Whole slide image cytometry: a novel method to detect abnormal DNA content in Barrett's esophagus

Yinhai Wang[1], Damian T McManus[2,3], Kenneth Arthur[3], Brian T Johnston[2], Andrew J Kennedy[2], Helen G Coleman[4], Liam J Murray[4] and Peter W Hamilton[3]

Barrett's esophagus (BE) is a precursor of esophageal adenocarcinoma (EAC). Both low-grade dysplasia (LGD) and high-grade dysplasia (HGD) are associated with an increased risk of progression to EAC. However, histological interpretation and grading of dysplasia (particularly LGD) is subjective and poorly reproducible. This study has combined whole slide imaging with DNA image cytometry to provide a novel method for the detection of abnormal DNA content through image analysis of tissue sections. A total of 20 cases were evaluated, including 8 negative for dysplasia (NFD), 6 LGD, and 6 HGD. Feulgen-stained esophageal sections were scanned in their entirety. Barrett's mucosa was interactively chosen for automatic nuclei segmentation where irrelevant cell types were ignored. The combined DNA content histogram for all nuclei within selected image regions was then obtained. In addition, three histogram measurements were computed, including xER-5C, 2cDI, and DNA-MG. Visual evaluation suggested the shape of DNA content histograms from NFD, LGD, and HGD cases exhibiting identifiable differences. The histogram measurements, xER-5C, 2cDI, and DNA-MG, were shown to be effective in differentiating metaplastic from dysplastic cases with statistical significance. Moreover, they also successfully separated NFD, LGD, and HGD patients with statistical significance. Whole slide image cytometry is a novel and effective method for the detection of abnormal DNA content in BE. Compared with histological review, it is more objective. Compared with flow cytometry and cytology-preparation image cytometry, it is low cost, simple to use, only requires a single 1 μm section, and facilitates selection of tissue and topographical correlation. Whole slide image cytometry can detect differences in DNA content between NFD, LGD, and HGD patients in this cross-sectional study. Abnormal DNA content detection by whole slide image cytometry is a promising biomarker of progression that could affect future diagnostics in BE.

The incidence of esophageal adenocarcinoma (EAC) has increased dramatically in the west over the past 30–40 years.[1–3] In the United States from 1975 to 2001, the incidence of EAC rose from 4 to 23 cases per million.[4] In the United Kingdom, a similar increase in EAC[5] is observed, although the absolute incidence is much higher than in the United States.[6] Globally, it is the eighth most common cancer with nearly 456 000 new cases diagnosed in 2012 (3% of the total). Esophageal cancer incidence rates are highest in Eastern Asia and lowest in Western Africa, but this partly reflects varying data quality worldwide.[7] The majority of EAC patients present with late-stage disease and survival is poor, with a 5-year relative survival rate of <15%.[8,9]

Barrett's esophagus (BE) is well accepted as a precursor to EAC.[10,11] Dysplasia in BE is associated with an increased risk

of progression to EAC, and is used in clinical practice as the 'gold standard' predictive biomarker for invasive carcinoma.[12,13] In many centers, periodic endoscopic surveillance is recommended for patients with confirmed Barrett's esophagus with specialized intestinal metaplasia (SIM). The latest radiofrequency ablation (RFA) clinical trial[14] confirmed, for the first time, that RFA treatment reduces neoplastic progression in patients with low-grade dysplasia (LGD). The detection of high-grade dysplasia (HGD) in BE patients may lead to radical therapeutic interventions such as esophagectomy or endoscopic mucosal resection, and RFA treatment.[15–17]

Some recent studies investigated the use of fluorescence *in situ* hybridization (FISH) for the detection of dysplasia,[18] and the potential molecular biomarkers for EAC

[1]Finland Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland; [2]Belfast HSC Trust, Belfast, UK; [3]Centre for Cancer Research and Cell Biology (CCRCB), Queen's University Belfast, Belfast, UK and [4]Centre for Public Health, Queen's University Belfast, Belfast, UK
Correspondence: Dr Y Wang, PhD, Finland Institute for Molecular Medicine (FIMM), University of Helsinki, P.O. Box 20, Helsinki, FI-00014, Finland or Professor PW Hamilton, PhD, Centre for Cancer Research and Cell Biology, Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7BL, UK.
E-mail: y.wang@qub.ac.uk or p.hamilton@qub.ac.uk

progression,[19,20] including epigenetic phenomena.[21,22] Although these attempts may hold promise, the diagnostic interpretation and histological grading of key clinical decision points in BE, including negative for dysplasia (NFD), LGD, and HGD, remains problematic. In current clinical settings, this is still based on the histopathological assessment of routinely hematoxylin and eosin (H&E)-stained tissue sections. It is well acknowledged that histological assessment is subjective and incurs inter-/intra-observer variability.[8,12]

Aneuploid DNA content as measured by flow cytometry on nuclear suspensions has been strongly advocated as a marker of progression in BE and is the only other biomarker used by some in clinical practice.[13,23] However, the wide use of flow cytometry is limited because of its high cost and complex sample preparation procedure.[24,25] A significant amount of tissue needs to be sacrificed. Most importantly, flow cytometry is regarded as a 'black-box' method, where cells are disaggregated and analyzed irrespectively. Cytology-preparation image cytometry is an alternative approach for measuring DNA content. It uses two or more thick sections of FFPE (formalin fixed and paraffin embedded, e.g., 50 μm) and disaggregates cells into a monolayer preparation, followed by Feulgen staining.[26,27] Using digital image processing techniques, the integrated optical density for each nucleus is obtained followed by the construction and interpretation of DNA content histograms, similar to that of flow cytometry. This technology consumes valuable tissue samples, requires complex sample preparation, and is labor intensive, and hence not widely available for the frontline health care. Neither flow cytometry nor cytology-preparation image cytometry permits the correlation of DNA content data with tissue and cellular context. Pathologists and technicians do not have any control over which tissue region or cell groups to be analyzed.

Despite these limitations, the detection of abnormal DNA content by flow cytometry and cytology image cytometry[20,28,29] has been correlated with progression and recurrence of dysplasia in Barrett's patients. Abnormal DNA content remains one of only a few biomarkers to have been tested in longitudinal studies.

A practical and cost-effective method is needed for the detection of abnormal DNA content in biopsy samples at the time of routine pathological review. With the development of digital pathology and tissue slide scanning technology, it is possible to scan the entire histological tissue section at diagnostic resolution (e.g., 0.25 μm per pixel) creating large digital slide images. A typical tissue specimen can exceed 120 000 × 80 000 pixels in size at × 40 magnification (28 gigabytes of uncompressed data or ~ 1 GB JPEG compressed[30]). Whole slide imaging, together with appropriate viewing technology, is beginning to be used for a range of pathology applications, such as pathology education,[31,32] image analysis for biomarker discovery,[33,34] and primary diagnosis.[35–37]

In tissue-based BE research, some researchers have investigated alternative computerized morphometry and texture measurements as indicators for disease progression.[18,38,39]

Nonetheless, the majority of other studies are still focusing on the well-established DNA density measurements using tissue sections (e.g., 3–7 μm), especially the use of the commercially available ACIS system (Automated Cellular Imaging System, Dako, Denmark) from a field of view (FoV) of tissue sections.[40–44] Under a microscope at × 40 magnification, a FoV is digitally captured followed by image analysis to locate nuclei and to measure the integrated optical density (IOD) for each nucleus. DNA content histogram for FoV can then be constructed. Nonetheless, a FoV of a section does not always represent a whole tissue slide, and thereafter regional DNA content histograms do not always reflect the true DNA ploidy distribution of the whole tissue sample. With increased numbers of high-throughput whole slide image scanners (from, e.g., Leica Microsystems, Hamamatsu Photonics KK, and Philips) being used by researchers, pathologists and clinicians, it is possible to scan Feulgen-stained glass slides in their entirety. This study presents such a new cytometry method, namely whole slide image (WSI) cytometry. Using digitally scanned Feulgen slides, this study presents a histological image processing-based method for the detection of DNA content in endoscopic biopsy samples of metaplastic (NFD) and dysplastic (LGD and HGD) Barrett's mucosa.

MATERIALS AND METHODS
Ethics Statement
The study was undertaken under the auspices of Northern Ireland BioBank (NIBioBank Application NIB12-0039) that uses anonymized surplus diagnostic material from cancer patients and patients with premalignant conditions.

Materials
Approximately 0.4% of patients with BE develop adenocarcinoma annually.[45] Therefore, it can be difficult to identify LGD and HGD samples from the same time period/population as comparative BE samples. Samples used in this study had previously been histopathologically reviewed for selection in one of two biomarker studies that our working group have previously published.[20,46] To reduce difficulties for image segmentation for overlapping nuclei at later stages, tissue sections, nominally 1 μm in thickness (within limits of microtomes), were cut from each biopsy via a standard rotary microtome. Tissue sections were then H&E stained, followed by the pathological re-evaluation (by an expert pathologist DTMcM) with verification of the original diagnosis by the reporting pathologists as previously described.[20,46] The pathologist DTMcM confirmed the presence of BE-NFD and/or LGD or HGD in the levels used for the Feulgen stain. All cases displayed SIM, 8 were uncomplicated NFD cases, and 12 cases had areas of dysplasia, including 6 cases of LGD and 6 cases of HGD.

Another 1 μm section from each FFPE block was cut for Feulgen staining. Both H&E- and Feulgen-stained slides were subsequently scanned using an Aperio ScanScope CS whole slide scanner (Leica Microsystems, San Diego, CA, USA)

using the objective of 20×/0.75 Plan Apo with a doubler. The H&E-stained slides were digitally scanned at 40× magnification that gave the image resolution of 0.25 µm/pixel. The Feulgen-stained slides were scanned at the highest 83× magnification under oil immersion, generating digital slides with the resolution of 0.17 µm/pixel. Both H&E- and Feulgen-stained digital slides were compressed using standard JPEG compression at the compression quality of 70.

Scanned digital slides were archived at PathXL and slides viewed using their online digital slide viewer (PathXL, Belfast, UK). An experienced histopathologist (DTMcM) reviewed the H&E-stained slides online, and annotated tissue regions with NFD, LGD, and HGD using a freehand drawing tool within the PathXL software.

Methods

The proposed WSI cytometry method follows a two-step approach. First, a pathologist interactively selects a number of tissue regions of interest from a whole slide image using software tools. Second, individual nuclei from the selected regions are automatically identified and isolated via a nuclei segmentation method.

Three forms of output can then be generated. First, a DNA content histogram in the form of a 2D plot is produced. A graphical webpage, namely *CellMap*, is also constructed to illustrate the relationship between a DNA content histogram and individual cell profiles. Finally, three statistical measurements are calculated to numerically describe the degree of DNA content abnormality.

The following subsections present the technical details of these two technical steps and the three forms of output.

Tissue region selection

We chose Aperio's ImageScope software for initial tissue region selection. A pathologist is able to view the Feulgen-stained digital slide at any magnifications up to 83×, and extract any number of rectangular regions with arbitrary sizes using the Extract Region Tool from Aperio ImageScope software, as illustrated in Figure 1a. The extracted regions are subsequently saved as JPEG images.

It is unavoidable that the selected rectangle regions may contain unwanted cell types, such as lymphocytes and stromal cells, rather than stratified epithelial cells. Therefore, we developed an interactive software tool using Matlab (Math-Works, Natick, MA, USA) so that the pathologist could further select areas of interest from the rectangle regions, where unwanted cell types can be excluded. This software tool allows pathologists to make this interactive selection using four types of shape drawings, rectangle, ellipse, polygon, and freehand drawing. A snapshot of the tool is shown in Figures 1b and c.

Nuclei segmentation

This nuclei segmentation process localizes each nucleus and the exact boundaries of all identified nuclei in the digital
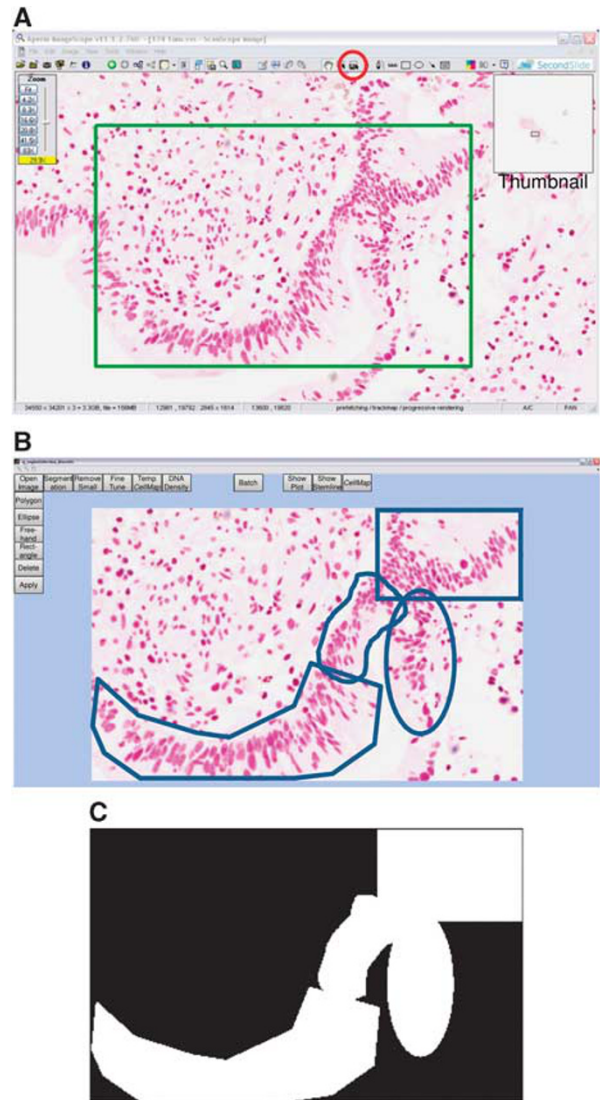

Figure 1 An illustration of region selection. (**a**) Screenshot showing the selection of a rectangular region (green) using the Extract Region tool (the red circle) from Aperio ImageScope software. (**b**) Screenshot of the annotation tool for the interactive selection of regions of interest. The list of buttons on the left enables users to draw rectangle, ellipse, polygon, or freehand drawing. The blue lines superimposed on top of the tissue region are illustrations of the four types of annotation. (**c**) After user clicks the 'Apply' button in (**b**), a binary image is automatically generated to demonstrate the user selected region (white) for further analysis.

image. Many nuclei segmentation methods from tissue sections are available in the literature, but they are largely focused on the processing of H&E-stained tissue for, e.g., prostate,[47] neuroblastoma,[48] and laryngeal tissue.[49] Given the morphological characteristics of Feulgen-stained esophageal tissue, we developed a marker-controlled watershed-based nuclei segmentation algorithm using the Matlab (version 8) software.[50]

To describe this algorithm briefly, a RGB color image is firstly converted into the CIE $L^*a^*b^*$ color space.[51] The a^*

color channel I_a is processed at 83× magnification for contrast enhancement using a contrast-limited adaptive histogram equalization method[52] and a foreground enhancement using morphological opening-by-reconstruction operations. The resultant image I_i is then filtered with a Sobel edge detector to get the gradient magnitude image I_{mag}. A marker controlled watershed algorithm uses both internal and external markers to limit the amount of oversegmentations. External markers should be located within the background (non-nuclei area) that is easily achieved using an Otsu's global thresholding on I_i followed by a distance transform. Internal markers should be subcomponents inside each nucleus that is achieved by applying an extended-maxima transform on I_i and followed by morphological tidying-ups. The gradient magnitude image I_{mag} is then modified using a minima imposition technique to enforce regional minima to occur only at marked locations using both internal and external markers. A watershed transform is applied on the modified gradient magnitude image to get the segmented nuclei image I_{nuc}. Finally, the shape and contour of the segmented I_{nuc} is fine-tuned using convex hulls[53] and a pair of forward and inverse Fourier descriptors[54,55] to get smooth boundaries.

To eliminate artifacts such as overlapping/touching nuclei and debris, only nuclei with the diameter of 10 to 56 pixels (at 83× magnification) are retained. An interactive manual removal procedure is also included so that artifacts missed by the image processing can be further removed. In this study, this manual removal procedure was performed by the pathologist DTMcM or YW and validated by DTMcM.

Examples of the nuclei segmentation results are shown in Figures 2c and 3b. A number of regions can be selected from each BE whole tissue slide for the segmentation of nuclei. They are further used for the construction of DNA content histograms.

Controls

Control cells are important to scale DNA measurements, as well as to define the location of DNA diploid (2N) and DNA tetraploid (4N) in subsequently generated DNA content histogram. As widely used in image cytometry studies,[56,57] lymphocytes were chosen for internal controls in this study.

Following the same tissue region selection and nuclei segmentation steps, control cells are identified. Regions of high density of lymphocytes are firstly identified using Aperio's ImageScope software and our in-house-developed region selection tool. Each single lymphocyte is then segmented. In this study, the lymphocyte region selection and segmentation were performed by DTMcM, or by YW under the supervision of DTMcM as above. A number of lymphocyte-rich regions can be selected for automated segmentation.

DNA content histogram

After the segmentation of nuclei, the boundaries of nuclei are identified. Using Feulgen-stained whole slide tissue scan, the immediate results obtained are the image color intensity in

the RGB color space for each pixel within an identified nucleus. The Aperio digital slide scanner has an objective lens in it. Therefore, similar to the use of traditional microscope, in Aperio scanned digital slide, the optical density (OD) of a pixel can be derived from its RGB color intensity.

Given a pixel I with the RGB component of (I_r, I_g, I_b), its gray scale value G can be obtained using the following linear combination:

$$G = \begin{bmatrix} 0.2989 & 0.5870 & 0.1140 \end{bmatrix} \cdot \begin{bmatrix} I_r \\ I_g \\ I_b \end{bmatrix} \quad (1)$$

For the digital slide scanned using Aperio's scanner, the color optical density for a pixel can be easily obtained using its corresponding RGB values[58,59]:

$$OD_m = log_{10}\frac{I_0}{I_m} \quad (2)$$

where $I_0 = 240$, and $m \in \{r,g,b\}$. Therefore the optical density OD of given pixel can be obtained:

$$OD = log_{10}\frac{I_0}{G} \quad (3)$$

By considering all the pixels within the boundary of a segmented nucleus, we have the IOD for the nucleus in the $1\,\mu$m tissue section:

$$IOD = \sum OD \quad (4)$$

Assuming nuclei are nearly round objects, to evaluate DNA distribution from tissue sections with different thickness, the correction factor α can be used to obtain the normalized integrated optical density ($nIOD$)[60].

$$\alpha = \begin{cases} \frac{1}{1-\frac{3r}{8t}} & r < t \\ \frac{1}{\frac{3t}{4r}-\frac{r}{8t}\left(\frac{t}{r}\right)^3} & r > t \end{cases} \quad (5)$$

where t is the thickness of the tissue section, and r is the radius of a nucleus. The authors also acknowledge the existence of other correction methods in the literature.[61]

$$nIOD = \frac{\alpha \cdot IOD}{\langle IOD_C \rangle} \quad (6)$$

where IOD_C is the integrated optical density for the control lymphocytes, and $\langle \cdot \rangle$ is the average operator. DNA content histogram can then be constructed using all the $nIOD$ values.

A regional DNA content histogram can be plotted using all the nIODs in an image region (Figures 2b–d). It is also possible to combine a number of regional DNA content histograms from a slide to form the multiregional DNA content histogram, illustrated in Figure 2e.

The x axis of the DNA content histogram has the unit of nIOD. We then superimpose the DNA density distribution of the internal control lymphocytes on top of the regional/multiregional DNA content histogram. It is expected that the nIODs of the lymphocytes follow a single Gaussian
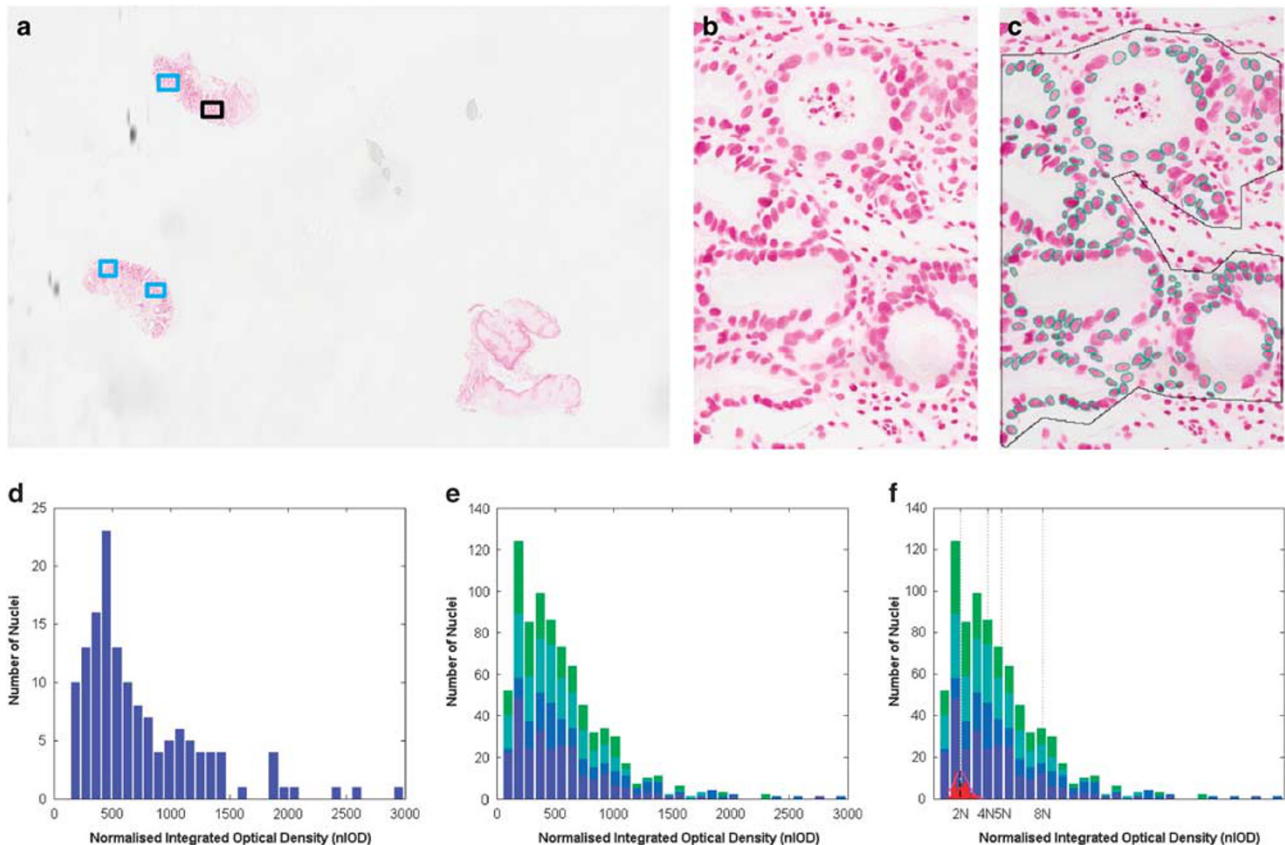
Figure 2 An example showing the construction of DNA content histogram. (**a**) A thumbnail view of a Feulgen-stained slide containing three pieces of tissue. The four rectangles indicate they were interactively selected for further analysis. (**b**) An image region as indicated as the black rectangle in (**a**). (**c**) The nuclei segmentation result superimposed on (**b**). (**d**) The regional DNA content histogram of (**b**). (**e**) The combined DNA content histogram using all the four image regions indicated as rectangles in (**a**). (**f**) The combined DNA content histogram with the x axes scaled using internal lymphocyte controls.

distribution, and the peak of the lymphocyte Gaussian peak reflects DNA diploid *aka* the location of 2N. Therefore, using simple scaling, the reference unit scale (e.g., 2N, 4N, 5N, and 8N) can be defined as the example shown in Figure 2f.

CellMap

DNA content histogram uses bars to represent the number of cells in each bin, whereas the proposed *CellMap* merges the actual segmented nuclei with the DNA content histogram in the form of a webpage. An example is shown in Figure 3. If we rotate the DNA content histogram (Figure 3c) clockwise to be a horizontal bar graph, the histogram shape of Figure 3c matches exactly with Figure 3d.

CellMap provides such a cross-referencing capability between DNA content histogram and individual cell morphology profile. The histogram distribution is embedded in the *CellMap*, and the exact location of each nucleus in the histogram is clearly presented for visual inspection.

DNA cell histogram can also be used to remove oversegmentation at the nuclei segmentation stage and for cross-referencing at later data analysis stage. When nuclei were

segmented, oversegmentation could happen and potentially lead to contaminated data and interpretation. Overlapping nuclei could be wrongly recognized as a single nucleus and included in histogram construction. Similarly, debris may also be wrongly recognized as a nucleus. To double check the correctness of nuclei segmentation, the real-time *CellMap* could help to further remove these oversegmentations. Debris is typically small and contains small amount of DNA content, and are therefore likely to appear at the top rows in the *CellMap*. Overlapping nuclei typically contains, e.g., ≥ double amount of DNA content, and hence are more likely to appear at the bottom rows in the *CellMap*. Simple visual examination could easily identify these oversegmentations. They can then be interactively removed from the segmentation result by simple button clicks.

Statistical measurements

Three DNA content histogram statistical measurements from cytology-preparation image cytometry were borrowed and implemented in this study:[56] the popularly used 5c exceeding ratio (xER-5C), 2c deviation index (2cDI), and DNA grade of
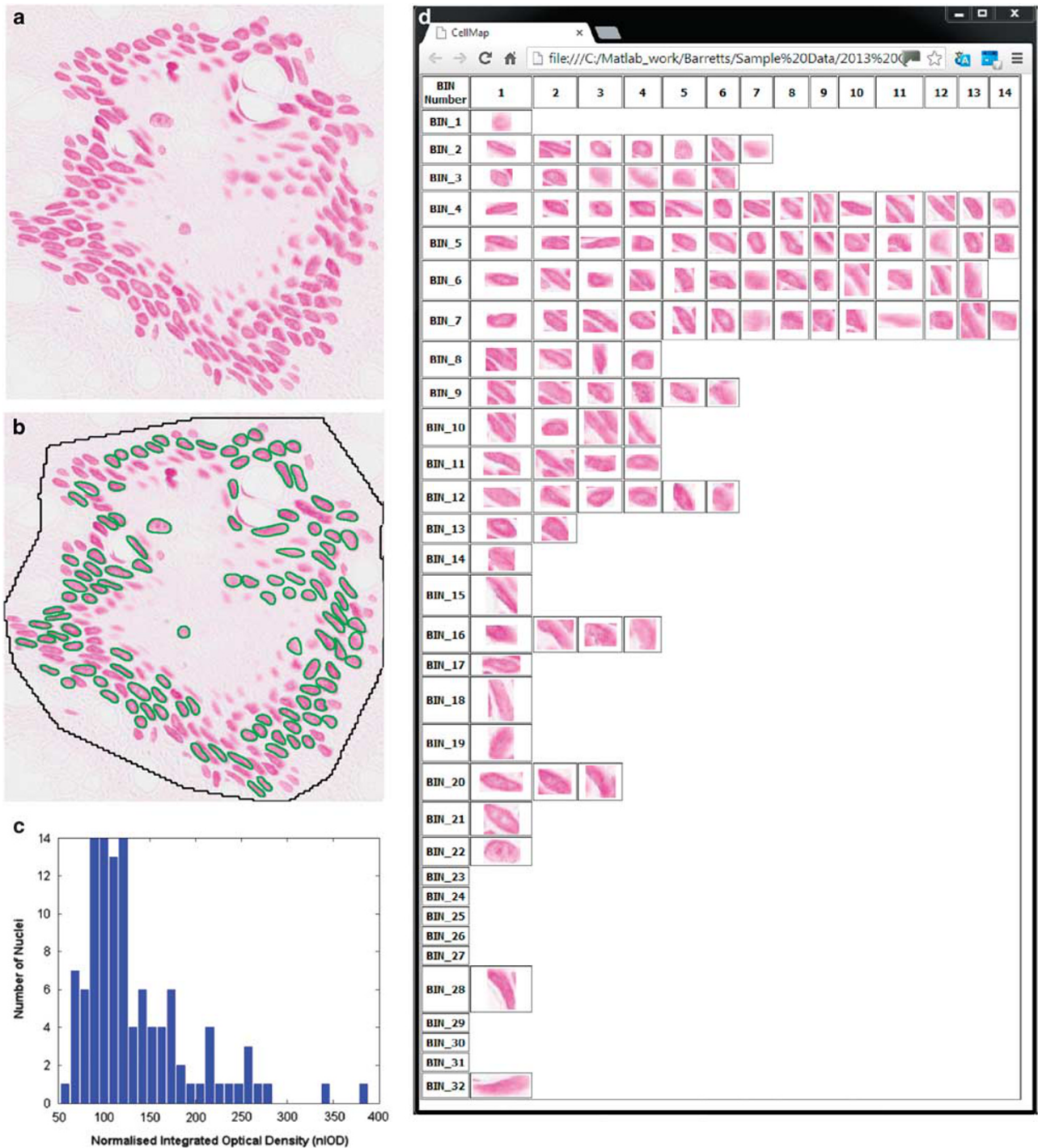
Figure 3 An example of a *CellMap*. (**a**) A small image region. (**b**) The nuclei segmentation results superimposed on (**a**). (**c**) The regional DNA content histogram of (**a**). (**d**) The *CellMap* of (**a**).

malignancy (DNA-MG). These three measurements were calculated using the nIOD value as the variable.

RESULTS

To evaluate the robustness of the proposed WSI cytometry method, we first visually evaluated the DNA content histograms from all the 20 cases that were histologically categorized to be NFD, LGD, and HGD. This evaluation was also assisted by *CellMap*. Second, we compared their statistical measurements using two sample *t*-tests.

DNA Content Histograms

After manual selection of a number of regions of interest indicated by the pathologist, the combined DNA content
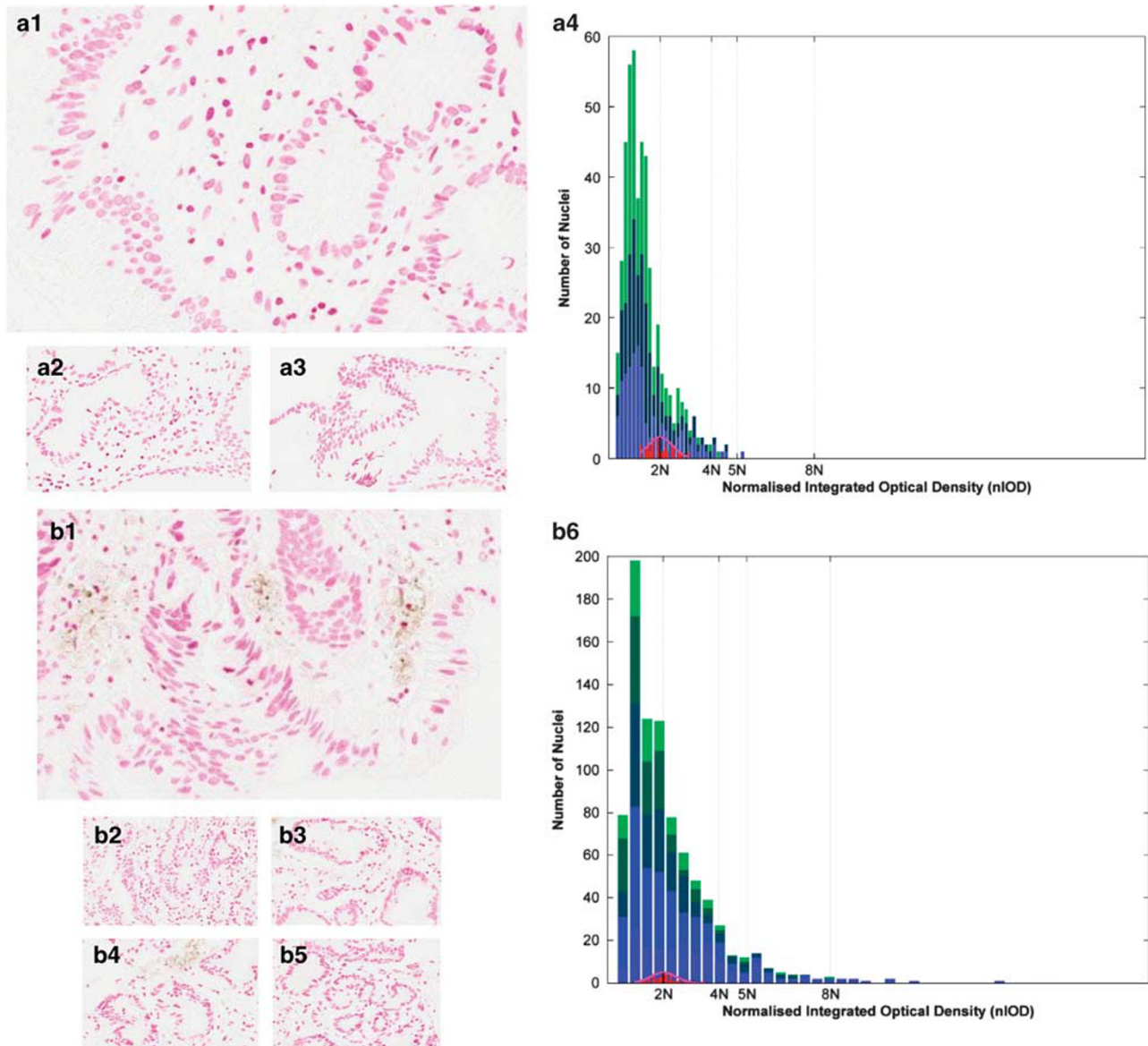
Figure 4 Three examples showing the differences of DNA content histogram among NFD, LGD, and HGD cases. (**a**) A NFD case with three regions (**a**1–**a**3) analyzed. (**a**4) The combined DNA content histogram of all three regions is shown. (**b**) A LGD case with five regions (**b**1–**b**5) analyzed. (**b**6) The combined DNA content histogram of all five regions is shown. (**c**) A HGD case with three regions (**c**1–**c**3) analyzed. (**c**4) The combined DNA content histogram of all three regions is shown.

histogram for each of the 20 cases was obtained. Examples from each of the NFD, LGD, and HGD cases are shown in Figure 4.

Visual assessment 1

For the NFD case illustrated in Figure 4a4, the peak of the histogram was located at the fifth bar (out of 32 bars), to the left of the 2N position defined by controls. The height of the bar at 4N position was lower than 2N. As a nondysplastic case, there were not any DNA octaploid cells; hence, the 8N location was not shown. With the help of the *CellMap*, we were able to easily visualize the actual cell morphologies for

each bar in the histogram. To use the NFD case illustrated in Figure 4a as an example, a number of cells from the 5th bin (peak of the histogram), 12th bin (the 2N position defined by controls), and 24th and 25th bins (4N position) are shown in Figure 5a. A quick visual assessment suggested cells selected from the fifth bin were smaller in size and paler in staining comparing with the cells from the 2N position. This trend was much more obvious when comparing the cells from the 2N position with the 4N position. The increase in cell size and stain intensity were more obvious as shown in the LGD and HGD case (Figures 5b and c) when comparing the cells taken out from 2N, 4N, and 8N positions.
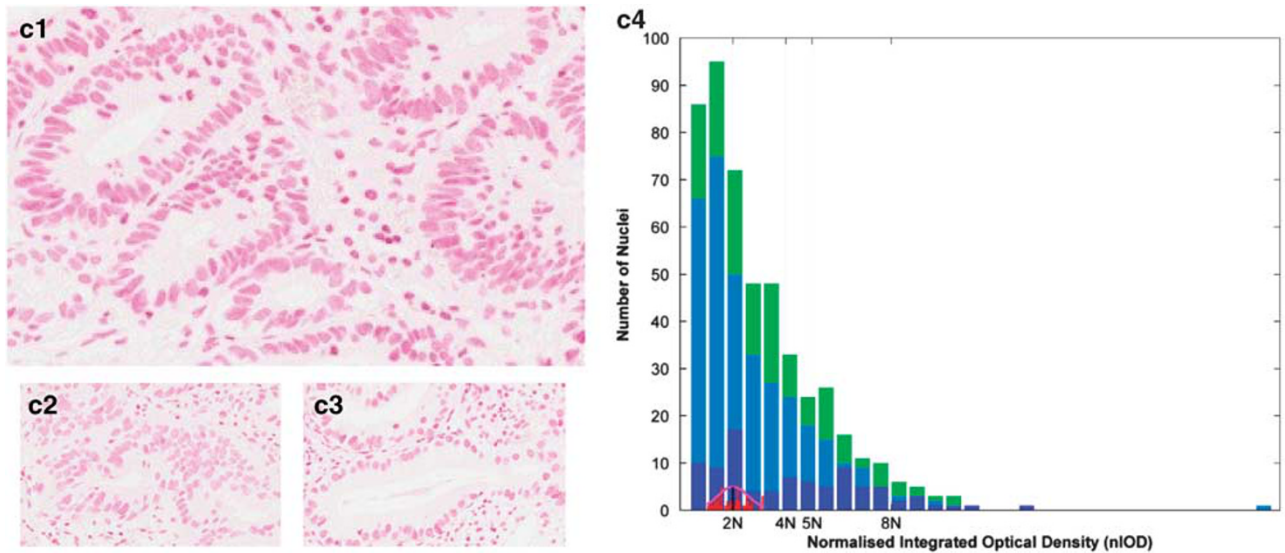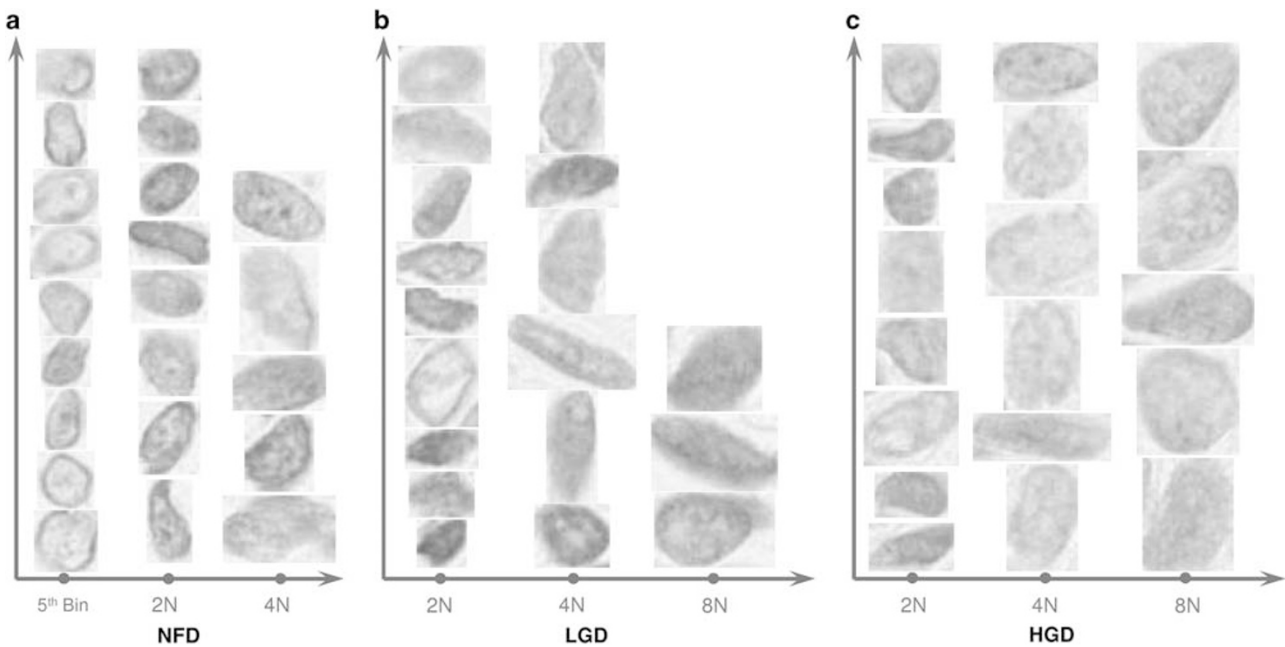
Figure 4 (Continued)

Figure 5 Example of segmented nuclei from three cases. (**a**) A NFD case example of nuclei from the fifth bin (the peak), the 2N and 4N position of the DNA content histogram. (**b**) A LGD case shows nuclei segmented at the 2N, 4N, and 8N position of the DNA content histogram. (**c**) A HGD case shows nuclei segmented at the 2N, 4N, and 8N position of the DNA content histogram.

Visual assessment 2
Visual evaluation also suggested the differences in the shapes of histograms from the three groups. If we use the 5N position, defined using internal controls, in each histogram as reference point, it showed that there were higher proportion of cells having bigger than 5N nIOD values in the LGD case (Figure 4b6) compared with the NFD case (Figure 4a4), and in the HGD case (Figure 4c4) compared with the LGD case.

These two visual evaluations were consistent with the pathologist's histological assessment.

The combined use of DNA content histogram and *CellMap* could also easily highlight DNA aneuploid cells for further interpretation, especially the cells with high DNA content, e.g., >8N. These cells were located to the right of a DNA content histogram that corresponds to the bottom rows in the corresponding *CellMap*. Closer visual examination might

Table 1 DNA content statistical measurements for 20 Barrett's esophagus cases

	Case no.	xER-5C	2cDI	DNA-MG
Dysplasia				
HGD	1	19.90	2.59	2.25
	2	18.20	2.08	1.98
	3	18.20	1.71	1.75
	4	11.07	1.25	1.42
	5	9.15	0.87	1.10
	6	7.35	0.74	0.97
LGD	7	8.02	0.84	1.07
	8	8.57	0.82	1.05
	9	5.98	0.82	1.05
	10	8.04	0.71	0.95
	11	5.95	0.71	0.94
	12	6.12	0.64	0.87
Metaplasia				
NFD	13	7.05	0.77	1.00
	14	6.33	0.69	0.92
	15	3.41	0.59	0.81
	16	4.57	0.51	0.73
	17	2.87	0.37	0.56
	18	2.52	0.37	0.56
	19	1.20	0.29	0.45
	20	0.21	0.27	0.42

suggest these cells to be overlapping cells that were missed during cell segmentation, cells that were hard to tell whether they are overlapping or simply staining artifacts, or indeed they are genuine aneuploid cells.

Finally, the DNA content histograms were generated using the same nIOD scale, with the maximum value of 2500 IOD units. Visual evaluation also suggested slight variations of the locations for reference points defined by internal lymphocytes controls (e.g., 2N, 4N and 8N locations) from case to case. This was likely to be caused by the variations in the amount of Feulgen stain, tissue thickness, or factors from other lab procedures. This observation indicated the importance of using internal rather than external controls.

Statistical Histogram Measurements

All three histogram measurements (xER-5C, 2cDI, and DNA-MG) were calculated for the 20 cases. The results are listed in Table 1. To test the robustness of these three measurements, we first evaluated whether these three measurements were able to distinguish metaplasia (NFD, 8 cases) from dysplasia

(LGD and HGD, 12 cases altogether) using two-sample *t*-tests. We then tested the abilities of these three measurements in differentiating NFD from LGD, and from HGD cases. Similarly, two-sample *t*-tests were performed and individual *P*-values were calculated.

Results showed that all three measurements successfully differentiated metaplastic from dysplastic cases with high degree of statistical significance ($P < 0.005$, Figures 6a–c), $P_{\text{xER-5C}} = 0.0008$, $P_{\text{2cDI}} = 0.0045$, and $P_{\text{DNA-MG}} = 0.0011$. *P*-values suggested that xER-5C is the most robust measurement followed by DNA-MG.

For the testing among NFD, LGD, and HGD cases, statistical significance was also achieved for each pair of comparisons (Figures 6d–f). High degrees of statistical significance ($P < 0.005$) were obtained when comparing NFD from LGD samples with $P_{\text{xER-5C}} = 0.0036$, $P_{\text{2cDI}} = 0.0040$, and $P_{\text{DNA-MG}} = 0.0049$. The evaluation of NFD and HGD cases also showed high degree of statistical significance for the measurement of xER-5C ($P_{\text{xER-5C}} = 0.0038$), and statistical significance for other two measurements ($P_{\text{2cDI}} = 0.0148$ and $P_{\text{DNA-MG}} = 0.0054$). Furthermore, the differences between LGD and HGD cases were also statistically significant for all the three measurements with $P_{\text{xER-5C}} = 0.0258$, $P_{\text{2cDI}} = 0.0446$, and $P_{\text{DNA-MG}} = 0.0338$.

DISCUSSION

Dysplasia is a predictive biomarker in BE for the increased risk for EAC progression. The treatment and clinical management for patients with NFD, LGD, and HGD are different. In some centers, patients with NFD undergo periodic endoscopic surveillance. The management of LGD patients is controversial; some guidelines recommend patients have frequent endoscopic surveillance, whereas a recent study suggests RFA treatment reduces the risk of neoplastic progression to HGD/EAC.[14] HGD patients can be managed by endoscopic mucosal resection and/or RFA. The histological classification for some key diagnostic stages in BE has been challenging, resulting in a fair interobserver agreement (LGD, $\kappa = 0.32$).[12] Flow cytometry and cytology-preparation image cytometry are not suitable for widespread clinical use. Studies also suggested the simple measurements of cross-sectional nuclear area for cytometric analysis. Nonetheless, the lack of correlation between nuclear size and DNA ploidy were shown in colorectal[62] and breast cancer.[63] In this study, we have developed a method for WSI cytometry. Through the detection and quantification of abnormal DNA content, the three proposed statistical measurements from this study were able to differentiate NFD cases from LGD and HGD. Moreover, it successfully separated cases with LGD from HGD and NFD cases.

WSI cytometry uses modern histological image processing technology to measure DNA content using whole tissue digital slides. Compared with manual histological review, the proposed method has the potential to be more objective. Compared with flow cytometry and cytology-preparation image cytometry, WSI cytometry is simple to use and only
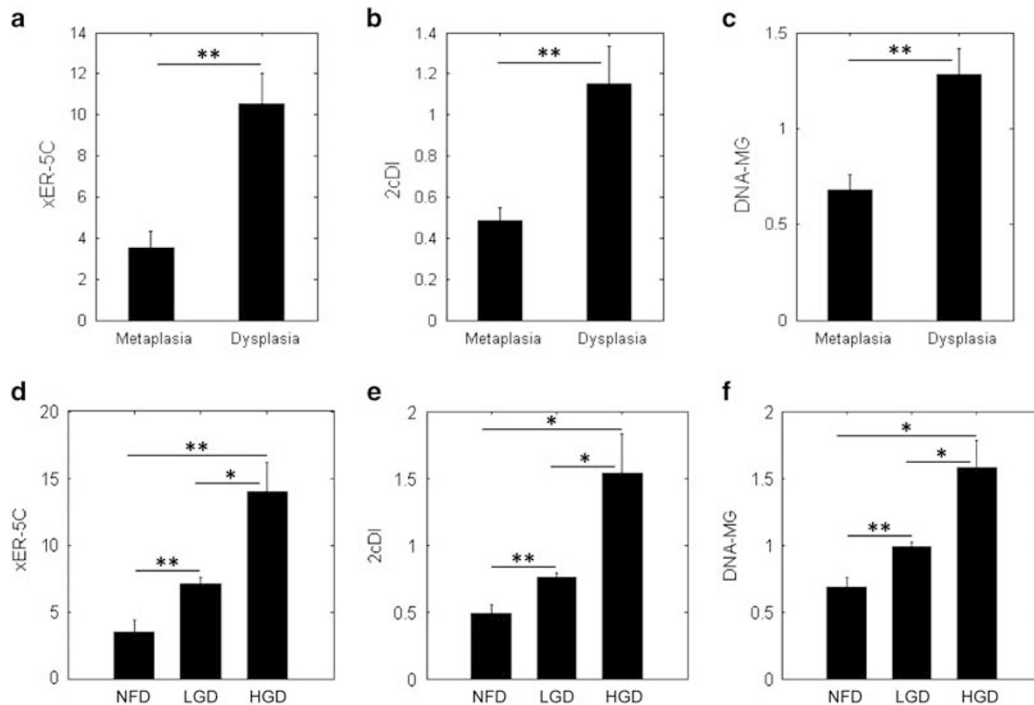
Figure 6 The *t*-test results showing the statistical significant differences between metaplastic and dysplastic cases (**a–c**) and among NFD, LGD, and HGD cases (**d–f**). (**a**) The measurement of 5c exceeding ratio between metaplastic and dysplastic cases. (**b**) The measurement of 2c deviation index between metaplastic and dysplastic cases. (**c**) The measurement of DNA grade of malignancy between metaplastic and dysplastic cases. (**d**) The measurement of 5c exceeding ratio among NFD, LGD, and HGD cases. (**e**) The measurement of 2c deviation index among NFD, LGD, and HGD cases. (**f**) The measurement of DNA grade of malignancy among NFD, LGD, and HGD cases. *$P < 0.05$ and **$P < 0.005$, statistical significance between the indicated groups.

requires one thin (e.g., 1 μm) tissue section. Compared with previous studies of analyzing a FoV from histological tissues, WSI cytometry is able to select multiple tissue regions and to combine (stack) regional histograms together to reflect the summary of heterogeneous regional characteristics.

Rather than analyzing the entire tissue and every cell, a previous study in tissue-based ploidy analysis[27] has discussed the advantages of selective sampling that permits cross-referencing between DNA content histogram with the original tissue context. WSI cytometry extended this cross-referencing ability with the use of *CellMap* where DNA content histogram is visually associated with individual cell profile. Cellular chromatin characteristics are now easily visualized within the context of DNA content histogram. In addition, a pathologist is able to combine his/her rich interpretation skills to interactively select tissue regions and cell types of interest improving the correlation of morphological changes with DNA content histograms.

WSI cytometry generates a wider peak width (the coefficient of variation (CV)) in the DNA content histogram compared with flow cytometry or cytology-preparation image cytometry. This has been reported previously and attributed to the intrinsic characteristics of performing DNA ploidy analysis using thin tissue sections.[43,64] Thick tissue sections (e.g., >10 μm) create difficulty in nuclei segmentation, especially in situations of touching and overlapping nuclei. Therefore, biased sampling error could happen where smaller and isolated nuclei are preferred.[27,65] To avoid nuclear oversegmentation, this study used ∼1 μm thin tissue sections. When a thin tissue section is applied (e.g., 1–5 μm), the IOD of a segment of a nucleus rather than the entire nucleus is measured. Hence, the sensitivity for the detection of aneuploidy is compromised.[66,67] Sectioned nucleus segments with various sizes (up to the true diameter of the nucleus) are obtained that in turn lead to a range of IOD values depending on the position where the nuclei are truncated. A collection of these IOD measurements consequently result in a wider peak in the histogram and a higher CV. Despite this limitation, our study has demonstrated that it is an effective technique with differences evident between NFD, LGD, and HGD cases, both on visual evaluation and by statistical analysis in these samples. The authors also acknowledge the small number of samples investigated in this preliminary study. Therefore, further longitudinal studies employing more samples are required to validate the visual evaluation of WSI cytometry DNA content histogram and to optimize threshold values in differentiating NFD from LGD, and LGD from HGD. It is also acknowledged that further studies with the comparison of not only pathological grading, but also actual patient outcomes data would provide stronger validation for the proposed WSI cytometry method.

DISCLOSURE/CONFLICT OF INTEREST
The authors declare no conflict of interest.

1. Blot WJ, McLaughlin JK. The changing epidemiology of esophageal cancer. Semin Oncol 1999;26:2–8.
2. Bosetti C, Levi F, Ferlay J et al. Trends in oesophageal cancer incidence and mortality in Europe. Int J Cancer 2008;122:1118–1129.
3. Botterweck AAM, Schouten LJ, Volovics A et al. Trends in incidence of adenocarcinoma of the oesophagus and gastric cardia in ten European countries. Int J Epidemiol 2000;29:645–654.
4. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. J Natl Cancer Inst 2005;97:142–146.
5. Lepage C, Rachet B, Jooste V et al. Continuing rapid increase in esophageal adenocarcinoma in England and Wales. Am J Gastroenterol 2008;103:2694–2699.
6. Jankowski JA, Provenzale D, Moayyedi P. Esophageal adenocarcinoma arising from Barrett's metaplasia has regional variations in the West. Gastroenterology 2002;122:588–590.
7. Ferlay J, Soerjomataram I, Ervik M et al. Globocan 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancerbase No. 11. Vol. 2013, International Agency for Research on Cancer: Lyon, France, 2012.
8. O'Connell JB, Maggard MA, Liu JH et al. A report card on outcomes for surgically treated gastrointestinal cancers: are we improving? J Surg Res 2004;121:214.
9. Polednak AP. Trends in survival for both histologic types of esophageal cancer in US. surveillance, epidemiology and end results areas. Int J Cancer 2003;105:98.
10. Spechler SJ, Goyal RK. The columnar-lined esophagus, intestinal metaplasia, and Norman Barrett. Gastroenterology 1996;110:614.
11. Yousef F, Cardwell C, Cantwell MM et al. The incidence of esophageal cancer and high-grade dysplasia in Barrett's esophagus: a systematic review and meta-analysis. Am J Epidemiol 2008;168:237–249.
12. Montgomery E, Goldblum JR, Greenson JK et al. Dysplasia as a predictive marker for invasive carcinoma in Barrett esophagus: a follow-up study based on 138 cases from a Diagnostic Variability Study. Hum Pathol 2001;32:379–388.
13. Moyes L, Going J. Still waiting for predictive biomarkers in Barrett's oesophagus. J Clin Pathol 2011;64:742–750.
14. Phoa KYN, van Vilsteren FG, Pouw RE et al. 1004 radiofrequency ablation in Barrett's esophagus with confirmed low-grade dysplasia: interim results of a European Multicenter Randomized Controlled Trial (SURF). Gastroenterology 2013;144:S–187.
15. Gatenby P, Ramus J, Caygill C et al. Routinely diagnosed low-grade dysplasia in Barrett's oesophagus: a population-based study of natural history. Histopathology 2009;54:814.
16. Fleischer D, Overholt B, Sharma V et al. Endoscopic radiofrequency ablation for Barrett's esophagus: 5-year outcomes from a prospective multicenter trial. Endoscopy 2010;42:781.
17. Lyday W, Corbett F, Kuperman D et al. Radiofrequency ablation of Barrett's esophagus: outcomes of 429 patients from a multicenter community practice registry. Endoscopy 2010;42:272–278.
18. Fritcher EGB, Brankley SM, Kipp BR et al. A comparison of conventional cytology, DNA ploidy analysis, and fluorescence in situ hybridization for the detection of dysplasia and adenocarcinoma in patients with Barrett's esophagus. Hum Pathol 2008;39:1128–1135.
19. McManus DT, Olaru A, Meltzer SJ. Biomarkers of esophageal adenocarcinoma and Barrett's esophagus. Cancer Res 2004;64:1561–1569.
20. Bird–Lieberman EL, Dunn JM, Coleman HG et al. Population-based study reveals new risk-stratification biomarker panel for Barrett's esophagus. Gastroenterology 2012;143:927–935.
21. Jin Z, Cheng Y, Gu W et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. Cancer Res 2009;69:4112–4115.
22. Li X, Galipeau PC, Sanchez CA et al. Single nucleotide polymorphism-based genome-wide chromosome copy change, loss of heterozygosity, and aneuploidy in Barrett's esophagus neoplastic progression. Cancer Prev Res 2008;1:413–423.
23. Sikkema M, Kerkhof M, Steyerberg EW et al. Aneuploidy and over-expression of Ki67 and P53 as markers for neoplastic progression in Barrett's esophagus: a case control study. Am J Gastroenterol 2009;104: 2673–2680.
24. Davey HM, Kell DB. Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. Microbiol Rev 1996;60:641–696.
25. Michelson AD. Flow cytometry: a clinical test of platelet function. Blood 1996;87:4925–4936.
26. Dunn JM, Mackenzie GD, Oukrif D et al. Image cytometry accurately detects DNA ploidy abnormalities and predicts late relapse to high-grade dysplasia and adenocarcinoma in Barrett's oesophagus following photodynamic therapy. Br J Cancer; 2010:1021608.
27. Fleskens SJHM, Takes RP, Otte-Höller I et al. Simultaneous assessment of DNA ploidy and biomarker expression in paraffin-embedded tissue sections. Histopathology 2010;57:14.
28. Reid BJ, Levine DS, Longton G et al. Predictors of progression to cancer in Barrett's esophagus: baseline histology and flow cytometry identify low- and high-risk patient subsets. Am J Gastroenterol 2000;95: 1669–1676.
29. Khan M, Bui HX, del Rosario A et al. Role of DNA content determination by image analysis in confirmation of dysplasia in Barrett's esophagus. Mod Pathol 1994;7:169–174.
30. Digital Slides and Third Party Data Interchange. Aperio Technologies 2006.
31. Hamilton PW, Wang Y, McCullough SJ. Virtual microscopy and digital pathology in training and education. APMIS 2012;120:305.
32. Wang Y, Williamson KE, Kelly PJ et al. Surfaceslide: a multitouch digital pathology platform. PLoS One 2012;7:e30783.
33. Isse K, Lesniak A, Grama K et al. Digital transplantation pathology: combining whole slide imaging, multiplex staining and automated image analysis. Am J Transplant 2012;12:27–37.
34. Wang Y, Savage K, Grills C et al. A TMA de-arraying method for high throughput biomarker discovery in tissue research. PLoS One 2011;6: e26007.
35. Yang L, Chen W, Meer P et al. High throughput analysis of breast cancer specimens on the grid. Medl Image Comput Comp Assistd Intervi 2007;10:617.
36. Wang Y, Crookes D, Eldin OS et al. Assisted diagnosis of cervical intraepithelial neoplasia (CIN). IEEE J Sel Topics Signal Process 2009; 3:112.
37. Wang Y, McCleary D, Wang C-W et al. Ultra-fast processing of gigapixel tissue microarray images using high performance computing. Anal Cell Pathol 2010;33:271–285.
38. Dunn J, Hveem T, Pretorius M et al. Comparison of nuclear texture analysis and image cytometric DNA Analysis for the assessment of dysplasia in Barrett's oesophagus. Br J Cancer 2011;105:1218–1223.
39. Bacus JW, Method and apparatus for automated analysis of biological specimens. U.S. Patent No. 5,526,258, Google Patents, 1996.
40. Sapi Z, Hendricks JB, Pharis PG et al. Tissue section image analysis of breast neoplasms. Evidence of false aneuploidy. Am J Clin Pathol 1993;99:714–720.
41. Huang Q, Yu C, Klein M et al. DNA index determination with Automated Cellular Imaging System (ACIS) in Barrett's esophagus: comparison with CAS 200. BMC Clin Pathol 2005;5:7.
42. Huang Q, Yu C, Zhang X et al. Comparison of DNA histograms by standard flow cytometry and image cytometry on sections in Barrett's adenocarcinoma. BMC Clin Pathol 2008;8:5.
43. Yu C, Zhang X, Huang Q et al. High-fidelity DNA histograms in neoplastic progression in Barrett's esophagus. Lab Invest 2007;87:466–472.
44. Zhang X, Huang Q, Goyal RK et al. DNA ploidy abnormalities in basal and superficial regions of the crypts in Barrett's esophagus and associated neoplastic lesions. Am J Surg Pathol 2008;32:1327–1335.

45. Bhat S, Coleman HG, Yousef F *et al.* Risk of malignant progression in Barrett's esophagus patients: results from a large population-based study. J Natl Cancer Inst 2011;103:1049–1057.

46. Ahmad J, Arthur K, Maxwell P *et al.* A cross sectional study of p504s, CD133, and Twist expression in the esophageal metaplasia dysplasia adenocarcinoma sequence. Dis Esophagus 2015;28:276–282.

47. Naik S, Doyle S, Agner S *et al.* Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium 2008: Paris, France, 14–17 May 2008, pp 284–287.

48. Gurcan MN, Pan T, Shimada H *et al.* Image analysis for neuroblastoma classification: segmentation of cell nuclei. Conf Proc IEEE Eng Med Biol Soc 2006;1:4844–4847.

49. Mouroutis T, Roberts SJ, Bharath AA. Robust cell nuclei segmentation using statistical modelling. Bioimaging 1998;6:79–91.

50. Fernand M. Topographic distance and watershed lines. Signal Process 1994;38:113–125.

51. Connolly C, Fleiss T. A study of efficiency and accuracy in the transformation from RGB to CIELAB color space. IEEE Trans Image Process 1997;6:1046.

52. Zuiderveld K. Contrast limited adaptive histogram equalization. Graphics Gems IV. Academic Press Professional: Paul Heckbert, San Diego, CA, USA, 1994, pp 474–485.

53. Barber CB, David PD, Hannu H. The quickhull algorithm for convex hulls. ACM Trans Math Softw 1996;22:469–483.

54. Gonzalez R, Woods R. Digital Image Processing2nd EditionPrentice Hall, 2002, pp 655–659.

55. Neumann A, Lorenz C. Statistical shape model based segmentation of medical images. Comput Med Imaging Graph 1998;22:133–143.

56. Bocking A, Giroud F, Reith A. Consensus report of the ESACP task force on standardization of diagnostic DNA image cytometry. European Society for Analytical Cellular Pathology. Anal Cell Pathol 1995;8:67–74.

57. Haroske G, Giroud F, Reith A *et al.* 1997 ESACP consensus report on diagnostic DNA image cytometry. Part I: basic considerations and recommendations for preparation, measurement and interpretation. European Society for Analytical Cellular Pathology. Anal Cell Pathol 1998;17:189–200.

58. Jonker A, Geerts W, Chieco P *et al.* Basic strategies for valid cytometry using image analysis. Histochem J 1997;29:347.

59. Krajewska M, Smith LH, Rong J *et al.* Image analysis algorithms for immunohistochemical assessment of cell death events and fibrosis in tissue sections. J Histochem Cytochem 2009;57:649–663.

60. Jutting U, Gais P, Rodenacker K *et al.* MIB-1, AgNOR and DNA distribution parameters and their prognostic value in neuroendocrine tumours of the lung. Image Analysis & Stereology 2011;19:39–43.

61. Bacus JW, Bacus JV. A method of correcting DNA ploidy measurements in tissue sections. Mod Pathol 1994;7:652–664.

62. Dangou JM, Kiss R, DePrez C *et al.* Heterogeneity of DNA ploidy, proliferation index and nuclear size in human colorectal carcinomas. Anal Quant Cytol Histol 1993;15:23–31.

63. Lindberg LG. Nuclear DNA ploidy in mammary carcinomas; using nuclear size as co-parameter reveals more complex patterns. Anal Cell Pathol 1992;4:389–394.

64. Suzuki R, Kohno H, Suzui M *et al.* An animal model for the rapid induction of tongue neoplasms in human c-Ha-ras proto-oncogene transgenic rats by 4-nitroquinoline 1-oxide: its potential use for preclinical chemoprevention studies. Carcinogenesis 2006;27:619–630.

65. Gschwendtner A, Mairinger T. DNA measurement on histologic slides. Does it work on human tissue? Anal Quant Cytol Histol 1997;19:294.

66. Susnik B, Poulin N, Phillips D *et al.* Comparison of DNA measurement performed by flow and image cytometry of embedded breast tissue sections. Anal Quant Cytol Histol 1995;17:163–171.

67. Freed JA. Improved correction of quantitative nuclear DNA (ploidy) measurements in tissue sections. Anal Quant Cytol Histol 1999;21:103–112.