HHS Public Access

Author manuscript

Microsc Microanal. Author manuscript; available in PMC 2022 March 22.

Published in final edited form as:

Microsc Microanal. 2021 August ; 27(4): 878–888. doi:10.1017/S1431927621000507.

Early Upper Aerodigestive Tract Cancer Detection Using Electron Microscopy to Reveal Chromatin Packing Alterations in Buccal Mucosa Cells

Oisín Bugter[1,2,†], **Yue Li**[3,†], **Anouk H.G. Wolters**[4], **Vasundhara Agrawal**[3], **Amil Dravid**[5], **Andrew Chang**[3], **Jose Hardillo**[1], **Ben N.G. Giepmans**[4], **Robert J. Baatenburg de Jong**[1], **Arjen Amelink**[6,7], **Vadim Backman**[3], **Dominic J. Robinson**[1,2,*]

[1]Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus MC Cancer Institute, Dr. Molewaterplein 40, Postbus 2040, 3000 CA Rotterdam, the Netherlands

[2]Center for Optical Diagnostics and Therapy, Erasmus MC Cancer Institute, Dr. Molewaterplein 40, Postbus 2040, 3000 CA Rotterdam, the Netherlands

[3]Department of Biomedical Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

[4]Department of Biomedical Sciences of Cells and Systems, University of Groningen, University Medical Center Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, the Netherlands

[5]Department of Electrical Engineering and Computer Science, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

[6]Department of Optics, the Netherlands Organization for Applied Scientific Research (TNO), Stieltjesweg 1, 2628 CK Delft, the Netherlands

[7]Department of Physics and Astronomy, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands

Abstract

A profound characteristic of field cancerization is alterations in chromatin packing. This study aimed to quantify these alterations using electron microscopy image analysis of buccal mucosa cells of laryngeal, esophageal, and lung cancer patients. Analysis was done on normal-appearing mucosa, believed to be within the cancerization field, and not tumor itself. Large-scale electron microscopy (nanotomy) images were acquired of cancer patients and controls. Within the nuclei, the chromatin packing of euchromatin and heterochromatin was characterized. Furthermore, the chromatin organization was quantified through chromatin packing density scaling. A significant difference was found between the cancer and control groups in the chromatin packing density scaling parameter for length scales below the optical diffraction limit (200 nm) in both the euchromatin ($p = 0.002$) and the heterochromatin ($p = 0.006$). The chromatin packing scaling analysis also indicated that the chromatin organization of cancer patients deviated significantly

*Author for correspondence: Dominic J. Robinson, d.robinson@erasmusmc.nl.
†The first two authors contributed equally to this work.

Disclosure Statement. The authors declare no potential conflicts of interest.

from the control group. They might allow for novel strategies for cancer risk stratification and diagnosis with high sensitivity. This could aid clinicians in personalizing screening strategies for high-risk patients and follow-up strategies for treated cancer patients.

Keywords

buccal mucosa; chromatin packing; early detection; electron microscopy; field cancerization; upper aerodigestive tract cancer

Introduction

Head and neck (HN), esophageal, and lung cancer are all frequently occurring types of tumors (Ferlay et al., 2015; Siegel et al., 2015; Torre et al., 2015). They are often referred to as upper aerodigestive tract (UADT) tumors. Worldwide, an estimated more than three million new cases were diagnosed in 2018, and more than two million patients died of the consequences (Global Cancer Observatory). These three tumor types share the two most important risk factors for their development: smoking and for the HN region and esophagus also alcohol (Domper Arnal et al., 2015; Malhotra et al., 2016; Marur & Forastiere, 2016). Tobacco and alcohol use cause physiological and mutagenic effects on the exposed mucosa of the UADT. Unfortunately, a large percentage of UADT tumors are diagnosed in the advanced stages of development, often limiting the treatment options and survival chances of diseased patients (Kollarova et al., 2007; Horeweg & de Koning, 2014; Siegel et al., 2015). Diagnosing more tumors in an early stage of development could have a significant positive impact (Hall et al., 1999; Pennathur et al., 2013; Siegel et al., 2015). Patients with early stage UADT tumors could benefit from complete surgical resection or curative radiotherapy, whereas the treatment of patients with high-stage tumors is sometimes not with curative intent (Horeweg & de Koning, 2014).

Early tumor detection by screening asymptomatic high-risk patients holds the potential to increase the survival rate of lung cancer patients substantially. Upper aerodigestive tract tumors, in theory, appear ideally suited to such screening because of (a) the association of identifiable risk factors, (b) the survival advantage of early diagnosis, and (c) the significant morbidity and mortality associated with the disease. At present, there are no national screening programs for HN, esophageal, or lung cancer in Western countries. Some issues screening initiatives are facing is the low sensitivity for early stage tumors, the substantial population of at-risk persons, and the risk of overdiagnosing due to false-positive results (Domper Arnal et al., 2015; de Koning et al., 2020).

A novel yet effective approach to cancer screening is focused on detecting changes in apparently histologically normal tissue described as field cancerization (FC). Field carcinogenesis is the notion that a multitude of physiological and nanoscale architectural alterations affect an entire organ or tract before ultimately resulting in a focal neoplasm in one area of the organ (Roy et al., 2010). There is evidence that FC of HN, lung, and esophageal cancers encompasses the entire UADT (Kopelovich et al., 1999). This concept is supported by the high incidence of second primary tumors in patients with UADT tumors (Krishnatreya et al., 2013; van de Ven et al., 2019). The FC tissue alterations occur

superficially in the epithelial layer, the basal membrane, and the vascularized papillary layer of the lamina propria. They include alterations in the microvasculature and the tissue nanoscale architecture, such as the organization of the chromatin, the cytoskeleton, and the size and structure of cell nuclei and organelle (Subramanian et al., 2009*b*; Evers et al., 2012). The accurate detection of FC could potentially be used to pre-screen for UADT tumors (Kopelovich et al., 1999; Sidransky, 2008; Angadi et al., 2012). Optical techniques have proven to be sensitive to detect the changes in the nano-organization of tissue (Roy et al., 2011). The detection of FC could potentially be used as a risk-stratification method to decrease the population size of persons at risk for UADT tumors.

A technique to assess and visualize the nano-anatomy of tissue, in our case, buccal mucosa, is electron microscopy (EM). One of the downsides of EM imaging at high resolution is typically a small field of view of a given tissue sample, which makes it difficult to correlate changes at high magnification to a wide tissue scale. To tackle this problem, we applied large-scale EM (nanotomy), which allows for ultrastructural examination of tissue, cells, organelles, and macromolecules in a single dataset (Kuipers et al., 2016; de Boer et al., 2020). A nanotomy dataset combines thousands of conventional EM images. Moreover, the software allows zooming in and out of the image from a total overview to nanoscale resolution in a “Google Earth” approach. Using this technique, the large areas of tissue are scanned and presented online. A significant advantage of nanotomy over conventional EM is unbiased data acquisition, presentation, and sharing at high resolution (Kuipers et al., 2016; de Boer et al., 2020).

A promising target to analyze as a possible alteration due to FC is the quantification of chromatin organization in the cell nucleus. As the carrier of genetic information, chromatin forms and regulates the nano-environment in which transcription happens (Phillips-Cremins et al., 2013). Higher-order chromatin organization is essential in regulating gene transcription, and abnormalities in such an organization are associated with a variety of diseases, including neurological disorders, autoimmune diseases, and cancer. While nuclear blebbing and chromatin condensation (>200 nm) has been identified by optical microscopy as a hallmark for carcinogenesis, little is known about chromatin organization before this stage of progression of cancer at an even smaller length scale (Rizvi et al., 2015). Due to the optical diffraction limit of approximately 200 nm, it requires super-resolution optical microscopy, nanoscale-sensitive optical spectroscopy, or microscopy techniques that utilize different illumination sources with a wavelength smaller than light, such as neutrons, electrons, and X-ray (Gerchman & Ramakrishnan, 1987; Le Gros et al., 2016).

High-resolution electron tomography experiments with DNA-specific staining showed that chromatin consists of disordered polymers with a diameter ranging from 5 to 24 nm with a differential packing density throughout the nucleus (Ou et al., 2017). A classical polymer is expected to exhibit self-similar, fractal behavior across all length scales, and the fractal dimension is determined by the balance between polymer–polymer and polymer–solvent interaction as well as constraint-processes such as, in the case of chromatin, loop formation in part driven by extrusion, transcriptional, and lamin-associated processes (Flory, 1949; De Gennes & Gennes, 1979). For chromatin, it has been reported that the chromatin

polymer forms spatially segregated packing domains with sub-Mb genomic size, sub-200 nm physical size, and internal fractal structure (Huang et al., 2020; Virk et al., 2020*b*). The packing configuration plays an important role in regulating gene transcription by dictating chromatin accessibility and the diffusion rate of transcriptional reactants such as transcription factors (Li et al., 2020; Virk et al., 2020*a*). Importantly, shown through computational modeling, optical studies, and EM studies, the fractal dimension of chromatin packing domains has a non-monotonic relationship with active transcription. It also plays a critical role in the regulation of transcriptional plasticity of cells and their access to the genomic landscape (Li et al., 2020). A mass fractal can be characterized by a power-law spatial autocorrelation function (ACF), ACF $\sim r^{(D-3)}$, with D being the fractal dimension. Previous work has shown that the chromatin of more aggressive tumors exhibits an increase in the fractal dimension of the chromatin (Cherkezyan et al., 2014).

In this work, we analyzed normal-appearing buccal mucosa from patients with tumors elsewhere in the UADT (larynx, esophagus, and lungs). We hypothesized that the ACF and the fractal dimension D of patients with cancer differ from the control group. In order to utilize the large-scale EM data, we developed a novel convolutional neural network (CNN) to segment the nucleus from the tissue bed in grayscale images. We further quantified the fractal dimension D of the euchromatin and the heterochromatin independently to investigate ultrastructural alterations in the buccal mucosa cells of patients with UADT cancers and control subjects. If proven that cancer patients have an altered chromatin packing in the mucosa adjacent to their tumor, this may be a step toward the development of novel and sensitive tools for cancer screening.

Materials and Methods

Subjects

This study was approved by the Medical Ethics Committee of the Erasmus MC (MEC-2017–551). Patients were recruited from the departments of “Otorhinolaryngology and Head and Neck Surgery,” “Gastroenterology and Hepatology,” and “Pulmonology” of the Erasmus MC Cancer Institute. Patients with UADT malignancies and non-oncologic control patients were included. In this study, we also include the lungs under the “umbrella term” UADT. The oncologic group consisted of patients with primary and untreated HN (all laryngeal cancer), esophageal, and lung cancer (all subsites and stages). The non-oncologic control group consisted of patients with chronic rhinosinusitis with or without nasal polyps, chronic obstructive pulmonary disease, or gastrointestinal diseases. The absence of occult, unexpected malignancies in this group was confirmed using an endoscopic examination or imaging. Patients with a medical history of malignancies were excluded. All patients signed an informed consent form before enrollment in this study. Patient and tumor-specific data such as date of birth, sex, substance abuse, tumor stage, and tumor type were collected using the electronic medical patient record.

Biopsy Procedure

The buccal mucosa biopsies were performed at the outpatient clinics. The biopsies were performed on normal-appearing mucosa and at least 5 cm from the primary tumor. First,

local buccal mucosa anesthesia was given with a submucosal injection of approximately 1 mL xylocaine 2%—adrenaline 1:80.000. Second, the buccal mucosa was slightly turned "inside out" manually to have optimal visualization to perform a 2 mm Ø punch biopsy. The sample was cut from the subcutaneous tissue with a curved iris scissor without lifting the sample with a pair of forceps or tweezers. In this way, no external pressure was applied on the biopsy, maintaining the mucosa organization as optimally as possible before placing the biopsy in a fixative. Finally, the pressure was applied to the biopsy location with a singular gauze for approximately 1 min. Sutures were not needed.

Sample Preparation, Image Acquisition, and Automated Annotation

The sample preparation and image acquisition protocols were described in detail previously (Kuipers et al., 2016). In short, fresh buccal mucosa samples were immediately fixated in small vials in 3 mL solution of 0.5% paraformaldehyde, 2% glutaraldehyde, and 0.1 M sodium cacodylate (pH 7.4) and stored at 4°C. They were washed in 0.1 M sodium cacodylate and postfixed in 1% osmium tetroxide and 1.5% potassium ferrocyanide. The samples were dehydrated in ethanol by incubations in increasing ethanol concentrations. They were then embedded in epoxy resin and sectioned with a diamond knife to ultrathin sections of ~80 nm. Sections were mounted on formvar coated copper grids and stained with 2% uranyl acetate in water and Reynold lead citrate.

Image acquisition was performed with a Supra 55-VM scanning EM (Carl Zeiss AG, Oberkochen, Germany) with ATLAS software (Fibics Incorporated, Ottawa, Canada). We used scanning EM with an external scan generator capable of acquiring multiple large fields of view at high resolution using scanning transmission EM detection. One image generated this way is equivalent to the fields of view of ~100 transmission EM images, which significantly reduces the amount of stitching needed. Samples were recorded at a 2.5 nm pixel size. Scans were stitched, and raw datasets were rendered as HTML files.

Nucleus Segmentation Using a CNN

The CNN employed to segment the nucleus from the large-scale STEM images in this work was initially trained with TEM cheek cell electron micrographs by Dravid (2019) (Fig. 1). The CNN model is based on the Deep Residual U-Net, a residual learning framework for substantially deeper networks. It has shown significant success in segmentation tasks such as road extraction (Zhang et al., 2018). The Deep Residual U-net relies on an encoder–decoder structure proposed in the original U-Net encoder–decoder model, whereby an image is downsampled to its features of interest via the left branch of the model as seen in Figure 1a, and then upsampled through the right branch with skip connections propagating details from the original image to ultimately produce an accurate segmentation mask (Ronneberger et al., 2015). The inclusion of residual blocks allows the network to be deeper, leading to higher accuracy and less need for additional training data (He et al., 2016).

Each residual block consists of a sequence of batch normalization (BatchNorm), rectified linear unit (ReLU), and 3 × 3 convolution. Batch normalization reduces variance by scaling down the intermediate input values between operations. This speeds up training by reducing jumps in these values and allowing the optimization process to be smoother. The ReLU

operation adds nonlinearity to the model to allow for learning more complex features. This is an activation function, meaning that it maps intermediate values between operations in the model based on a nonlinear function, allowing the model to better discriminate pixels that may belong to the amorphous nucleus. The truncated ResBlock of Figure 1 uses 64 filters in each convolution. These individual filters learn to detect specific features of the image, such as curves or corners. The next two ResBlocks use convolutions with 128 and 256 filters, respectively, followed by another block using 512 filters. The decoder proceeds with the opposite pattern: 256, 128, then 64 filters. Overall, these convolutions learn features of the input TEM image that will contribute to the segmentation.

The network was trained for 30 epochs. By the end of every epoch, each training example has been fed into the network. Based on these training examples, the network adjusts the parameters of the internal operations in order to learn a better mapping between each micrograph and its corresponding segmentation. Multiple epochs, or passes through the entire dataset, are required for the optimization process to iteratively adjust the parameters. The number of epochs was set to 30 in order for the model to learn an effective mapping but avoid a neural network's tendency to "memorize" non-generalizable oddities specific to the training examples. This phenomenon, known as "overfitting," occurs when training for too many epochs. Two input micrographs were fed in simultaneously at a time to train the model faster than with one input. However, the amount of video memory limits how many inputs could be processed at a time to two. The Adam optimization algorithm was then employed to adjust the parameters in order to converge to an optimal segmentation model (Kingma & Ba, 2014). The learning rate was set to 10^{-5} to control the magnitude of each optimization step and the standard binary cross-entropy loss function was used for monitoring the effect of this optimization process on adjusting the model's performance.

Accuracy was evaluated using the Dice coefficient, which compares which pixels were correctly predicted to be the nucleus between the original input and generated segmentation mask. Using 300 training examples, an accuracy of 96% was achieved. This was implemented in the Keras library with a Tensorflow backend on a machine equipped with an NVIDIA GTX 1080, Core i7 CPU, and 32 GB of RAM.

The electron micrographs containing cheek buccal mucosa cells were manually selected for downstream nucleus segmentation. Limited by the available RAM, the images were first downsized by a factor of 4 before feeding them to the CNN. The output masks were upsampled to recover the original resolution. Active contour algorithm (Matlab, MathWorks) was used to refine the boundary of the nucleus mask based on the morphological features in the original EM images.

Spatial Autocorrelation Function

STEM images of ~80 nm thin sections were used in the analysis of chromatin packing density alterations between cancer patients and controls. The bright-field contrast in STEM attenuates following Beer's law,

$$I(x, y) = I_0 e^{-\sigma \rho(x, y) t}, \quad (1)$$

where $I(x, y)$ is the STEM bright-field image intensity distribution, I_0 is the incident beam intensity, σ is the absorption coefficient of the biological sample to the incident beam, $\rho(x, y)$ is the density distribution, and t is the section thickness. In our experiment, I_0 and t were controlled to be constant for all images, and we assumed that the biological sample has a relative constant σ given current resolution. As a result, only the chromatin density $\rho(x, y)$ contributes to the final image intensity $I(x, y)$. To obtain the density fluctuation function, $\rho_\Delta(x, y)$, we took the negative logarithm of the STEM images and then subtracted the mean value from the image. At the same time, the incident beam intensity I_0 is canceled out.

Next, the two-dimensional ACF is calculated from the density fluctuation obtained from the STEM images using the Wiener–Khinchine relation (Cherkezyan et al., 2014):

$$B_\rho(x, y) = F^{-1}\left\{\left|F(\rho_\Delta(x, y))\right|^2\right\}, \tag{2}$$

where F^{-1} and F are the inverse Fourier and the Fourier transform, respectively, and the ρ_Δ represents the fluctuations in the chromatin density. To minimize the noise, a rotational average of $B_\rho(x, y)$ was taken to obtain the final form of the ACF, $B_\rho(r)$, representing the correlation of chromatin density as a function of spatial separation r. Notice that mathematically, a fractal structure can be characterized by a power-law ACF, $B_\rho(r) \sim r^{D-3}$, with D being the fractal dimension. To analyze the chromatin packing structure from the experimental ACF obtained, we fit the ACF to the Whittle–Matérn (WM) family of correlation functions (Rogers et al., 2014). WM is defined as the product of a power-law and a modified Bessel function of the second kind ($K_{(D-3)/2}$) of the order (D–3)/2:

$$B_\rho(r) = A_\rho\left(\frac{r}{l_n}\right)^{D-3/2} K_{D-3/2}\left(\frac{r}{l_n}\right). \tag{3}$$

In (3), A_ρ, l_n, and D are fitting parameters. A_ρ is the density fluctuation amplitude; l_n is correlation length indicating the characteristic length of chromatin heterogeneity; the dimension D controls the shape of ACF, such that $D \to \infty$ for Gaussian; $D = 4$ for exponential; $3 < D < 4$ for stretched exponential; and $D < 3$ for power-law shape of the ACF. Particularly, when $D < 3$, the biological medium can be considered a mass fractal and D takes the special meaning of the fractal dimension.

Each nucleus was segmented using the CNN. As heterochromatin and euchromatin are genetically distinct and have a different affinity to osmium and uranyl/lead staining, they were segmented from the nucleus using automated grayscale thresholding. In special cases where the chromatin in the cell nucleus was composed primarily of the euchromatin, we did not conduct ACF analysis for the heterochromatin, as the resulting ACF would not have been representative of the true statistics of the chromatin packing. Mean ACF was calculated from all the cells of the same patient for both euchromatin and heterochromatin. For each mean ACF, the D value for euchromatin and heterochromatin was obtained by WM fitting from $r = 79$ nm to $r = 200$ nm in correspondence to the section thickness. In particular, the fitting range of D is constrained by $5/3 \leq D \leq 4$, the boundary values represent physiological values reported in published work (Cherkezyan et al., 2014). In addition, the D value was

also calculated for each cell. Overall, 253 cells from 20 patients were analyzed (control: $N_{patient} = 6$, $N_{cell} = 68$; lung cancer: $N_{patient} = 1$, $N_{cell} = 10$; esophagus cancer: $N_{patient} = 6$, $N_{cell} = 78$; HN cancer: $N_{patient} = 7$, $N_{cell} = 97$). Figure 2 summarizes the methods utilized for the image analysis and evaluating D.

Statistical Analysis

The quantitative variables were non-normally distributed, and thus, the values were expressed as the median and interquartile range (IQR). Differences between groups were analyzed with the Mann–Whitney U-test. Categorical data were reported as frequencies and percentages and differences between groups were analyzed using the chi-squared test or Fisher's exact test. For ACF analysis, a two-sample Kolmogorov–Smirnov (K-S) test was performed to evaluate the difference of ACF values between the control and cancer groups, as the K-S test is sensitive not only the median of the ACFs but also to the shape, which is the characteristic of the underlying structure of chromatin packing. For D values calculated from mean ACFs per patient, a two-sided Wilcoxon rank-sum test was employed. For D values calculated from individual ACF per cell, student t-test was used. Statistical analysis was performed using SPSS (IBM Corp., Armonk, NY, USA) and MATLAB (MathWorks, Natick, MA, USA). The cut-off for significance was set at $p < 0.05$.

Results

Large-Scale EM Datasets for Buccal Mucosa Biopsies

Normal appearing buccal mucosa biopsies from 20 patients were included in this study: 14 patients with cancer (7 HN, 6 esophageal, and 1 lung cancer) and 6 non-oncologic controls. Their baseline characteristics are presented in Table 1. There was a non-significant difference in the percentage of males between the cancer and control groups (66.7 and 78.6%). The median age between the cancer and control groups was also similar (69.0 [IQR 66.0–72.8] versus 62.0 [IQR 57.2–74.5] years). Only the HN cancer group had a significantly higher amount of pack years than the control group. Six patients (42.9%) of the cancer group had a stage I, two patients (14.3%) a stage II, five patients (35.7%) a stage III, and one patient (7.1%) a stage IV carcinoma.

The large-scale EM datasets of the buccal mucosa biopsies are available to view online at http://www.nanotomy.org/OA/. The nanotomy images show the non-keratinizing epithelial layer of the buccal mucosa and the superficial part of the lamina propria including the capillaries. The basal layer is located in between the two layers (Fig. 3).

Nucleus Segmentation Pipeline with a Pre-Trained CNN

Figure 1 shows the implemented residual U-Net architecture. A 640 × 640 × 1 grayscale EM image with a cheek cell nucleus is downsampled through convolutions within each residual block, ResBloc (Fig. 1b) is employed to obtain a smaller-scale image consisting of the features of interest that can benefit the accuracy of the model. The right side of the model CNN consists of upsampling operations, which restore the size of the image. The skip connections pass fine details, which are lost through downsampling, over to the right side of

the network to relay the structure of the STEM micrograph in the final binary segmentation mask.

Characterizing Chromatin Packing Using ACF for Cheek Cells in the Control and Cancer Groups

We utilized the grayscale image intensity in the STEM micrographs to characterize the spatial heterogeneity of chromatin density distribution. We quantified the relative magnitudes and length scales of all spatial fluctuations in the degree of chromatin compaction via its ACF, $B_\rho(r)$. Upon comparison of chromatin density correlation between the controls and cases representing various cancers, we have found a qualitative difference in chromatin packing at subdiffractional length scales (Figs. 4a, 4b). Using K-S tests, we confirmed that the ACFs for the overall control ($n = 6$) and cancer groups ($n = 14$) are statistically significantly different ($p < 0.001$) for both the euchromatin and the heterochromatin at all length scales. The mean ACFs for individual diagnostic groups are shown in Figures 4c and 4d.

Quantifying Chromatin Packing Alterations Using Packing Scaling Factor D

We further quantified the chromatin packing scaling D below the optical diffraction limit in terms of D by fitting the ACFs to WM-family of functionals (Fig. 5). The fitting range was chosen to be from 80 to 200 nm (optical diffraction limit). As the resin sections for STEM imaging have a similar finite thickness of around 80 nm, the ACF will overestimate the correlation of chromatin density fluctuation below this length scale due to the projection average. For the overall control group, we obtained a median D value of 1.72 (IQR 1.67–1.90) for the euchromatin and 2.24 (IQR 1.67–2.41) for the heterochromatin. For the overall cancer group, we obtained a median D value of 2.24 (IQR 2.12–2.48) for the euchromatin and 3.00 (IQR 2.52–4) for the heterochromatin. We employed Wilcoxon rank-sum to quantify the alteration in D distribution for the control and the cancer groups and obtained a significant statistical difference for both the euchromatin ($p = 0.002$) and the heterochromatin ($p = 0.005$). Particularly, $5/3 < D < 3$ is indicative of underlying massfractal chromatin structure, $D > 3$ suggests a non-fractal distribution of chromatin density that is consistent with a stretched exponential function. Significantly, we observed that cancer patients showed an increase in median packing scaling D by 26.6% for the euchromatin, and by 33.9% for the heterochromatin. In addition, the median value of D for the heterochromatin of the cancer patients equals 3, suggesting a fundamental alteration in chromatin packing, which can have a substantial effect in regulating gene transcription and phenotypic plasticity (Virk et al., 2020*a*). Since a small number of patients were employed in the analysis, we further quantified the chromatin packing scaling for every cell within each group and evaluated the statistical difference between control patients and cancer patients (Figs. 5e, 5f). Overall, we observed a significant difference between chromatin packing scaling for the control group ($n = 76$) and the cancer group ($n = 219$) in both the euchromatin ($p = 0.010$) and the heterochromatin ($p = 0.007$).

Discussion

This study is, to our knowledge, the first to utilize large-scale EM images to investigate FC tissue changes associated with cancer of the UADT. We aimed to investigate whether these changes could be used to discriminate cancer patients from controls. Utilizing spatial correlation function analysis, we characterized chromatin packing below the optical diffraction limit and identified significant changes between the control and cancer groups in both the euchromatin and the heterochromatin contents. Moreover, we showed that the difference in fractal dimensions calculated from the spatial correlation function agrees with the well-known hallmarks of cancer but manifested at a much smaller length scale. These results highlight that the alterations in chromatin observed in the field of a tumor represent an early stage event of carcinogenesis.

Elevated D is a hallmark of cancer cells and has been reported through multiple lines of work using Partial-Wave Spectroscopic (PWS) microscopy in colorectal, lung, and breast cancers (Subramanian et al., 2009*a*; Gladstein et al., 2018). From a theoretical perspective, the spatial arrangement of chromatin packing affects large-scale gene expression patterns through a number of physical regulators, such as chromatin volume concentration, accessible surface area, and chromatin packing scaling D (Huang et al., 2020; Virk et al., 2020*a*). Particularly, chromatin with an increased D in the tumor field can select for a higher transcriptional and adaptive potential. This ability, in turn, may facilitate the prognosis of tumor cells independent of tumor type, stage, demographic factors, and molecular transformations (Virk et al., 2020*a*). Recently, utilizing EM, Cherkezyan et al. reported the existence of significant alterations in chromatin packing in the colorectal cancer field in both human and animal models (Cherkezyan et al., 2014). They identified a profound shift of chromatin organization: the chromatin of cancer subjects had chromatin with a higher fractal dimension or, in some cases, adopted a non-fractal configuration, while the control subjects had a normal fractal chromatin structure. In the same vein, optical nanosensing showed an increase in the fractal dimension of chromatin in the field cancerization associated with a variety of malignancies (Bauer et al., 2017). In the present study, we have also identified similar alterations in chromatin packing comparing the cancer group with the control group.

A previous series of studies by our group had the same aim as the present study: discriminating cancer patients from controls by detecting FC tissue changes (Bugter et al., 2018*a*, 2018*b*, 2019). In these studies, *in vivo* multidiameter single-fiber reflectance (MDSFR) spectroscopy measurements were performed on the buccal mucosa of a larger cohort of UADT cancer patients, including 20 included in the present study. This provides an opportunity to compare the EM images analysis results with the scattering results from optical spectroscopy. A logical hypothesis would be that an increased D in cell nuclei will correlate with an increased power of scattering of light during spectroscopic measurements. Though not significantly different, we indeed found a tendency for the scattering power of μ's to be higher for cancer patients. Possibly, this difference was not significant because, although the patients included were the same, the measured tissue volumes were not. In the EM study, the cell nuclei were segmented and isolated, while in the spectroscopic studies, the optical properties of the complete cells of the mucosal top layer were averaged. By studying the ultrastructural changes in chromatin organization, it is possible to detect the

initial stages of various kinds of cancers. It is important to note that in this work, instead of a single cancer model, we assessed three types of cancers and their controls: head/neck, esophageal, and lung cancer. The direction of alterations in chromatin packing in terms of D is consistent for all cancer groups compared with the control group. This probably reflects the fact that these three organs that encompass the UADT all have their embryological origin in the early foregut (Metzger et al., 2011). The UADT also encompasses the oral cavity which is a predominant and prevalent site of development of (pre)malignancies, because it comes into direct contact with many carcinogens.

Although not explored in this work, there are several implications of an increasing D in chromatin packing in FC. Besides buccal mucosa cells investigated in this work, there is a plethora of different types of cells along the spectrum of differentiation within buccal mucosa imaged in the large-scale EM dataset. A similar trend in chromatin packing for basal and pickle layers is expected. Therefore, this dataset gives us the opportunity to study the effects of differentiation progression on higher-order chromatin organization. It is expected that as the phenotypic plasticity decreases during differentiation, the chromatin packing scaling would also decrease. Although the CNN was trained to segment the nucleus of epithelial cells, transfer learning can be implemented to tune the CNN to adapt quickly to other types of cells.

The clinical applicability of the method and results described in the present study will probably not be in the shape of a diagnostic tool for routine use. At present, the analysis is too time-consuming and not cost-effective. However, our findings did confirm the presence of ultrastructural field effect changes in the buccal mucosa of patients with distant UADT tumors and that these changes can be used to differentiate them from non-oncologic controls. This paves the way for existing optical techniques, like MDSFR spectroscopy or PWS microscopy to utilize and quantify these changes so that they may be applied in clinical practice. These techniques are easy-to-use, fast, and noninvasive and might be used to screen for distant tumors or aid surgeons in achieving adequate tumor resection margins.

There are several limitations to our study design that should be considered. First, the number of patients included was limited, and more patients need to be recruited to validate our current findings. This was especially true for the lung cancer group with only one patient. Second, the segmentation of the euchromatin and the heterochromatin relies solely on the EM image intensity. Labeling molecular markers, such as histone modifications, are required for a more rigorous and accurate separation of those two compartments. The resin sections of tissue biopsy in our study have a finite thickness of ~80 nm, and projections instead of 3D tomography of the sections were used. Due to the intensity averaging along the z-direction, the ACFs calculated from the projection images do not reflect the chromatin packing at length scales below the thickness of the section, and therefore, restricted our analysis to larger length scales. In order to investigate the chromatin alterations at a finer scale, future studies can incorporate thinner sections or electron tomography.

Conclusions

We identify significant quantitative nanoscale alterations in chromatin packing in FC for UADT cancers. Through large-scale EM datasets and high-throughput image processing using CNN, we confirm that the ultrastructural field effect changes of the nuclear organization are a hallmark of cancer. We propose that the CNN segmentation pipeline and the downstream nanoscale nuclear abnormalities identified here can be employed as biomarkers for FC. The large-scale EM nanotomy datasets combined with semi-automated data analysis might overcome previous clinical applicability issues, such as the long duration of the analysis. Optical techniques like MDSFR spectroscopy or PWS microscopy could also utilize the abnormalities in chromatin packing for the diagnosis and risk stratification of cancer.

Acknowledgments.

The authors have no further people to acknowledge for any contributions to this manuscript.

Funding. The Dutch Cancer Society (2014–7074), ZonMw (40–43500-98–4121 and 91111.006) [a Dutch Technology for Life Sciences (DTL) enabling technology grant], and the National Institutes of Health (R01CA200064, R33CA225323, and R01CA225002) partly funded this study. They had no role in the study conception, design, data collection, analysis, data interpretation, or writing of the manuscript. The corresponding author had full access to all data and final responsibility for the decision to submit for publication.

References

Angadi PV, Savitha JK, Rao SS & Sivaranjini Y (2012). Oral field cancerization: Current evidence and future perspectives. Oral Maxillofac Surg 16(2), 171–180. [PubMed: 22354325]

Bauer GM, Stypula-Cyrus Y, Subramanian H, Cherkezyan L, Viswanathan P, Zhang D, Iyengar R, Bagalkar S, Derbas J, Graff T, Gladstein S, Almassalha LM, Chandler JE, Roy HK & Backman V (2017). The transformation of the nuclear nanoarchitecture in human field carcinogenesis. Future Sci OA 206, 2017–0027.

Bugter O, Hardillo JA, Baatenburg de Jong RJ, Amelink A & Robinson DJ (2018a). Optical pre-screening for laryngeal cancer using reflectance spectroscopy of the buccal mucosa. Biomed Opt Express 9(10), 4665–4678. [PubMed: 30319894]

Bugter O, Spaander MCW, Bruno MJ, Baatenburg de Jong RJ, Amelink A & Robinson DJ (2018b). Optical detection of field cancerization in the buccal mucosa of patients with esophageal cancer. Clin Transl Gastroenterol 9 (4), 152. [PubMed: 29712897]

Bugter O, van Brummelen SE, van der Leest KH, Aerts J, Maat A, Baatenburg de Jong RJ, Amelink A & Robinson DJ (2019). Towards the optical detection of field cancerization in the buccal mucosa of patients with lung cancer. Trans Oncol 12(12), 1533–1538.

Cherkezyan L, Stypula-Cyrus Y, Subramanian H, White C, Cruz MD, Wali RK, Goldberg MJ, Bianchi LK, Roy HK & Backman V (2014). Nanoscale changes in chromatin organization represent the initial steps of tumorigenesis: A transmission electron microscopy study. BMC Cancer 14(1), 189. [PubMed: 24629088]

de Boer P, Pirozzi NM, Wolters AHG, Kuipers J, Kusmartseva I, Atkinson MA, Campbell-Thompson M & Giepmans BNG (2020). Large-scale electron microscopy database for human type 1 diabetes. Nat Commun 11(1), 2475. [PubMed: 32424134]

De Gennes P- G & Gennes P-G (1979). Scaling Concepts in Polymer Physics. Ithaca, NY: Cornell University Press.

de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, Lammers JJ, Weenink C, Yousaf-Khan U, Horeweg N, van ‘t Westeinde S, Prokop M, Mali WP, Mohamed Hoesein FAA, van Ooijen PMA, Aerts J, den Bakker MA, Thunnissen E, Verschakelen J, Vliegenthart R, Walter JE, Ten Haaf K, Groen HJM & Oudkerk M (2020). Reduced lung-cancer

mortality with volume CT screening in a randomized trial. N Engl J Med 382(6), 503–513. [PubMed: 31995683]

Domper Arnal MJ, Ferrandez Arenas A & Lanas Arbeloa A (2015). Esophageal cancer: Risk factors, screening and endoscopic treatment in western and eastern countries. World J Gastroenterol 21(26), 7933–7943. [PubMed: 26185366]

Dravid A (2019). Employing deep networks for image processing on small research datasets. Microsc Today 27(1), 18–23.

Evers D, Hendriks B, Lucassen G & Ruers T (2012). Optical spectroscopy: Current advances and future applications in cancer diagnostics and therapy. Future Oncol 8(3), 307–320. [PubMed: 22409466]

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D & Bray F (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136, E359–E386. [PubMed: 25220842]

Flory PJ (1949). The configuration of real polymer chains. J Chem Phys 17(3), 303–310.

Gerchman S & Ramakrishnan V (1987). Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy. Proc Natl Acad Sci USA 84(22), 7802–7806. [PubMed: 3479765]

Gladstein S, Damania D, Almassalha LM, Smith LT, Gupta V, Subramanian H, Rex DK, Roy HK & Backman V (2018). Correlating colorectal cancer risk with field carcinogenesis progression using partial wave spectroscopic microscopy. Cancer Med 7(5), 2109–2120. [PubMed: 29573208]

Global Cancer Observatory (GCO). https://gco.iarc.fr/.

Hall SF, Groome PA, Rothwell D & Dixon PF (1999). Using TNM staging to predict survival in patients with squamous cell carcinoma of head & neck. Head Neck 21(1), 30–38. [PubMed: 9890348]

He K, Zhang X, Ren S & Sun J (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Horeweg N & de Koning H (2014). The importance of screening for lung cancer. Expert Rev Respir Med 8(5), 597–614. [PubMed: 25158921]

Huang K, Li Y, Shim AR, Virk RKA, Agrawal V, Eshein A, Nap RJ, Almassalha LM, Backman V & Szleifer I (2020). Physical and data structure of 3D genome. Sci Adv 6(2), eaay4055.

Kingma DP & Ba J (2014). Adam: A method for stochastic optimization. arXiv:1412.6980.

Kollarova H, Machova L, Horakova D, Janoutova G & Janout V (2007). Epidemiology of esophageal cancer–An overview article. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 151(1), 17–20. [PubMed: 17690734]

Kopelovich L, Henson DE, Gazdar AF, Dunn B, Srivastava S, Kelloff GJ & Greenwald P (1999). Surrogate anatomic/functional sites for evaluating cancer risk: An extension of the field effect. Clin Cancer Res 5(12), 3899–3905. [PubMed: 10632318]

Krishnatreya M, Rahman T, Kataki AC, Das A, Das AK & Lahkar K (2013). Synchronous primary cancers of the head and neck region and upper aero digestive tract: Defining high-risk patients. Indian J Cancer 50(4), 322–326. [PubMed: 24369209]

Kuipers J, Kalicharan RD, Wolters AH, van Ham TJ & Giepmans BN (2016). Large-scale scanning transmission electron microscopy (nanotomy) of healthy and injured zebrafish brain. J Vis Exp 111, 53635.

Le Gros MA, Clowney EJ, Magklara A, Yen A, Markenscoff-Papadimitriou E, Colquitt B, Myllys M, Kellis M, Lomvardas S & Larabell CA (2016). Soft X-ray tomography reveals gradual chromatin compaction and reorganization during neurogenesis in vivo. Cell Rep 17(8), 2125–2136. [PubMed: 27851973]

Li Y, Eshein A, Virk R, Eid A, Wu W, Frederick J, VanDerway D, Gladstein S, Huang K & Anthony N (2020). Nanoscale chromatin imaging and analysis (nano-ChIA) platform bridges 4-D chromatin organization with molecular function. Sci Adv 6(2), eaay4055.

Malhotra J, Malvezzi M, Negri E, La Vecchia C & Boffetta P (2016). Risk factors for lung cancer worldwide. Eur Respir J 48(3), 889–902. [PubMed: 27174888]

Marur S & Forastiere AA (2016). Head and neck squamous cell carcinoma: Update on epidemiology, diagnosis, and treatment. Mayo Clin Proc 91(3), 386–396. [PubMed: 26944243]

Metzger R, Wachowiak R & Kluth D (2011). Embryology of the early foregut. Semin Pediatr Surg 20(3), 136–144. [PubMed: 21708333]

Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH & O'shea CC (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science 357(6349), eaag0025.

Pennathur A, Gibson MK, Jobe BA & Luketich JD (2013). Oesophageal carcinoma. Lancet 381(9864), 400–412. [PubMed: 23374478]

Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong C-T, Hookway TA, Guo C & Sun Y (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153(6), 1281–1295. [PubMed: 23706625]

Rizvi MA, Zaki M, Afzal M, Mane M, Kumar M, Shah BA, Srivastav S, Srikrishna S, Peerzada GM & Tabassum SJ (2015). Nuclear blebbing of biologically active organoselenium compound towards human cervical cancer cell (HeLa): In vitro DNA/HSA binding, cleavage and cell imaging studies. Eur J Med Chem 90, 876–888. [PubMed: 25535953]

Rogers JD, Radosevich AJ, Yi J & Backman V (2014). Modeling light scattering in tissue as continuous random media using a versatile refractive index correlation function. Selected topics in quantum electronics. IEEE J Sel Top Quantum Electron 20(2), 173–186.

Ronneberger O, Fischer P & Brox T (2015). U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241. Springer.

Roy HK, Hensing T & Backman V (2011). Nanocytology for field carcinogenesis detection: Novel paradigm for lung cancer risk stratification. Future Oncol 7(1), 1–3. [PubMed: 21174531]

Roy HK, Subramanian H, Damania D, Hensing TA, Rom WN, Pass HI, Ray D, Rogers JD, Bogojevic A, Shah M, Kuzniar T, Pradhan P & Backman V (2010). Optical detection of buccal epithelial nanoarchitectural alterations in patients harboring lung cancer: Implications for screening. Cancer Res 70(20), 7748–7754. [PubMed: 20924114]

Sidransky D (2008). The oral cavity as a molecular mirror of lung carcinogenesis. Cancer Prev Res (Phila) 1(1), 12–14. [PubMed: 19138931]

Siegel RL, Miller KD & Jemal A (2015). Cancer statistics, 2015. CA Cancer J Clin 65(1), 5–29. [PubMed: 25559415]

Sokol E, Kramer D, Diercks GFH, Kuipers J, Jonkman MF, Pas HH & Giepmans BNG (2015). Large-scale electron microscopy maps of patient skin and Mucosa provide insight into pathogenesis of blistering diseases. J Invest Dermatol 135(7), 1763–1770. [PubMed: 25789704]

Subramanian H, Pradhan P, Liu Y, Capoglu IR, Rogers JD, Roy HK, Brand RE & Backman V (2009a). Partial-wave microscopic spectroscopy detects subwavelength refractive index fluctuations: An application to cancer diagnosis. Opt Lett 34(4), 518–520. [PubMed: 19373360]

Subramanian H, Roy HK, Pradhan P, Goldberg MJ, Muldoon J, Brand RE, Sturgis C, Hensing T, Ray D, Bogojevic A, Mohammed J, Chang JS & Backman V (2009b). Nanoscale cellular changes in field carcinogenesis detected by partial wave spectroscopy. Cancer Res 69(13), 5357–5363. [PubMed: 19549915]

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J & Jemal A (2015). Global cancer statistics, 2012. CA Cancer J Clin 65(2), 87–108. [PubMed: 25651787]

van de Ven S, Bugter O, Hardillo JA, Bruno MJ, Baatenburg de Jong RJ & Koch AD (2019). Screening for head and neck second primary tumors in patients with esophageal squamous cell cancer: A systematic review and meta-analysis. United European Gastroenterol J 7(10), 1304–1311.

Virk R, Wu W, Almassalha LM, Bauer GM, Li Y, VanDerway D, Frederick J, Zhang D, Eshein A & Szleifer I (2020a). Disordered chromatin packing regulates ensemble gene expression and phenotypic plasticity. Biophys J 118 (3), 549a–550a.

Virk RKA, Wu W, Almassalha LM, Bauer GM, Li Y, VanDerway D, Frederick J, Zhang D, Eshein A, Roy HK, Szleifer I & Backman V (2020b). Disordered chromatin packing regulates phenotypic plasticity. Sci Adv 6(2), eaax6232.

Zhang Z, Liu Q & Wang Y (2018). Road extraction by deep residual U-net. IEEE Geosci Remote Sens Lett 15(5), 749–753.

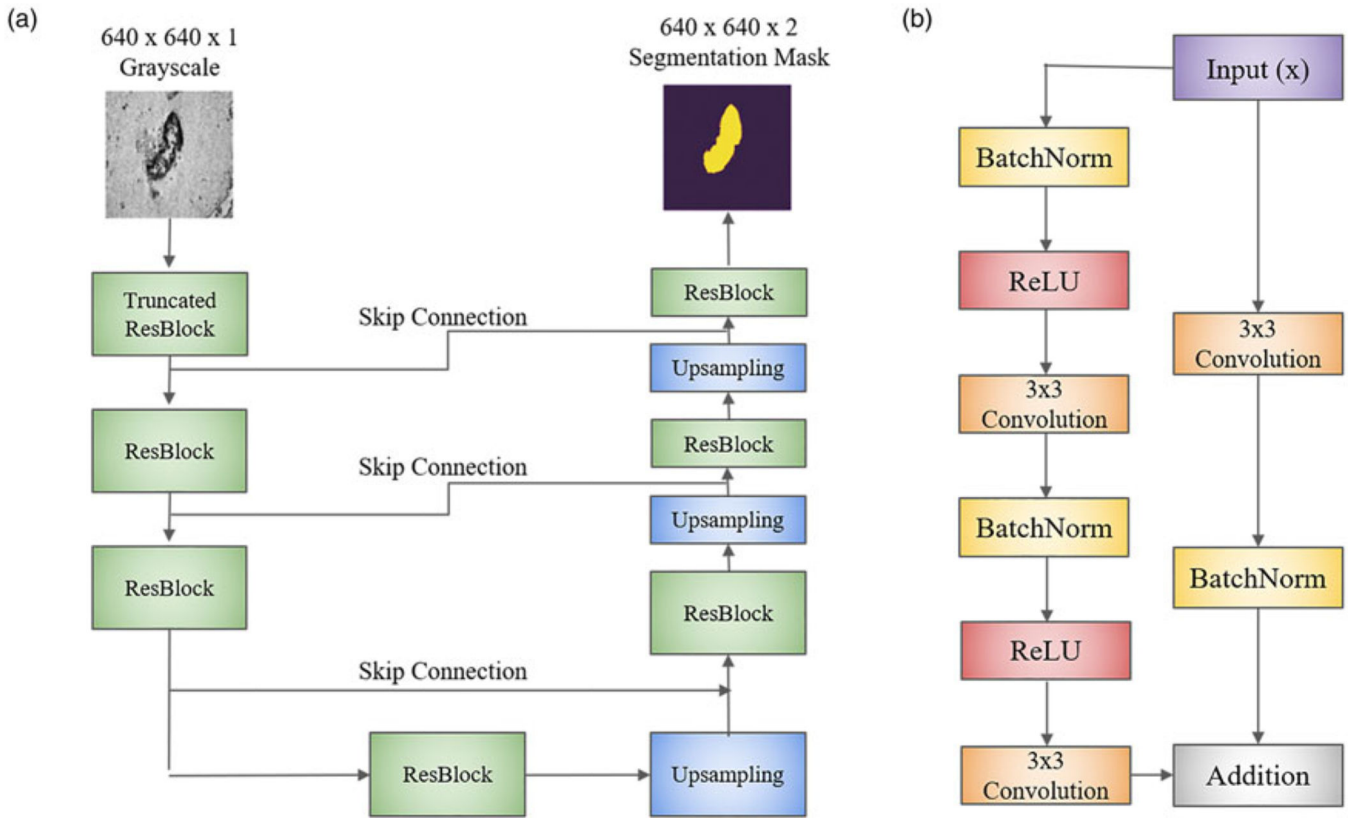

Fig. 1.

Architecture of the CNN employed in segmenting the nucleus. (**a**) The U-net is one classic way to arrange these operations in a logical way for transforming images, that is, segmenting, denoising, or performing super-resolution. In a U-net, several convolutional blocks with nonlinear functions at the end, referred to as Res-blocks (**b**) in the figure which will be detailed next, are arranged in a sequence. After each block, the image is downsampled which allows for convolution to be performed at a higher and higher level in the image. After three convolutions and downsamples, the transformed image is then passed to the right-hand side of the network and upsampled iteratively. After each upsample, the fine details are passed back into the image through a skip-connection before being convolved with a new set of filters and output into a binary mask. Figure reproduced from Virk et al. (2020b) with permission.

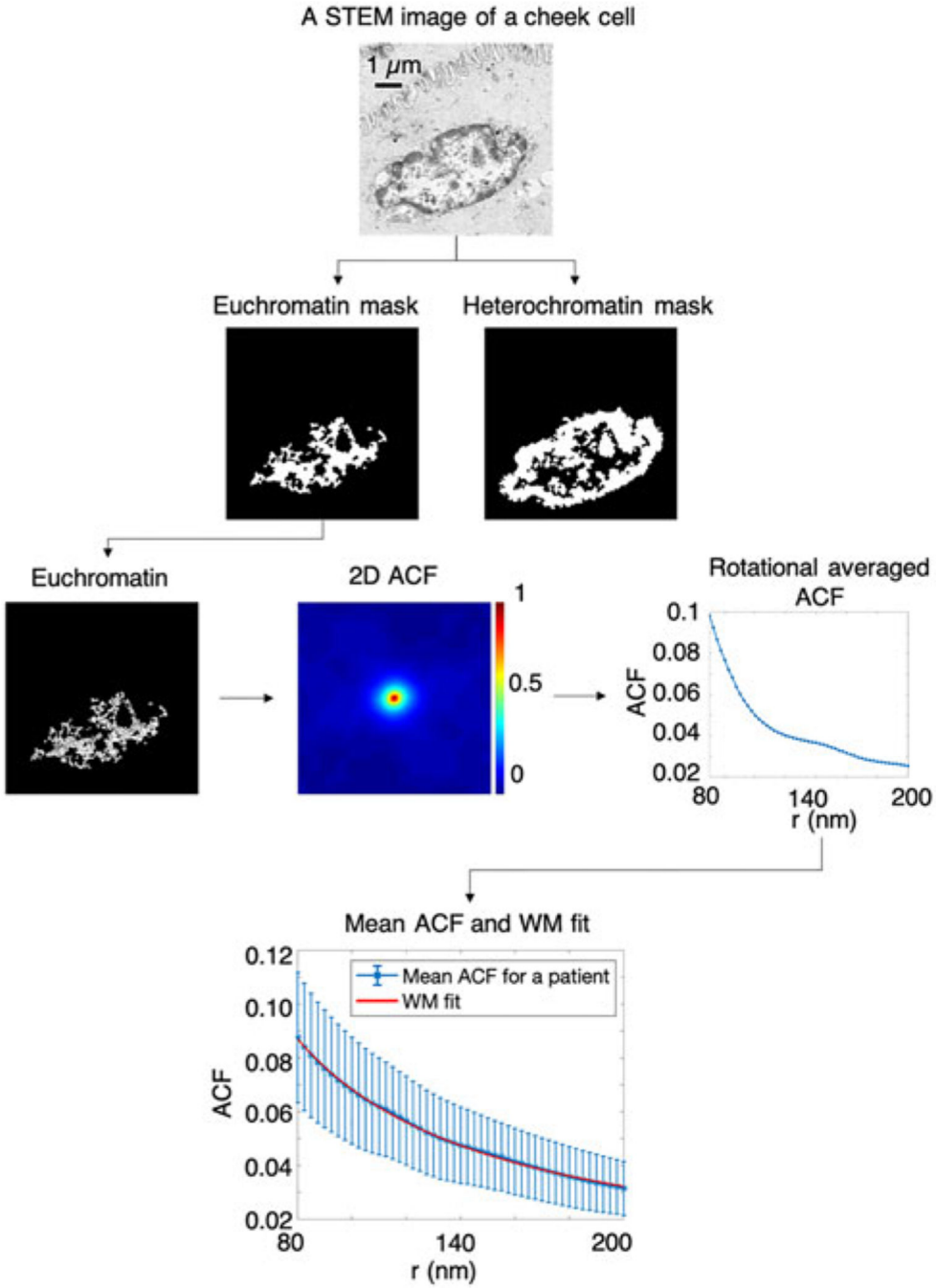

Fig. 2.

Workflow of obtaining chromatin packing *D* from an STEM image of a cheek cell. The nucleus is segmented using the CNN trained specifically for this task with electron micrographs. The masks for euchromatin and heterochromatin are further created from the mask for the nucleus using automatic grayscale intensity thresholding. While the heterochromatin is mainly located in the periphery of the nucleus, some penetrates the nuclear interior space. On the other hand, the euchromatin primarily distributes in the interior of the nucleus. 2D ACF is calculated for both euchromatin and heterochromatin

separately, and the rotational average is employed to remove the noise from the ACF curve. Mean ACF is obtained by averaging ACFs of individual cells for each patient, and WM fitting is used to quantify chromatin packing in terms of D. The fitting range is from r = 80 nm to r = 200 nm, as the section has a finite thickness of 80 nm, and the ACF curve is smoothed for length scales below 80 nm due to projection.

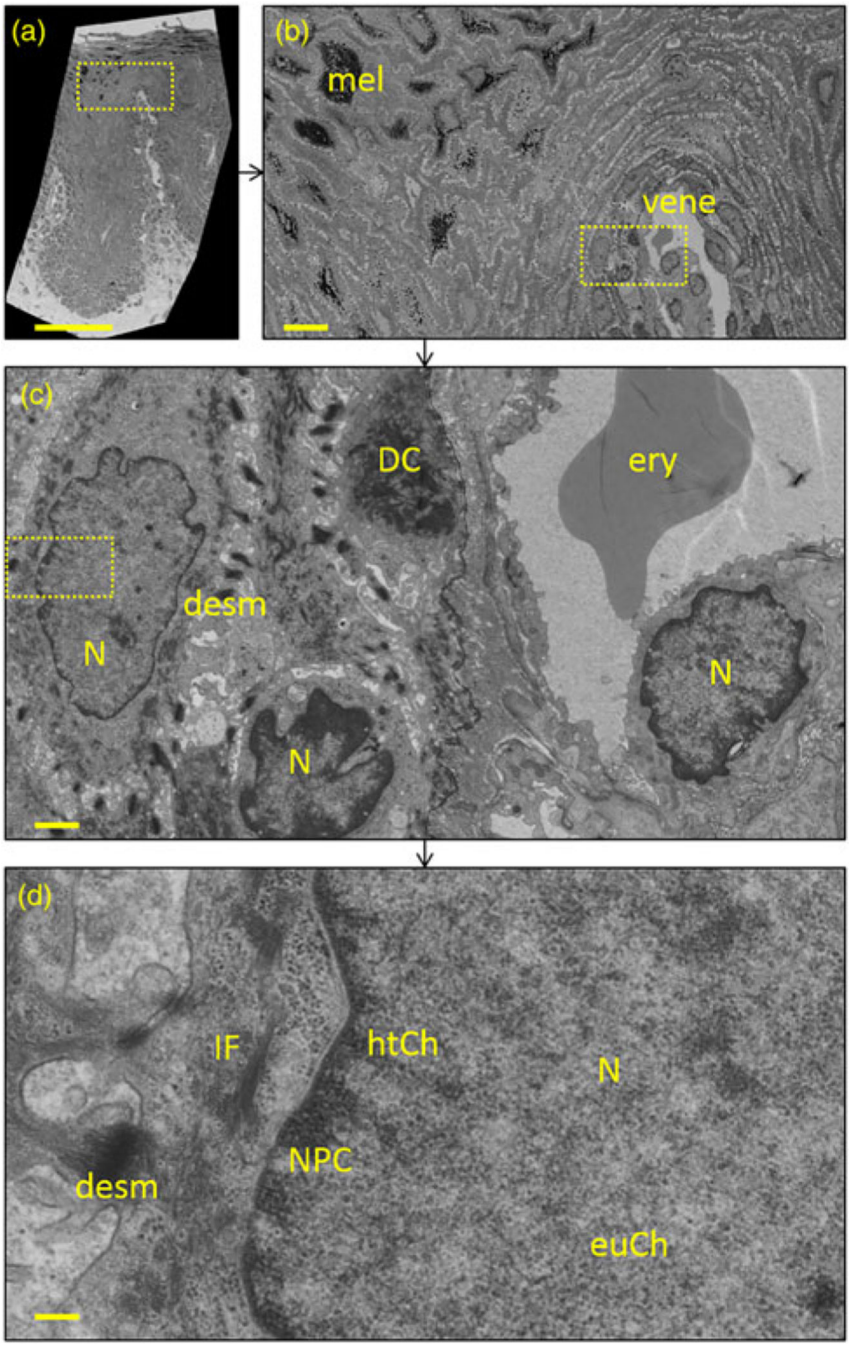

Fig. 3.

Nanotomy of buccal mucosa of an esophageal cancer patient. A large area of 0.2 × 0.4 mm is recorded that 2.5 nm pixel size. This allows to compare histological known landmarks up to macromolecular complexes. (**a**) Overview allows to discriminate layers of the epithelium of the mucosa: stratum basale, stratum spinosum, stratum intermedium, and stratum superficiale (for reference to the different layers, see Sokol et al., 2015). (**b**) Zooming in allows to recognize a vene, melanocytes (mel), and other mesoscale structures. (**c**) Further zooming discerns other cell types such as an erythrocyte (ery); organelles such

as nuclei (N), desmosomes (desm), as well as a desmosome cluster (DC). At the maximum resolution, macromolecules such as intermediate filaments (IUF), nuclear pore complexes (NPC), as well as euchromatin (euCh) and heterochromatin (htCh) can be identified. Note that this is a poor presentation of the data present: all data are available at high resolution in zoomable datasets via http://www.nanotomy.org/OA/Bugter2020BOE/index.html. The dataset used for this figure is 2016–194. Bars: 100, 10, 1, and 0.2 um.

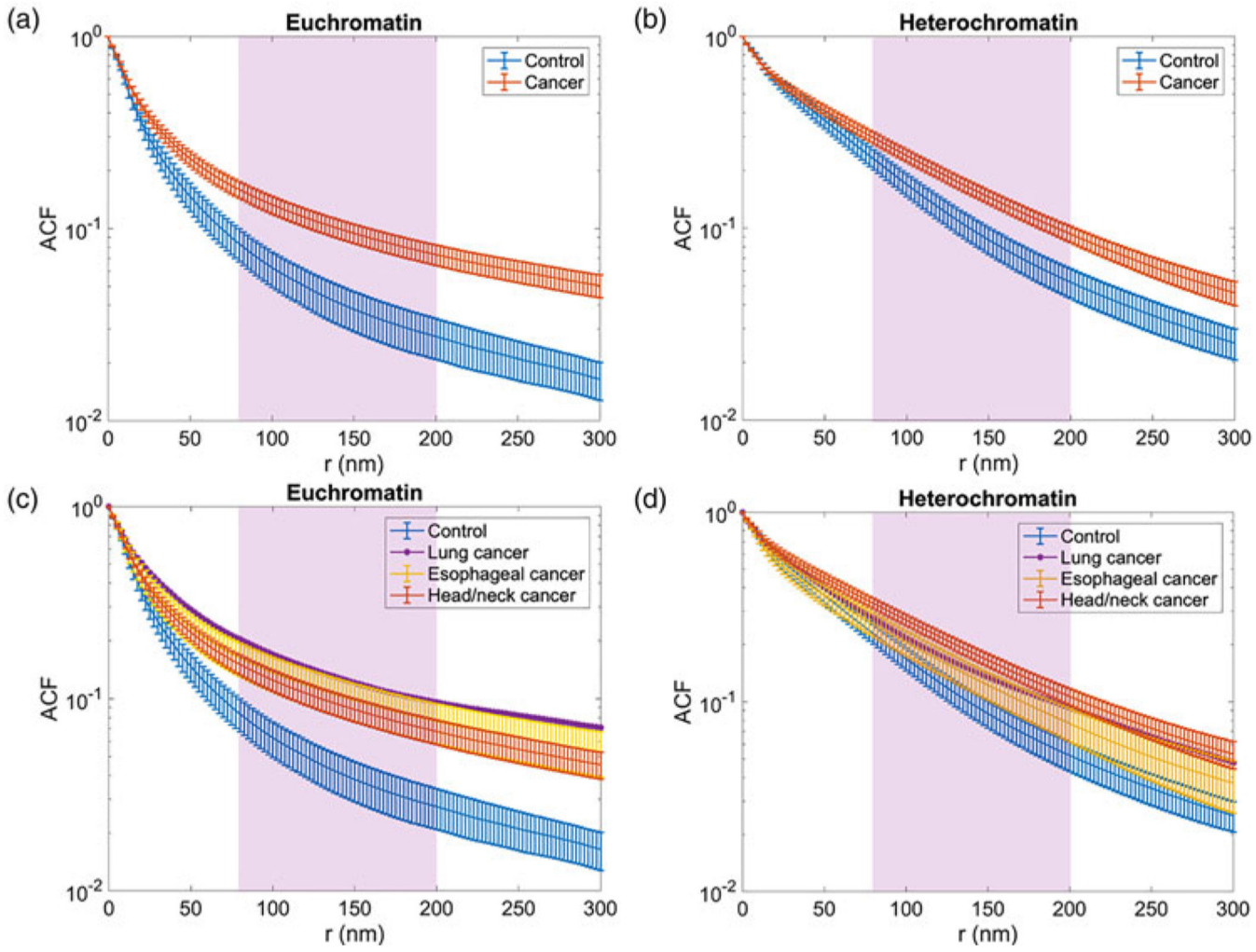

Fig. 4.

Differentiation of control versus cancer cells based on chromatin packing alterations using ACF analysis. ACFs for the euchromatin (**a**) and the heterochromatin (**b**) for the all control (blue) and the all cancer (red) groups show distinct chromatin packing within 80 nm < *r* < 200 nm (purple shaded region) with $p < 0.001$. The breakdown of ACFs by diagnosis is shown in (**c**) for the euchromatin and (**d**) for the heterochromatin. The ACFs were calculated from averaging the mean ACFs per patient, and the error bars represent the standard error.

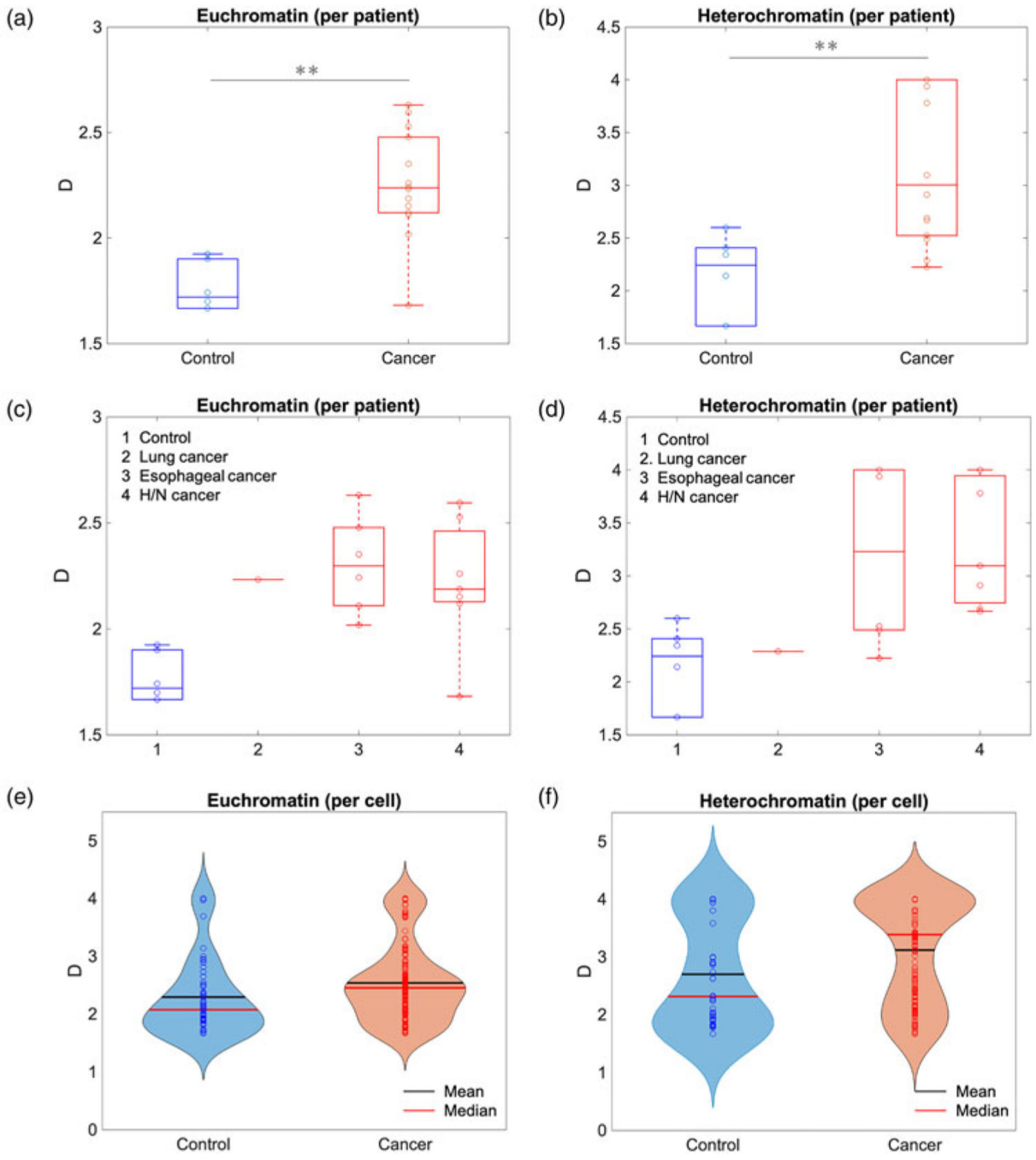

Fig. 5.

Quantifying chromatin packing alterations using packing scaling *D*. Mean ACF per patient was employed in calculating *D* within the length scales under the diffraction limit. For all control and cancer patients, the difference in *D* is statistically significant for the euchromatin (**a**) with *p*-value = 0.002 and the heterochromatin (**b**) with *p*-value = 0.005. For each diagnostics category, the head/neck cancer shows a significant difference compared to the control group, while the esophageal only exhibits moderate difference for the euchromatin (**c**) with *p*-value = 0.015 for H/N cancer and *p*-value = 0.004 for esophageal cancer and

the heterochromatin (**d**) with *p*-value = 0.003 for H/N cancer and *p*-value = 0.037 for esophageal cancer. Chromatin packing scaling distribution calculated from each cell for (**e**) the euchromatin and (**f**) the heterochromatin. The black line denotes the mean value and the red line represents the median value.

Table 1.

Baseline Characteristics of Included Patients.

				Tumor sublocations of cancer group					
	Control (*n* = 6)	**Cancer (*n* = 14)**	***p*-value**	**HN (*n* = 7)**	***p*-value (*n* = 6)**	**Esophagus**	***p*-value**	**Lung (*n* = 1)**	***p*-value**
Male sex, *n* (%)	4 (66.7)	11 (78.6)	0.613	7 (100.0)	0.192	4 (66.7)	1.000	0 (0.0)	0.429
Age, median (IQR)	62.0 (57.2–74.5)	69.0 (66.0–72.8)	0.284	69.2 (66.3–73.2)	0.568	70.1 (66.3–75.5)	0.200	64.3	0.617
Smoking			0.067		0.099		0.059		0.459
Never	3 (50.0)	1 (7.1)		-		1 (16.7)		-	
Past	1 (16.7)	8 (57.1)		3 (42.9)		5 (83.3)		-	
Current	2 (33.3)	5 (35.7)		4 (57.1)		0 (0.0)		1 (100.0)	
Smoking PY, median (IQR)	6.3 (0.0–28.5)	30.0 (11.9–36.3)	0.081	35.0 (30.0–40.0)	0.031 *	21.2 (7.5–30.6)	0.413	24.0	0.604
Tumor stage									
I		6 (42.9)		5 (71.4)		1 (16.7)		0	
II		2 (14.3)		0		1 (16.7)		1 (100.0)	
III		5 (35.7)		2 (28.6)		3 (50.0)		0	
IV		1 (7.1)		0		1 (16.7)		0	
Tumor type, n (%)									
PCC		12 (85.7)		7 (100.0)		3 (50.0)		1 (100.0)	
Adeno		2 (14.3)		0		3 (50.0)		0	

HN, head and neck; IQR, interquartile range; PY, pack years; SCC, squamous cell carcinoma. *p*-value is compared with the control group.

* *p*-value < 0.05.