

Early detection of lung cancer using artificial intelligence-enhanced optical nanosensing of chromatin alterations in field carcinogenesis

Ali Daneshkhah^{1*}, Sravya Prabhala^{1*}, Parvathi Viswanathan^{2*}, Hariharan Subramanian^{1,2}, Jianan Lin², Andrew Chang¹, Ankit Bharat⁴, Hemant Kumar Roy³, Vadim Backman¹

¹ Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA

² Nanocytomic LCC, Northwestern University, Evanston, IL, USA

³ Boston Medical Center, Boston, MA, USA

⁴ Canning Thoracic Institute, Department of Surgery, Northwestern University, Feinberg School of Medicine, 420 East Superior Street, Chicago, Illinois 60611

*Authors Contributed equally to the manuscript

Abstract

Supranucleosomal chromatin structure, including chromatin domain conformation, is involved in the regulation of gene expression and its dysregulation has been associated with carcinogenesis. Prior studies have shown that cells in the buccal mucosa carry a molecular signature of lung cancer among the cigarette-smoking population, the phenomenon known as field carcinogenesis or field of injury. Thus, we hypothesized that chromatin structural changes in buccal mucosa can be predictive of lung cancer. However, the small size of the chromatin chain (approximately 20 nm) folded into chromatin packing domains, themselves typically below 300 nm in diameter, preclude the detection of the alterations in intradomain chromatin conformation using diffraction-limited optical microscopy. In this study, we developed an optical spectroscopic statistical nanosensing technique to detect chromatin packing domain changes in buccal mucosa as a lung cancer biomarker, chromatin-sensitive partial wave spectroscopic microscopy (csPWS). Artificial intelligence (AI) was applied to the csPWS measurements of chromatin alterations to enhance diagnostic performance. Our AI-enabled buccal csPWS nanocytology of 179 patients at two clinical sites distinguished Stage-I lung cancer versus cancer-free controls with an area under the ROC curve (AUC) of 0.93 ± 0.06 for Site 1 (in-state location) and 0.82 ± 0.11 for Site 2 (out-of-state location).

Introduction

Cancer screening tests should ideally identify cancer before symptoms have appeared and while the tumor is small to effectively increase the chance of treatment and reduce the mortality rate. Lung cancer is the leading cause of cancer deaths across races and genders in the U.S. with an overall five-year survival rate of 22.9% which is notably lower than colorectal (65.1%), breast (90.6%), and prostate cancers (96.8%)¹. However, if lung cancer is detected at an early stage, it is highly curable through surgical resection. The five-year survival rate for late-stage (distant) non-small lung cancer (NSLC) is less than 8% but increases to 64% if detected at a localized stage and reaches 90% if detected at Stage-IA². Low-dose computed tomography (LDCT) has been established as the gold standard for lung cancer screening and is associated with a 20% decrease in mortality among patients screened with the technique. Accessibility, cost, stigma, and lack of adherence to LDCT guidelines are among the major challenges limiting its impact; unfortunately, only about 5% of the LDCT-eligible population undergoes screening³, resulting in 55% of lung cancer cases being detected at an advanced stage where the survival rate is below 8%⁴. We propose to utilize a minimally invasive accessible, sensitive, and accurate pre-screening test for LDCT with high sensitivity (Se) to early-stage lung cancer to effectively implement the LDCT procedure on a broader scale and enhance the efficacy of lung cancer screening.

Screening methods other than LDCT such as chest X-rays and sputum cytology have proven unsatisfactory when evaluated in large-scale clinical screening settings⁵. New methods based on standard protein biomarkers used for the detection of cancer do not provide sufficient sensitivity and specificity (Sp)⁶. Recently, there has been significant interest in the development of protocols that rely on tumor secretions in the blood, such as liquid biopsy. Tests being developed by companies including Grail, Freenome, Guardant, Delfi, and Thrive identify cancer by analyzing circulating tumor DNA (ctDNA) or tumor-derived circulating free DNA(cfDNA) properties such as gene mutations, methylation, and fragmentation⁷⁻¹¹. Although initial results have shown promise in the detection of various cancers, including lung cancer, the sensitivity to Stage-I and smaller lesions drops precipitously below a clinically acceptable level. It has been suggested that this limitation is not primarily due to a technology limitation but might be related to the biology of the source and type of biomarker. Smaller lesions secrete less tumor ctDNA (~1 ctDNA/ 10mL of blood), while tumor heterogeneity can only be modeled through many tumor-byproduct biomarkers, which makes it challenging to find the needed quantities of ctDNA in a clinically practical blood sample¹². For example, the overall sensitivity of the Grail multi-cancer early detection (MCED) test may drop from 90.1% [95% confidence interval (CI): 87.5% to 92.2%] in Stage-IV patients to 16.8% [95% CI: 14.5% to 19.5%] in Stage-I patients¹³. Liquid biopsy can be a powerful tool for non-screenable cancers (pancreatic, etc.) but for cancers with established screening protocols, such as colorectal and lung, methods to detect treatable early-stage lesions are still urgently needed. To address these issues and develop an effective pre-screening test for LDCT, we optimized three crucial aspects 1) biomarker source, 2) biomarker type, and 3) enabling technology.

An ideal biomarker source for the development of a large-scale prescreening test should be obtained in a minimally invasive procedure from an accessible specimen, with an easy-to-implement and reproducible protocol, and provide high sensitivity to small treatable lesions¹⁴. Our approach to finding this biomarker source relies upon the application of a well-established phenomenon, field carcinogenesis, also known as field effect, or field of injury, which was first introduced six decades ago¹⁵. According to field carcinogenesis, the genetic/epigenetic alterations

leading to neoplastic cell transformation are distributed diffusely throughout the “field of injury” at even the pre-malignant stage¹⁵⁻²³. For example, cells across the aero-digestive mucosa (field of injury) are exposed to carcinogens in tobacco and accumulate genetic changes among smokers^{21,22}. Due to the stochastic nature of these molecular changes, some cells may eventually give rise to a tumor clone. Thus, cells throughout the entire aero-digestive mucosa harbor biomarkers of carcinogenesis regardless of their proximity to a lung tumor. The buccal mucosa is widely recognized as a “molecular mirror” for lung cancer because of field carcinogenesis^{16,21,23,24} and we considered it as our biomarker source for two reasons. First, buccal brushings are easily performed and uniquely suited for an at-home test or primary care office, as opposed to “liquid biopsies” that can hardly be self-administered. Next, the biomarkers of field carcinogenesis are sensitive to early (e.g., Stage-I) cancers, regardless of tumor size, which is diagnostically superior to other sources such as blood or breath, as the latter depends on a load of secretions by the tumor and thus are more sensitive to large tumors than small ones.

Discovering a suitable lung cancer biomarker type from buccal mucosa is the next major task. Biomarkers obtained from genetic changes are negatively impacted by the extremely high number of genetic alterations and astonishing tumor heterogeneity which hampers the application of downstream biomarkers for the detection of small lesions. The highly dynamic chromatin structure is known as a substrate that regulates gene expression by adjusting the accessibility to transcription factors (TF) and RNA polymerases (RNAPs)^{25,26}, and thereby can be used to possibly predict the risk of cancer at an early stage²⁵⁻²⁷.

To understand what types of chromatin structure may foster carcinogenesis, we first needed to calculate a quantifiable metric of chromatin structure. We and others have reported that chromatin is organized as a variety of packing domains²⁸⁻³⁰. Packing domains represent a unique chromatin structure that can regulate gene accessibility. At the smallest length scale, DNA wraps around a histone and forms ~11 nm nucleosome complexes of “beads on a string” which are further folded into the curvilinear chromatin chain, between 5 and 24 nm³¹. These chromatin chains are packed together in various structural compactions and densities forming irregular blocks of larger packing domains. The packing domains have heterogeneous morphological properties with an average radius of 80 nm and genomic size of about 200 kbp and within these domains, chromatin shows a polymeric fractal-like behavior (i.e. the mass scaling behavior within domains follows a near-power-law relationship) along with radially decreasing mass density from the center to the periphery³². Chromatin packing scaling D is defined by estimating the number of base pairs (N) scaling with the radius of the occupied volume (R) as $N \propto R^D$. The value of D falls between 1.8 and 2.9 across packing domains²⁹. A higher D value may indicate a packing domain with an increased chromatin heterogeneity and a decreased gene connectivity scaling resulting in more frequent longer-distance contacts^{33,34}. Chromatin domain structures with a higher D have been linked to fostering further upregulation of initially upregulated genes and concomitant suppression of downregulated genes^{25,33}. In turn, these processes result in transcriptional patterns with greater transcriptional malleability and intercellular transcriptional heterogeneity. As neoplastic cells must keep developing new traits in response to constraints (e.g., hypoxia, immune system attack, new microenvironment), they benefit from transcriptional plasticity. Tumor cells that can more efficiently upregulate critical pro-survival pathways for a given level of stress through transcriptional malleability and heterogeneity have a higher likelihood of attaining a rare transcriptional state that is critical for cancer cell survival, thus further carrying this transcriptional phenotype through replication and increasing the probability of their progeny to acquire other gene mutations, some of which may be beneficial to tumorigenesis. Thus, transcriptional plasticity-

facilitating chromatin states (including higher chromatin packing domain D) may play a critical role in creating a “proneoplastic positive feedback loop” and therefore serve as a marker for neoplastic progression³⁴.

Chromatin structural changes occur from across the chromatin chain to domains at length scales from ~20 to ~300 nm which are too small to be observed by conventional optical microscopy. Our group utilized statistical spectroscopic techniques with optical microscopy measurement and developed csPWS technology to detect these nanoscale chromatin structural changes in the buccal mucosa. csPWS is a fast, reliable, and nanoscale-sensitive optical spectroscopic technique that can detect buccal chromatin conformation changes with a sensitivity between 23 and 334 nm³⁵. The key innovation in csPWS is statistical nanosensing in which sub-diffractive structures, while not resolvable through conventional optical microscopy, are detectable through analysis of the spatial variations of the refractive index (RI) via the spectroscopic analysis of scattered light interference within each of the microscopic resolution voxels^{24,36-41}. The output of csPWS microscopy is an image where a pixel describes chromatin structural heterogeneity by estimating packing scaling D for each coherence volume centered around it^{29,32,42}.

We searched for a suitable approach to thoroughly capture the complex biological interaction between lung cancer and packing scaling D in the nucleus of the buccal mucosa. Packing scaling D describes a quantitative statistical measurement from 3D packing of the chromatin polymer within a self-similar domain. However, local physical conditions such as nuclear crowding density, genomic size (N_d), domain volume fraction, and domain intracellular positioning (peripheral vs interior, etc.) are also important physical regulators that help determine chromatin connectivity, accessibility, and transcriptional heterogeneity and in doing so, gene expression²⁵. As packing scaling D is not the only predictor of the plasticity-fostering conformation, calculating the average D will not fully capture the complexity of chromatin regulatory mechanisms of gene expression. Thus, we utilized advanced machine learning algorithms and artificial intelligence (AI) to distinguish the lung cancer biological footprints engraved on the nucleus D images. Such a novel and potent “hybrid” AI+etiologic biomarker approach is possible by developing neural network (NN) layers informed with mechanistic data obtained from chromatin structure alterations and the packing scaling D image. Thus, we coupled our novel csPWS microscopy with a knowledge-based AI approach and achieved high sensitivity in the detection of early-stage lung cancer.

Material and Methods

csPWS nanocytology involves collection, shipment, and preparation of buccal samples followed by csPWS image acquisition, and evaluation of the nucleus chromatin packing scaling D image using an AI-driven approach.

(i) *Patient Recruitment.* Patients were recruited through Institutional Review Board-approved studies where written informed consent was obtained from all participants. The cohort comprised 96 patients with histologically confirmed lung cancer within the last year (case population) and 83 patients with a negative LDCT scan within the last year (control populations). 167 patients were over 45 years of age, nine patients were 27 to 44 years of age, and the age of three patients was unknown. Exclusion criteria were family history of lung cancer, exposure to chemotherapy and radiation in the past three months, pregnant/lactating women, and inability to give informed consent. Our data were obtained at two different sites: Site 1, Northwestern Memorial Hospital (NMH) in Chicago, Illinois, US, and Site 2, Boston Medical Center (BMC) in Boston, Massachusetts, US. The control population included non-smokers, low-risk and high-risk

smokers, and patients with benign nodules. The lung cancer patients included all stages but were predominantly Stage-I patients (62% for Site 1 data and 76% for Site 2 data).

(ii) *Sample collection.* Buccal samples were collected in the primary care physician's office through a buccal swab procedure using a minimally invasive standard of care (Cytobrush, CooperSurgical, Inc., Trumbull, CT, USA). The patients rinsed their mouths with water three times before the physician placed the bristles on the inside of one buccal surface followed by a top to down motion including brush rotation. Next, the impregnated swabs were dipped into 1.5ml vial tubes (Neptune Scientific, San Diego, USA) containing 750 ml of 25% ethanol (collection buffer). The samples were then packaged and shipped to the central lab for PWS microscopy.

(iii) *Sample shipment.* The Site 2 samples were shipped through the air from an out-of-state location while the Site 1 samples were shipped by ground transportation from an in-state location. The samples were maintained at a temperature below 10°C during transport using a custom-built transport kit and were received at the central facility within 24 hours of sample collection. The transportation kit included an outer corrugated box (Uline, Pleasant Prairie, WI, USA) and polar pack refrigerants (SONOCO Thermosafe, Arlington Heights, IL, USA) and temperature was monitored using a temperature indicator (Timestrip, Cambridge, UK). The sealed vial was packaged using an inner Styrofoam container and absorbent sheets to avoid possible leakage under refrigerated conditions.

(iv) *Sample preparation.* Clinical samples were prepared within 24 hours of collection based on the approaches reported earlier.⁴² In brief, the samples in 25% ethanol were spray deposited on a Fisher brand Superfrost microscope slide (Fisher Scientific, Hampton, NH, USA) using our custom-built cell deposition system to form non-overlapping monolayer buccal cells. The sample slide was air-dried prior to cytological fixation with 95% ethanol (Thermo Fisher Scientific, Waltham, MA, USA) followed by csPWS microscopy.

(v) *Standard Operating Procedure (SOP).* We developed a well-designed SOP to capture buccal csPWS nucleus chromatin structure changes. Our goal was to ensure a fast, robust, reliable, and repeatable protocol with small variability of physical features of the cells acquired by csPWS from each patient. To minimize the complexity at the collection site, we carried out the cell fixation and sample deposition at the central lab instead of the primary care office⁴³. For each patient, a total of >30 cells were collected, where the sample size number was determined by power analysis with the confidence interval (CI) on mean Σ restricted to be less than 5% of the difference between cancer and the control population⁴³. We created a sample transport solution of 25% ethanol and used our custom-built cell deposition device to spray deposit non-deformed, non-overlapping monolayer buccal cells with clear nuclear boundaries on the glass slide. An airdrying step enhanced the attachment of cells to the glass followed by fixation with 95% ethanol and csPWS microscopy. The csPWS microscope was controlled via custom software with a graphical user interface (GUI). The imaging procedure began by scanning the whole slide using a 10X air objective. A semi-automated slide-map module was developed to rapidly generate a low-magnification image of a slide by collecting and stitching individual slide region images. This assisted a trained user blinded to the diagnostic information in selecting over 30 buccal cells across the entire slide in a timely manner. Our cell screening protocol selected non-folded and non-overlapping cells with clear nucleus boundaries. The csPWS spectral acquisition was performed for the cells in a liquid medium (95% ethanol) using a liquid-dipping 40X optical objective (Nikon, Melville, NY, USA) to match the RI between the buccal cell and liquid cover. Cs-PWS acquisition algorithm automatically acquired spectral data for selected cells, and the analysis algorithm rapidly

generated the processed spectral data. These processes facilitated reliable and reproducible results, making csPWS suitable for larger studies that span more clinical sites.

(vi) *csPWS microscopy*. Conventional microscopy systems are unable to resolve structures smaller than 200 nm (half the wavelength of light). Our lab developed csPWS, an optical statistical spectroscopic nanosensing approach for the detection of chromatin packing domain changes in the nucleus of buccal mucosa. This enables us to distinguish chromatin structural changes among histologically normal buccal cells that may carry a signature of cancer. csPWS detects chromatin packing domain changes through statistical spectroscopic approaches that evaluate the scattering of light within the nucleus of the mucosa. The detailed principles of csPWS microscopy have been provided in^{24,37-39,41,43,44} and here we briefly describe them. csPWS acquires a high magnification of monochromatic spectrally resolved images between wavelengths of 450-700 nm and distinguishes the non-resolvable sub-diffractive length scales through the assessment of scattered light. For a given cell location of $\mathbf{r}$, the RI n is proportional to the local macromolecular density (ρ) of proteins, DNA, RNA etc. The refraction increment is constant and nearly independent of macromolecular chemical composition and is estimated by eq-1^{43,45}.

$$(\rho): n(\mathbf{r}) = n_{\text{media}} + \alpha\rho(\mathbf{r}) \quad (1)$$

csPWS benefits from liquid cover microscopy to nearly match the RI between the buccal cell and liquid cover (shown in **Fig. 1 a**) while creating a mismatch between the cell-glass interface. Thus, the scattering of light from a reference wave due to the variation of RI is proportional to the density of the intracellular macromolecules in the coherence volume determined by the depth of field (DOF) longitudinally and the axial plane for each pixel. csPWS calculates the standard deviation of the interface spectra (Σ) obtained from the spectrum of reference wave and scattering of all RI changes due to macromolecular nanoscale density variations across different wavelengths. Σ is proportional to the Fourier transform of the autocorrelation function (ACF) of $\rho(\mathbf{r})$ integrated over the Fourier transform of the coherence volume. Each csPWS image stack is normalized by the reference wave that is acquired at the interface of the glass and cover media from a blank region on the slide. These subdiffractive variations in RI distribution measured as Σ can be characterized by the chromatin packing scaling factor. The relation between Σ and D could be explained as $\Sigma \propto (D - D_0)\delta n$, where $D_0 \sim 1.33$ is the fractal dimension of a random 2D-polymer and δn is the standard deviation of the correlation length scale⁴⁴. Simulation and experimental results confirmed a high correlation between Σ and packing scaling D associated with each domain^{25,29,32,42}. We estimated the packing scaling D value for the domains in each pixel (350 nm by 350 nm) based on the analytical framework for quantifying chromatin structure with spectral microscopy provided in⁴². csPWS length scale sensitivity depends on illumination-collection geometry. We designed a small-to-moderate numerical aperture (NA) of light incidence of 0.6, and light collection NA of 0.8 for csPWS. This illumination setting ensures a uniform intensity across the sample plane due to the Köhler alignment⁴⁶ and delivers a chromatin length scale sensitivity of 23-334 nm (exact value depends on the sample structure and thickness)^{35,47}. The larger length scales do not affect csPWS output signal^{43,44}. Electron microscopy data has revealed buccal chromatin structures significantly alter at this length scale range⁴⁸. Thus, csPWS nanocytology is mainly sensitive to length scales that are not resolvable by conventional optical microscopy while carrying a profound signature of field carcinogenesis and alterations in chromatin packing domains.

(vii) *csPWS Instrument*: The design and schematic of the csPWS instrument and the optical path for collecting buccal cell data are shown in **Fig. 1a**. The csPWS optical system (shown in **Fig. 1b**)

is built on a commercial microscope (Nikon Instruments, Melville, NY, USA) using a Nikon Eclipse Ni-E microscope body with modifications made to include a Xenon lamp (Exceliatas, Tampa, FL, USA). The light is fed to an acousto-optic tunable filter with a switching speed of 50 μ s, a bandwidth of 3nm, and a spectral range of 450-700nm (Gooch and Housego, UK). The light passes through objective lenses (Nikon, Melville, NY, USA) that are attached to an automated objective turret and onto a sample that is positioned with a nanomotion piezo-stage (Prior, Rockland, MA, USA). The data is recorded through a digital CMOS camera, ORCA Flash 2.8 (Hamamatsu, Bridgewater, NJ, USA) thereby enabling hyperspectral imaging. We attain high throughput and automated csPWS acquisition by utilizing Kohler illumination for uniform sample illumination. We use a high-speed, high-resolution RGB camera (Thorlabs Inc., Newton, NJ, USA) for low-magnification slide mapping. The workflow of csPWS buccal nanocytology is presented in **Fig 1c**. Packing scaling D images of the nucleus of 30 to 40 isolated non-overlapping buccal cells were acquired for each patient in liquid using a liquid-dipping 40X optical objective (Nikon, Melville, NY, USA) to match the RI between the buccal cell and liquid cover (shown **Fig 1.a**). AI and advanced machine learning algorithms were used to distinguish characteristics of field carcinogenesis and the alterations in the packing scaling D image.

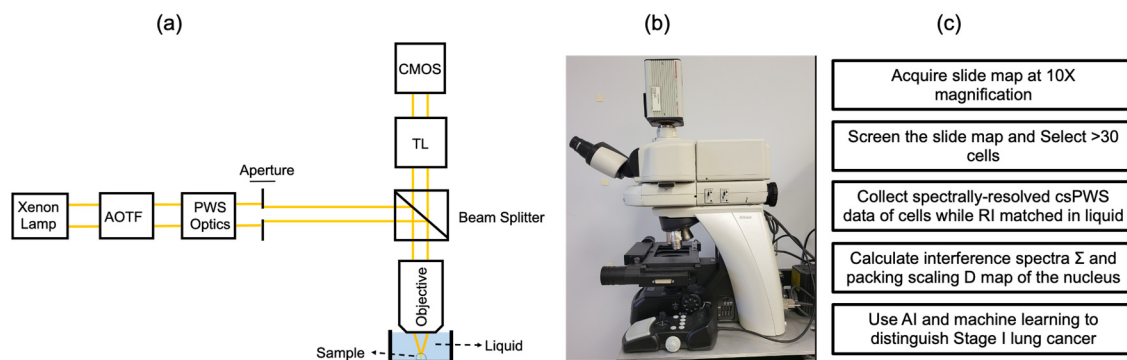


Figure 1. (a) Schematic of csPWS instrument (b) csPWS instrument (c) workflow of csPWS buccal nanocytology

(viii) *AI analysis of packing scaling D*. We used AI on csPWS data to see whether we can detect field carcinogenesis of buccal mucosa of patients with lung cancer and distinguish alterations in buccal chromatin packing domains that indicates tumor initiation and progression. Our AI-driven approach consisted of nucleus segmentation, preprocessing, feature learning, and classification of csPWS images as shown in **Fig. 2**. Nearly 7000 buccal csPWS D images (960 $\times$ 720 pixels) from 179 patients was evaluated in this study. Nuclear segmentation was conducted by a trained user blinded to the diagnostic information and the outlier cells with deformed shapes were excluded. Next, the nuclei images were resized and passed through min-max normalization in our preprocessing subsection. For the feature extractor unit, we incorporated transfer learning on VGG16 architecture, a convolutional neural network (CNN) pre-trained on 14 million images belonging to 1,000 different labels from the ImageNet data set. Features were extracted from multiple convolution layers of the architecture (Conv 2 through Conv 5) and flattened. The mean and standard deviation of the flattened features across all cells belonging to a patient were calculated to create a feature vector. We used a multiple instance learning (MIL) method of instance-level feature aggregation that facilitated the usage of patient-level clinical ground truths. MIL enables a seamless process integration for pathologists⁴⁹ and for that reason was used in our approach. To further reduce the dimension of the patient feature, we conducted a recursive feature elimination method using a random forest algorithm, thus selecting a panel of 40 features with

enhanced classification properties. We used a parameter-tuned random forest classifier to determine the patients with lung cancer from the control populations using a patient-wise analysis of the diagnostic features obtained by CNN. The tuning of the classifier model for optimal parameters was performed using grid search by searching through iterations of multiple configurations, of which the model configuration with minimal error for our dataset was used for classification. For a robust evaluation of model performance on our relatively small dataset, we calculated our metrics AUC, sensitivity, and specificity using a stratified 4-fold cross-validation method with 5 iterations.

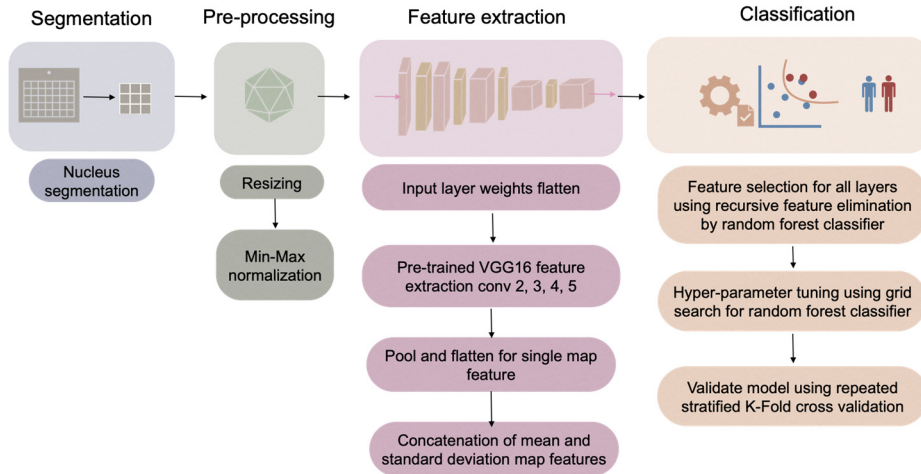


Figure 2. Workflow and architecture of the feature extraction and classification steps

Results

Patient Recruitment and Demographics. We analyzed csPWS D images of clinical buccal samples in a double-blind case-control study from two different clinical sites stationed in an in-state and an out-of-state location. Most patients identified with lung cancer were at Stage-I: 26 of the 42 (62%) at Site 1, and 41 of 54 (76%) at Site 2. The percentage of female patients with lung cancer was higher at Site 2 (71%) compared to Site 1 (56%). Minority populations were enriched at Site 2 with Caucasians forming only 49% of the population compared to 80% at Site 1. PWS D image of two histologically normal buccal cells (confirmed by reflectance image) shows an example of intercellular domain variation and an overall increase of D in a cell belonging to patients with lung cancer in comparison with a smoking control (**Fig. 3**).

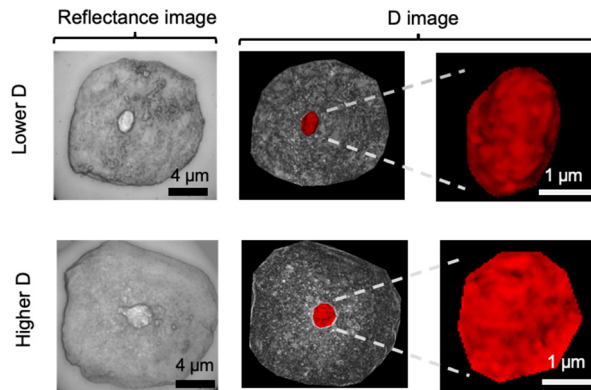


Figure 3. Reflectance image (left) the D image distribution (right) in cells from a control patient (top) and from a patient with lung cancer (bottom).

Population-based study showed a similar trend where the overall packing scaling D was higher in patients with lung cancer compared to the healthy control. Patient demographics including age, pack-years of smoking, gender, and race, and their association with D were evaluated with the significance criterion of the p-value using regression analysis for the patients with lung cancer and the control population (shown in **Table 1**).

Table 1. Demographic factor distribution for Sites 1&2. The Average value and standard deviation are reported

Demographic Factor	Control Parameter		Cancer Parameter		Effect on D	
	Site 1	Site 2	Site 1	Site 2	Site 1	Site 2
Age	59 ± 11	61 ± 11	67 ± 12	67 ± 7	0.02	0.47
Pack-Years	35 ± 26	39 ± 34	37 ± 30	41 ± 30	0.22	0.50
Gender (% Female)	49%	42%	56%	71%	0.69	0.19
Race (% Caucasian)	80%	49%	80%	49%	0.62	0.84

Our regression modeling did not find any statistically significant relationship between gender, race, and smoking pack-years with packing scaling D across both sites as shown in **Table 1**. While we did not observe a significant relationship between the age and packing scaling D in Site 2 (p-value = 0.47), age presented a statistically significant relationship with packing scaling D at Site 1. We further interrogated packing scaling D data at Site 1 and determined a statistically significant negative correlation may exist between age and D (**Fig. 4**). Patients with lung cancer presented significantly (p<0.001) higher overall D than the normal population while lung cancer cases were more common among the elder population. Thus, an age adjustment for addressing the negative correlation between age and D would further enhance the overall D diagnostic performance. Thus, our reporting diagnostic performance at Site 1 may be an underestimation.

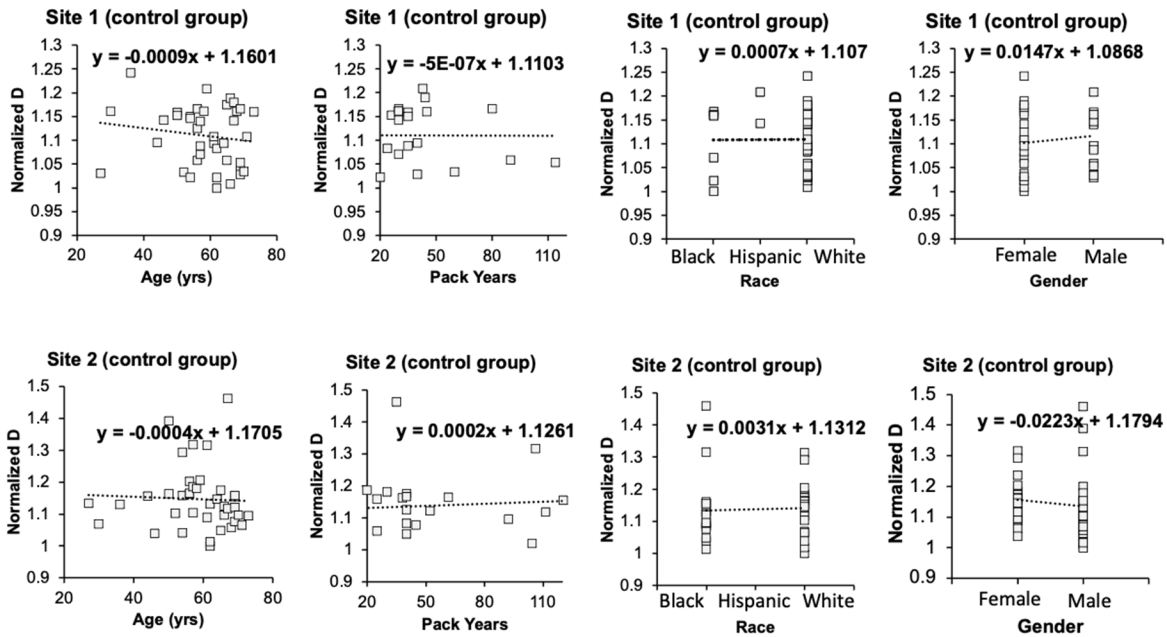


Figure 4. Linear regression analysis assessed the impact of demographic factors on average packing scaling D with the control population

AI-enabled csPWS for Early-stage Lung Cancer Detection. We evaluated the average packing scaling D among the entire 179 patients (data normalized by year and site) and determined that average D significantly increased among patients with lung cancer (p-value <0.001). The average packing scaling D provides a statistical descriptor of chromatin structure, however, factors such as chromatin connectivity, accessibility, and transcriptional heterogeneity also determine gene expression. Local physical conditions such as nuclear crowding density, genomic size, domain volume fraction, and domain intracellular positioning are important determinants of a plasticity-fostering conformation²⁵. Thus, average packing scaling D does not fully describe the complex regulatory role of chromatin structure in gene expression. We utilized deep learning and machine learning algorithms to capture the biological complexity of chromatin structure and the statistical marker, D. Specifically, we used transfer learning-based deep CNN modeling along with a random forest classifier on the D image of the buccal nucleus to distinguish patients with lung cancer from the control population. Our AI-driven csPWS nanocytology distinguished patients with Stage-I lung cancer from the control population with AUC of 0.92 ± 0.06 (Se = 92%, Sp = 89%) at Site 1 and AUC of 0.82 ± 0.11 (Se = 78%, Sp = 83%) as Site 2 as shown in **Table 2**. Our approach maintained its diagnostic performance for the detection of later-stage lung cancer and showed an AUC of 0.90 ± 0.07 (Se = 86%, Sp = 89%) and 0.82 ± 0.09 (Se = 78%, Sp = 87%) for detection of all stage lung cancer in Site 1 and Site 2, respectively.

Table 2. AUC comparison between Site 1&2 for control, Stage-I, and all-Stage lung cancer

AUC comparison	AI-enabled csPWS Nanocytology	
	Site 1 (#Control, #Case)	Site 2 (#Control, #Case)
Stage-I	0.92 (± 0.04) (40,26)	0.82 (± 0.11) (43,40)
Stage-I&II	0.92 (± 0.04) (40,33)	0.82 (± 0.04) (43,44)
All-Stages (I, II, III, & IV)	0.90 (± 0.07) (40,42)	0.82 (± 0.09) (43,54)

Discussion

We demonstrated that AI-enabled buccal csPWS nanocytology can be an effective modality for the detection of Stage-I lung cancer via the assessment of chromatin structural changes. Buccal csPWS nanocytology enables a fast, minimally invasive, accessible, and sensitive approach for the detection of early curable stage lesions (I&II) equivalent to advanced stage lesions (III, IV). High diagnostic performance was achieved at two different clinical sites with different socio-demographic populations and different proximity to the central lab which indicates the robustness of buccal csPWS nanocytology.

From a clinical perspective, the early detection of lung cancer has significant importance. Lung cancer patients' survival rate declines significantly as cancer progresses. Recent work has sought to develop new clinical methods to detect lung cancer, but these attempts often fail for the purpose of early-stage diagnosis. For example, the novel liquid biopsy technology has shown promising diagnostic performance for the detection of lung cancer, but the performance for the detection of early-stage lung cancer notably drops due to the limitations associated with the biology of the biomarker source¹². We started our search for a new biomarker source and biomarker type based on the molecular biology of lung cancer. Lung cancer is developed by the dysregulation of diverse

mutational processes, at different rates resulting in genetic heterogeneity in tumors. The number of possible molecular/gene alterations in lung cancer is extremely high with notable tumor heterogeneity which negatively impacts the efficacy of tumor byproduct biomarkers obtained from blood. Downstream biomarkers are expected to lose their sensitivity at the early stage of lung cancer where the tumor is small due to the staggering tumor heterogeneity. Fewer mutations per Mbp and fewer clones alongside the high tumor heterogeneity induced by the small lesions would negatively impact the diagnostic performance for the detection of early-stage lung cancer. Later-stage mutations may not exist at a high level in small tumor lesions. The AI-enabled technologies that evaluate hundreds of biomarkers will face challenges as most of those products will not be produced sufficiently from a small tumor. Thus, their sensitivity drops for the detection of small lesions^{12,50,51}. Therefore, we decided to look for a new biomarker source. The field carcinogenesis introduces buccal chromatin structures as an accessible “substrate” on which the tumor originates, thus biomarkers obtained from it will not depend on the size of the tumor. These realities motivated us to leverage the combination of field carcinogenesis with a novel optical spectroscopic statistical nanosensing technology to propose a new type of biomarker: nucleus chromatin structure from buccal mucosa.

There is growing evidence advocating for lung cancer field carcinogenesis through assessment of more accessible sites from the aero-digestive mucosa¹⁵⁻²⁰. For example, similar genomic and epigenomic changes to cancer are observed in the distal lung and across entire airway epithelial cells^{18,52-54}. Other studies showed smoking induces similar genetic changes in the bronchial airway and epithelia of the nasal, mucosa^{55,56}. The elevated incidence of synchronous and metachronous neoplastic lesions with the increased ratio of secondary cancer^{15,57-62}, similar chromosomal, gene methylation, and gene mutations (EGFR and p53) abnormalities in locations exposed to the same carcinogens are widely explained with field carcinogenesis^{18,57,63-67}. The utility and clinical application of field carcinogenesis are confirmed through the evaluation of protein expression, and miRNA^{18,68}. More specifically, buccal mucosa has been known as the “molecular mirror” of bronchial epithelium for lung cancer^{54,69}. A variety of genetic/epigenetic changes in the buccal mucosa are concordant with those in lung cancer^{20-22,70}. Changes in gene/protein expression of buccal mucosa have traditionally been used for diagnostic applications while we assessed chromatin structural changes for the detection of Stage-I lung cancer⁷¹. The highly complex and dynamic chromatin structure is widely known as ubiquitous and a common denominator for genetic alterations in biological cells and cancer-promoting transformation^{72,73}. Chromatin structure determines the microenvironment enabling neoplasia, thus, can be a predictor of cancer in histologically normal cells even prior to the formation of a tumor, and unlike downstream biomarkers, should retain sensitivity to small tumors. Cytometry measurement⁷⁴ and recent spectroscopic studies^{16,23,25,26} have indicated structural changes and field cancerization in the oral cavity of patients with lung cancer which advocate the diagnostic application of chromatin structural changes. Electron microscopy image analysis determined alterations in buccal chromatin packing at a length scale between 80 nm to 200 nm⁴⁸ is a profound and significant characteristic of field cancerization.

These are the reasons we developed buccal csPWS nanocytology to interrogate the chromatin structural changes at the length scale of 23-334 nm and detect early-stage lung cancer. csPWS senses chromatin structural changes based on the principle that light scatters substantially more in small dense chromatin packing domains than in large loosely packed chromatin domains. csPWS achieves its high sensitivity to chromatin structural changes via a spectroscopic nanosensing statistical approach. We enhanced the signals obtained from intra-cellular macromolecule density

changes in the nucleus by matching the RI at the cell/liquid interface and creating a mismatch at the cell/glass interface. The packing scaling D image calculated by csPWS⁴² provides statistical characterizing data from the nucleus chromatin packing domain changes which are used as a proxy for the prediction of plasticity-fostering conformation.

The role of packing scaling D in defining chromatin conformation and regulation of gene activity is a complex phenomenon. Alongside D, other structural predictors such as the domain size, chromatin volume concentration (CVC), average nuclear density, domain volume fraction, and their location in the nucleus (peripheral vs interior) are a predictor of the plasticity-fostering chromatin conformation^{25,29,32,42}. Thus, calculating the average D for each buccal nucleus will not fully capture the role of chromatin structure impact on tumor genesis. As such, we used an AI-driven approach to capture the packing scaling D image associated with the plasticity-fostering conformation that can lead to tumor genesis and is the predictor of Stag-I lung cancer.

Our AI-driven biomarkers are developed with layers informed by mechanistically driven chromatin structural changes and have advantages over conventional biomarker discovery methods such as 1) single hypothesis-driven biomarkers and 2) the AI-driven “black-box” approaches. A single hypothesis-driven biomarker approach cannot fully capture the complexity of biological interactions while the “black-box” approach will fail to deliver an accurate diagnosis in a relatively limited sample size due to the lack of mechanistic rationale. In this work, we bridged the two approaches while taking advantage of their strengths and mitigating their weaknesses for the detection of early-stage lung cancer.

We assessed AI-enabled buccal csPWS nanocytology across two different sites with different proximity to our central lab and different socioeconomic status and realized a high diagnostic performance for detecting Stage-I lung cancer (AUC>0.82, Se > 78%, Sp > 83%) which is a significant improvement over the current state of art technologies (Se = 34-50%).^{13,50} In addition to high sensitivity to small lesions and Stage-I lung cancer, the diagnostic performance was independent of the tumor size and maintained for stages II, III, and IV. This high sensitivity to early-stage and late-stage lung cancer is possibly due to field carcinogenesis and our design of the biomarker source (buccal mucosa) and biomarker type (chromatin structure).

To interrogate the robustness of our approach we specifically constructed this study from two different sites, one with affluent economic conditions (Site 1) near our central lab and the other one from a safety net hospital (Site 2) in a far out-of-state location. This design could assist us to evaluate the robustness of csPWS in different settings and shipment conditions. We treated/analyzed data from these sites individually rather than combining them. While there was a slight difference in the performance between the two sites, the diagnostic performance was still notably better than the currently available techniques for Stage-I lung cancer for both settings. Future studies may include more sample collection sites and control for possible chromatin degradation in prolonged out-of-state shipments and the hospital socioeconomic variables to measure this method's performance more robustly.

We envision buccal csPWS nanocytology as a sensitive, accurate, fast, and accessible pre-screening method for the LDCT procedure. Currently, only about 5% of the eligible population undergo LDCT screening which indicates challenges in the effective implementation of LDCT procedure on a broad scale³. Assessment of other tests such as Pap smear and colorectal Cologuard showed that pre-screening can increase adherence to the screening protocol^{75,76}. In a similar manner to what a Pap smear did for cervical cancer, we aimed to develop a pre-screening test for

LDCT procedure to possibly enhance access and adherence to the screening program. Our well-designed SOP enables an at-home self-administered collection and preparation of clinical samples which can notably enhance screening accessibility. If the buccal csPWS nanocytology is positive, a patient can be directed to a more expensive/invasive LDCT procedure. Successful deployment of this pre-screening test in clinical practice may potentially enable the screening of a larger portion of the asymptomatic at-risk population and identifies patients that urgently need to undergo LDCT screening. Additionally, large-scale randomized trials indicate that LDCT is triggered by benign lesions that frequently result in false positives⁷⁷⁻⁷⁹. Given the low prevalence of lung cancer among the LDCT-eligible smoking population, this exposes many patients to overdiagnosis, unnecessary invasive procedures, anxiety, and overexposure to radiation. A csPWS test with high sensitivity to Stage-I lung cancer prior to LDCT has the potential to increase the downstream prevalence of cases among the population undergoing LDCT thereby increasing positive predictive value (PPV) and enhancing the effectiveness of the screening protocol.

Our study and analysis had some limitations. This clinical study used a case-control design with a limited number of patients and, therefore, could not provide a definitive assessment of performance for our diagnostic method. However, it shows that our AI-driven nanosensing technology can be a promising approach to the early detection of lung cancer. Future work may build upon this study and involve a large-scale analysis to advance the model and technique. Because most patients in this study have a history of smoking, future studies could validate our findings with a non-smoker population. Additionally, there are some potentially confounding variables that future work must examine. The possible impact of shipment on chromatin degradation is unknown and needs to be determined in future work. The contribution of neoplastic signals from different layers of the buccal mucosa is not known while the exact complex structures of packing domains and packing scaling D are not established. This suggests that further optimization may lead to improved diagnostics.

Acknowledgment

This work was supported by National Institutes of Health Grants (NIH) grants U54CA268084, R01CA228272, and R01CA225002, and National Science Foundation grant (NSF) grant EFMA-1830961.

Author Contribution

VB, HS, HKR, and AB designed the project. SP, AD, and PV conducted statistical and data analysis. AD and PV prepared the manuscript while all authors contributed and provided feedback. AB and HKR collected clinical data from two sites. JL collected csPWS data.

Competing Interests

The authors declare no competing interests.

Data availability

The raw datasets generated during and/or analyzed during the study are available from the corresponding author on reasonable request.

References:

- 1 SEER. *Annual Report to the Nation 2022: Overall Cancer Statistics*, <<https://seer.cancer.gov/index.html>> (2022).
- 2 Chen, Y. Y. & Huang, T. W. Prognostic factors of patients with pathologic stage I lung adenocarcinoma. *J Thorac Dis* **10**, S1115-S1118, doi:10.21037/jtd.2018.04.37 (2018).
- 3 Fedewa, S. A. *et al.* State Variation in Low-Dose Computed Tomography Scanning for Lung Cancer Screening in the United States. *J Natl Cancer Inst* **113**, 1044-1052, doi:10.1093/jnci/djaa170 (2021).
- 4 SEER. *Cancer of the lung and bronchus - Cancer stat facts, SEER.*, <<https://seer.cancer.gov/statfacts/html/lungb.html>> (
- 5 Bach, P. B., Silvestri, G. A., Hanger, M., Jett, J. R. & American College of Chest, P. Screening for lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* **132**, 69S-77S, doi:10.1378/chest.07-1349 (2007).
- 6 Michael J. Duffy EMAIL logo , E. P. D. O. I. a. J. C. Circulating tumor DNA (ctDNA) as a pan-cancer screening test: is it finally on the horizon? *Clinical Chemistry and Laboratory Medicine (CCLM)* **59**, doi:<https://doi.org/10.1515/cclm-2021-0171> (2021).
- 7 Phallen, J. *et al.* Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aan2415 (2017).
- 8 Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926-930, doi:10.1126/science.aar3247 (2018).
- 9 Shen, S. Y., Burgener, J. M., Bratman, S. V. & De Carvalho, D. D. Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. *Nat Protoc* **14**, 2749-2780, doi:10.1038/s41596-019-0202-2 (2019).
- 10 Chen, X. *et al.* Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nature Communications* **11**, 3475, doi:10.1038/s41467-020-17316-z (2020).
- 11 Cristiano, S. *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385-389, doi:10.1038/s41586-019-1272-6 (2019).
- 12 Fiala, C. & Diamandis, E. P. Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection. *BMC Medicine* **16**, doi:10.1186/s12916-018-1157-9 (2018).
- 13 Klein, E. A. *et al.* Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Annals of Oncology* **32**, 1167-1177, doi:10.1016/j.annonc.2021.05.806 (2021).
- 14 Vansteenkiste, J., Dooms, C., Mascaux, C. & Nackaerts, K. Screening and early detection of lung cancer. *Ann Oncol* **23 Suppl 10**, x320-327, doi:10.1093/annonc/mds303 (2012).
- 15 Danely P. Slaughter M.D., H. W. S. M. D., Walter Smejkal M.D. "Field cancerization" in oral stratified squamous epithelium. Clinical implications of multicentric origin. *Cancer* **6**, 963-968, doi:[https://doi.org/10.1002/1097-0142\(195309\)6:5<963::AID-CNCR2820060515>3.0.CO;2-Q](https://doi.org/10.1002/1097-0142(195309)6:5<963::AID-CNCR2820060515>3.0.CO;2-Q) (1953).
- 16 Radosevich, A. J. *et al.* Buccal spectral markers for lung cancer risk stratification. *PLoS ONE* **9**, e110157, doi:10.1371/journal.pone.0110157 (2014).
- 17 Lochhead, P. *et al.* Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression. *Mod Pathol* **28**, 14-29, doi:10.1038/modpathol.2014.81 (2015).
- 18 Steiling, K., Ryan, J., Brody, J. S. & Spira, A. The Field of tissue injury in the lung and airway. *Cancer Prev Res* **1**, 396-403, doi:10.1158/1940-6207.capr-08-0174 (2008).

- 19 Hosgood, H. D. *et al.* Variation in oral microbiome is associated with future risk of lung cancer among never-smokers. *Thorax* **76**, 256-263, doi:10.1136/thoraxjnl-2020-215542 (2021).
- 20 Kolbasina, N. A. *et al.* Lung cancer increases H₂O₂ concentration in the exhaled breath condensate, extent of mtDNA damage, and mtDNA copy number in buccal mucosa. *Heliyon* **6**, e04303, doi:10.1016/j.heliyon.2020.e04303 (2020).
- 21 Saba, R., Halytsky, O., Saleem, N. & Oliff, I. A. Buccal Epithelium, Cigarette Smoking, and Lung Cancer: Review of the Literature. *Oncology* **93**, 347-353, doi:10.1159/000479796 (2017).
- 22 Shani, T. *et al.* Chromosomal numerical aberrations in apparently normal oral mucosa of heavy smokers affected by lung cancer. *Oral Oncol* **46**, 96-99, doi:10.1016/j.oraloncology.2009.10.010 (2010).
- 23 Bugter, O. *et al.* Towards the Optical Detection of Field Cancerization in the Buccal Mucosa of Patients with Lung Cancer. *Transl Oncol* **12**, 1533-1538, doi:10.1016/j.tranon.2019.07.018 (2019).
- 24 Roy, H. K. *et al.* Optical detection of buccal epithelial nanoarchitectural alterations in patients harboring lung cancer: implications for screening. *Cancer Res* **70**, 7748-7754, doi:10.1158/0008-5472.Can-10-1686 (2010).
- 25 Virk, R. K. A. *et al.* Disordered chromatin packing regulates phenotypic plasticity. *Sci Adv* **6**, eaax6232, doi:10.1126/sciadv.aax6232 (2020).
- 26 Almassalha, L. M. *et al.* The greater genomic landscape: the heterogeneous evolution of cancer. *Cancer research* **76**, 5605-5609 (2016).
- 27 Cherkezyan, L. *et al.* Nanoscale changes in chromatin organization represent the initial steps of tumorigenesis: a transmission electron microscopy study. *BMC Cancer* **14**, 189, doi:10.1186/1471-2407-14-189 (2014).
- 28 Xu, J. *et al.* Super-resolution imaging reveals the evolution of higher-order chromatin folding in early carcinogenesis. *Nature Communications* **11**, doi:10.1038/s41467-020-15718-7 (2020).
- 29 Li, Y. *et al.* Nanoscale chromatin imaging and analysis platform bridges 4D chromatin organization with molecular function. *Sci Adv* **7**, doi:10.1126/sciadv.abe4310 (2021).
- 30 Ide, S., Imai, R., Ochi, H. & Maeshima, K. Transcriptional suppression of ribosomal DNA with phase separation. *Sci Adv* **6**, doi:10.1126/sciadv.abb5953 (2020).
- 31 Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, doi:10.1126/science.aag0025 (2017).
- 32 Li, Y. *et al.* Analysis of three-dimensional chromatin packing domains by chromatin scanning transmission electron microscopy (ChromSTEM). *Scientific Reports* **12**, doi:10.1038/s41598-022-16028-2 (2022).
- 33 Gladstein, S. *et al.* Measuring nanoscale chromatin heterogeneity with partial wave spectroscopic microscopy. *Methods Mol Biol* **1745**, 337-360, doi:10.1007/978-1-4939-7680-5_19 (2018).
- 34 Almassalha, L. M. *et al.* Macro-genomic engineering via modulation of the scaling of chromatin packing density. *Nature Biomedical Engineering* **1**, 902-913, doi:10.1038/s41551-017-0153-2 (2017).
- 35 Cherkezyan, L. *et al.* Review of interferometric spectroscopy of scattered light for the quantification of subdiffractional structure of biomaterials. *J Biomed Opt* **22**, 30901, doi:10.1117/1.Jbo.22.3.030901 (2017).

- 36 Konda, V. *et al.* Nanoscale markers of esophageal field carcinogenesis: potential implications for esophageal cancer screening. *Endoscopy* **45**, 983-988, doi:10.1055/s-0033-1344617 (2013).
- 37 Backman, V. & Roy, H. K. Advances in biophotonics detection of field carcinogenesis for colon cancer risk stratification. *Journal of Cancer* **4**, 251-261, doi:10.7150/jca.5838 (2013).
- 38 Damania, D. *et al.* Insights into the field carcinogenesis of ovarian cancer based on the nanocytology of endocervical and endometrial epithelial cells. *International Journal of Cancer* **133**, 1143-1152, doi:10.1002/ijc.28122 (2013).
- 39 Subramanian, H. *et al.* Optical methodology for detecting histologically unapparent nanoscale consequences of genetic alterations in biological cells. *Proc Natl Acad Sci U S A* **105**, 20118-20123, doi:10.1073/pnas.0804723105 (2008).
- 40 Subramanian, H. *et al.* Partial-wave microscopic spectroscopy detects subwavelength refractive index fluctuations: an application to cancer diagnosis. *Opt Lett* **34**, 518-520, doi:10.1364/ol.34.000518 (2009).
- 41 Subramanian, H. *et al.* Nanoscale cellular changes in field carcinogenesis detected by partial wave spectroscopy. *Cancer Res* **69**, 5357-5363, doi:10.1158/0008-5472.CAN-08-3895 (2009).
- 42 Eid, A. *et al.* Characterizing chromatin packing scaling in whole nuclei using interferometric microscopy. *Opt Lett* **45**, 4810-4813, doi:10.1364/OL.400231 (2020).
- 43 Subramanian, H. *et al.* Procedures for risk-stratification of lung cancer using buccal nanocytology. *Biomed Opt Express* **7**, 3795-3810, doi:10.1364/boe.7.003795 (2016).
- 44 Cherkezyan, L. *et al.* Interferometric spectroscopy of scattered light can quantify the statistics of subdiffractional refractive-index fluctuations. *Physical Review Letters* **111**, doi:10.1103/physrevlett.111.033903 (2013).
- 45 Davies, H. G., Wilkins, M. H. F., Chayen, J. & La Cour, L. F. The Use of the Interference microscope to determine dry mass in living cells and as a quantitative cytochemical method. *Journal of Cell Science* **s3-95**, 271-304, doi:10.1242/jcs.s3-95.31.271 (1954).
- 46 Chandler, J. E. *et al.* High-speed spectral nanocytology for early cancer screening. *Journal of Biomedical Optics* **18**, doi:Artn 11700210.1117/1.Jbo.18.11.117002 (2013).
- 47 Cherkezyan, L., Subramanian, H. & Backman, V. What structural length scales can be detected by the spectral variance of a microscope image? *Optics Letters* **39**, 4290, doi:10.1364/ol.39.004290 (2014).
- 48 Bugter, O. *et al.* Early upper aerodigestive tract cancer detection using electron microscopy to reveal chromatin packing alterations in buccal mucosa cells. *Microsc Microanal* **27**, 878-888, doi:10.1017/s1431927621000507 (2021).
- 49 Perincheri, S. *et al.* An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. *Modern Pathology* **34**, 1588-1595, doi:10.1038/s41379-021-00794-x (2021).
- 50 Liu, M. C. *et al.* Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol* **31**, 745-759, doi:10.1016/j.annonc.2020.02.011 (2020).
- 51 Putcha, G., Gutierrez, A. & Skates, S. Multicancer screening: one size does not fit all. *JCO Precision Oncology*, 574-576, doi:10.1200/po.20.00488 (2021).

- 52 Dakubo, G. D., Jakupciak, J. P., Birch-Machin, M. A. & Parr, R. L. Clinical implications and utility of field cancerization. *Cancer Cell International* **7**, 2, doi:10.1186/1475-2867-7-2 (2007).
- 53 Nelson, M. A., Wymer, J. & Clements, N., Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. *Cancer Lett* **103**, 115-121, doi:10.1016/0304-3835(96)04202-4 (1996).
- 54 Bhutani, M. *et al.* Oral epithelium as a surrogate tissue for assessing smoking-induced molecular alterations in the lungs. *Cancer Prev Res* **1**, 39-44, doi:10.1158/1940-6207.capr-08-0058 (2008).
- 55 Sridhar, S. *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* **9**, 259, doi:10.1186/1471-2164-9-259 (2008).
- 56 Gomperts, B. N., Walser, T. C., Spira, A. & Dubinett, S. M. Enriching the Molecular Definition of the Airway “Field of Cancerization:” Establishing new paradigms for the patient at risk for lung cancer. *Cancer Prev Res* **6**, 4-7, doi:10.1158/1940-6207.Capr-12-0470 (2013).
- 57 Kopelovich, L. *et al.* Surrogate anatomic/functional sites for evaluating cancer risk: an extension of the field effect. *Clin Cancer Res* **5**, 3899-3905 (1999).
- 58 Tepperman, B. S. & Fitzpatrick, P. J. Second respiratory and upper digestive tract cancers after oral cancer. *Lancet* **2**, 547-549, doi:10.1016/s0140-6736(81)90938-7 (1981).
- 59 Partridge, M., Emilion, G., Pateromichelakis, S., Phillips, E. & Langdon, J. Field cancerisation of the oral cavity: comparison of the spectrum of molecular alterations in cases presenting with both dysplastic and malignant lesions. *Oral oncology* **33**, 332-337 (1997).
- 60 Hittelman, W. N. *et al.* Detection of chromosome instability of tissue fields at risk: in situ hybridization. *Journal of Cellular Biochemistry* **63**, 57-62 (1996).
- 61 Bedi, G. C., Westra, W. H., Gabrielson, E., Koch, W. & Sidransky, D. Multiple head and neck tumors: evidence for a common clonal origin. *Cancer research* **56**, 2484-2487 (1996).
- 62 Johnson, B. E. Second lung cancers in patients after treatment for an initial lung cancer. *JNCI: Journal of the National Cancer Institute* **90**, 1335-1345 (1998).
- 63 Silvestri, G. A. *et al.* A Bronchial genomic classifier for the diagnostic evaluation of lung cancer. *New England Journal of Medicine* **373**, 243-251, doi:10.1056/nejmoa1504601 (2015).
- 64 Tang, X. *et al.* EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res* **65**, 7568-7572, doi:10.1158/0008-5472.CAN-05-1705 (2005).
- 65 Varella-Garcia, M. *et al.* Spectral karyotyping detects chromosome damage in bronchial cells of smokers and patients with cancer. *Am J Respir Crit Care Med* **176**, 505-512, doi:10.1164/rccm.200609-1329OC (2007).
- 66 Lydiatt, W. M. *et al.* Molecular support for field cancerization in the head and neck. *Cancer* **82**, 1376-1380 (1998).
- 67 Wistuba, II *et al.* Deletions of chromosome 3p are frequent and early events in the pathogenesis of uterine cervical carcinoma. *Cancer Res* **57**, 3154-3158 (1997).
- 68 Rahman, S. M. *et al.* Proteomic patterns of preinvasive bronchial lesions. *Am J Respir Crit Care Med* **172**, 1556-1562, doi:10.1164/rccm.200502-274OC (2005).

- 69 Sidransky, D. The oral cavity as a molecular mirror of lung carcinogenesis. *Cancer Prev Res* **1**, 12-14, doi:10.1158/1940-6207.capr-08-0093 (2008).
- 70 Bae, D.-J., Jun, J. A., Chang, H. S., Park, J. S. & Park, C.-S. Epigenetic Changes in Asthma: Role of DNA CpG Methylation. *Tuberculosis and Respiratory Diseases* **83**, 1, doi:10.4046/trd.2018.0088 (2020).
- 71 Wali, R. K. *et al.* Buccal microRNA dysregulation in lung field carcinogenesis: gender-specific implications. *Int J Oncol* **45**, 1209-1215, doi:10.3892/ijo.2014.2495 (2014).
- 72 Zhao, S., Allis, C. D. & Wang, G. G. The language of chromatin modification in human cancers. *Nat Rev Cancer* **21**, 413-430, doi:10.1038/s41568-021-00357-x (2021).
- 73 Almassalha, L. M. *et al.* The global relationship between chromatin physical topology, fractal structure, and gene expression. *Scientific reports* **7**, 41061 (2017).
- 74 Kemp, R. A. & Turic, B. Can early lung cancer be detected from buccal mucosa scrapings? *Chest* **128**, 154S, doi:10.1378/chest.128.4_meetingabstracts.154s-a (2005).
- 75 Prince, M., Lester, L., Chiniwala, R. & Berger, B. Multitarget stool DNA tests increases colorectal cancer screening among previously noncompliant Medicare patients. *World J Gastroenterol* **23**, 464-471, doi:10.3748/wjg.v23.i3.464 (2017).
- 76 Khanna, N. & Phillips, M. D. Adherence to care plan in women with abnormal Papanicolaou smears: a review of barriers and interventions. *J Am Board Fam Pract* **14**, 123-130 (2001).
- 77 Board, P. S. a. P. E. *Lung cancer screening (PDQ®)–Health professional version*, <<https://www.cancer.gov/types/lung/hp/lung-screening-pdq>> (2022).
- 78 Bach, P. B. *et al.* Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA* **307**, 2418-2429, doi:10.1001/jama.2012.5521 (2012).
- 79 National Lung Screening Trial Research, T. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* **365**, 395-409, doi:10.1056/NEJMoa1102873 (2011).