for monitoring of potentially malignant oral lesions. Oral Oncol 2016;60:103–11.
[11] Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. Cell Profiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 2006;7:R100.
[12] R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available from: https://www.R-project.org.
[13] Price D, Knerr S, Personnaz L, Dreyfus G. Pairwise Neural Network Classifiers with Probabilistic Outputs; 1995. p. 1109–16.
[14] Meyer D, Buchta C. proxy: Distance and Similarity Measures. R package version 0.4-15 [Internet]. 2015. Available from: http://CRAN.R-project.org/package=proxy.
[15] Fayyad U, Irani K. Multi-interval discretization of continuous-valued attributes for classification learning. IJCAI 1993:1022–9.
[16] 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.
[17] Speight PM, Abram TJ, Floriano PN, James R, Vick J, Thornhill MH, et al. Interobserver agreement in dysplasia grading: towards an enhanced gold standard for clinical pathology trials. Oral Surg Oral Med Oral Pathol Oral Radiol 2015;120:474–82.
[18] 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.
[19] 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.
[20] Kutler DI, Singh B, Satagopan J, Batish SD, Berwick M, Giampietro PF, et al. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). Blood 2003;101:1249–56.
[21] Masserot C, Peffault de Latour R, Rocha V, Leblanc T, Rigolet A, Pascal F, et al. Head and neck squamous cell carcinoma in 13 patients with Fanconi anemia after hematopoietic stem cell transplantation. Cancer 2008;113:3315–22.
[22] Abram TJ, Pickering CR, Lang AK, Bass NE, Raja R, Meena C, et al. Risk stratification of oral potentially malignant disorders in Fanconi Anemia patients using autofluorescence imaging and cytology-on-A chip assay. Transl Oncol 2018;11:477–86.