Correction of stain variations in nuclear refractive index of clinical histology specimens

Shikhar Uttam
Rajan K. Bista
Douglas J. Hartman
Randall E. Brand
Yang Liu
Correction of stain variations in nuclear refractive index of clinical histology specimens

Shikhar Uttam,² Rajan K. Bista,² Douglas J. Hartman,² Randall E. Brand,² and Yang Liu²

¹ University of Pittsburgh, Department of Medicine and Department of Bioengineering, Biomedical Optical Imaging Laboratory (BOIL), Pittsburgh, Pennsylvania 15232
² University of Pittsburgh School of Medicine, Department of Pathology, Pittsburgh, Pennsylvania 15213
³ University of Pittsburgh, Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, Pittsburgh, Pennsylvania 15232

Abstract. For any technique to be adopted into a clinical setting, it is imperative that it seamlessly integrates with well-established clinical diagnostic workflow. We recently developed an optical microscopy technique—spatial-domain low-coherence quantitative phase microscopy (SL-QPM) that can extract the refractive index of the cell nucleus from the standard histology specimens on glass slides prepared via standard clinical protocols. This technique has shown great potential in detecting cancer with a better sensitivity than conventional pathology. A major hurdle in the clinical translation of this technique is the intrinsic variation among staining agents used in histology specimens, which limits the accuracy of refractive index measurements of clinical samples. In this paper, we present a simple and easily generalizable method to remove the effect of variations in staining levels on nuclear refractive index obtained with SL-QPM. We illustrate the efficacy of our correction method by applying it to variously stained histology samples from animal model and clinical specimens. © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.3650306]

Keywords: microscopy; refractive index; interferometry; spectroscopy.

1 Introduction

Pathology remains the gold standard for cancer diagnosis. Conventional pathology relies on the microscopic examination of cell morphology and tissue architectural characteristics using a regular bright-field microscope. Due to the diffraction-limited optical resolution, however, it can only detect structural alterations at micrometer-scale, which may not be present in the early stages of disease progression. Recently, the cellular structural properties at the nanoscale level have shown great potential for detecting cancer in histologically normal cells with a greater sensitivity than conventional pathology. Changes in these nanostructural properties are highly sensitive to molecular alterations associated with tumorigenesis,¹,² thereby providing potential cancer diagnostic information.

Quantitative phase microscopy has emerged as a powerful tool in quantifying the subcellular structure,³–⁷ in which the changes in the refractive index can be accurately quantified to characterize these structural properties.³–⁷ We recently developed a novel optical microscopy system, spatial-domain low-coherence quantitative phase microscopy (SL-QPM) that can detect refractive index within a single cell nucleus with nanoscale sensitivity.¹⁰ We have demonstrated the superior ability of SL-QPM-derived nanoscale refractive index from the cell nucleus to detect cancer from cells labeled as “normal” or “indeterminate” by expert pathologists.¹⁰,¹¹ Our clinical studies have shown great promise to improve the cancer detection of conventional pathology in multiple tumor types, including breast, esophageal, pancreatic, and colorectal cancer.¹²–¹⁵

To facilitate the clinical translation of the SL-QPM system into traditional pathology laboratories and integration with the pathology workflow, we adapted our SL-QPM system for clinical histology specimens (glass-slide–based) prepared according to routine clinical protocols without any additional processing.¹⁶ Consequently, SL-QPM can be easily integrated with the existing workflow of current pathology laboratories. In addition, the ability to work with standard clinical histology specimens also permits the direct correlation with the current gold-standard of conventional pathological features. Thus, the SL-QPM-derived new diagnostic markers could be used in conjunction with standard morphology-based pathology.

Standard clinical histology specimens use hematoxylin and eosin (H&E) stains to enhance the contrast between the cell nucleus and cytoplasm. Hematoxylin is a nuclear dye along with an oxidizing agent. The oxidized and alum-enriched hematoxylin binds with nuclei acids and nucleoproteins (e.g., histones), producing a deep purplish blue color.¹⁶–¹⁸ The cytoplasm on the other hand is generally eosinophilic and is counterstained by the alcoholic solution of eosin to an orange-pink color.¹⁶–¹⁸ The amount of stain levels in the cell nucleus and cytoplasm can vary significantly based on personal preferences of individual pathologists as well as stain variability. In a cell nucleus, the variations in the amount of H&E stain manifest as variations in the H&E uptake by the cell nucleus. As a result, cell nuclei with low uptake appear darkly stained, while cells with high uptake appear lightly stained.
The variability in the H&E stain could lead to stain-induced changes in the nuclear refractive index, thus affecting the measurements of the nuclear refractive index by our SL-QPM system. This may in turn compromise our ability to detect cancer in routine clinical specimens with high accuracy and sensitivity. In this paper, we present a simple and easy to use correction method to remove the stain-induced refractive index variations in the cell nucleus. The proposed method is based on a simple modification to the well-established empirical linear model relating changes in dry cell nuclear mass density to the corresponding nuclear refractive index changes. We first develop the correction model using a set of calibration tissue histology samples, and then validate the model through an independent testing sample set. We further evaluate the efficacy of our proposed method for histology specimens from a mouse model of colorectal carcinogenesis and human breast tissue biopsies.

2 Materials and Methods

2.1 Spatial-Domain Low-Coherence Quantitative Phase Microscopy

The SL-QPM system has been described in detail in our previous publications.10,11,19 It uses reflection-mode common-path interferometry equipped with spectroscopic detection and a low spatial-coherence thermal light source to generate an optical path length (OPL) image of the cell nucleus under observation. The reflectance-mode common-path interferometry configuration suppresses external noise allowing SL-QPM to be sensitive to nanometer changes in optical path length. Furthermore, the use of a broadband source along with a low numerical aperture objective results in a low spatially-coherent illumination that serves as a virtual aperture to produce a speckle-free OPL image. Finally, spectroscopic detection allows us to analyze the spectrum of the interference signal between the backscattered reference wave and the backscattered sample wave to generate an OPL image of the cell nucleus under observation.

The SL-QPM system records a three-dimensional spectroscopic interference intensity data cube \( [I(x, y, k); k = \text{free space wavenumber}, (x, y) \text{ corresponds to the specific pixel of the image}] \). The spectroscopic interference data cube is mathematically approximated by

\[
I(x, y, k) = |E_r(x, y, k)|^2 + |E_s(x, y, k)|^2 + 2|E_r(x, y, k)||E_s(x, y, k)|\cos[\phi(x, y, k)].
\]  

(1)

where \( E_r(x, y, k) \) and \( E_s(x, y, k) \) are the electric fields of the reference and sample beam, respectively, and \( \phi(x, y, k) \) is the phase difference between them. The pixel-wise Fourier transform of \( I(x, y, k) \) along the direction—after removing the bias term—gives us \( I_F(x, y, z') \), where \( z' \) is the optical path length. The amplitude of \( I_F(x, y, z') \), \( |I_F(x, y, z')| \), is used to find the prominent peak corresponding to the OPL of interest, \( z_p \). The phase map of the cell nucleus at the OPL of interest is then given by

\[
\Phi(x, y) = \angle I(x, y, z')]_{z=z_p}.
\]  

(2)

It is important to note that this phase map captures the phase difference. The total phase, however, has to also account for absolute phase \( \Phi_0(x, y) = kz_p \). We, therefore, write the total phase as

\[
\Phi(x, y) = \Phi_0(x, y) + \Phi_1(x, y),
\]  

(3)

and the corresponding OPL image as

\[
\text{OPL}(x, y) = \Phi(x, y)/(2k).
\]  

(4)

The factor 2 in the denominator of Eq. (4) accounts for the double path length due to the reflectance-mode configuration. The free-space wavenumber \( k \) corresponds to \( \lambda_0 = 550 \text{ nm} \). We use this wavelength because it is the center wavelength of the source and has the best signal efficiency.3 Given our knowledge of sample thickness \( L \), the refractive index image of the cell nucleus is finally obtained using

\[
\text{OPL}(x, y) = n(x, y)L.
\]  

(5)

Here, \( n(x, y) \) is the refractive index at a specific location for the sample under observation.

2.2 Correction Model for Stain-Induced Refractive Index Variation

The relation between stain and refractive index follows from the observation that most samples of interest can be optically characterized by a complex refractive index \( (m = n + i\kappa) \). The real part—known simply as the refractive index, \( n \)—affects the phase during light propagation inside the sample and the imaginary part, \( \kappa \) is responsible for attenuation of light within the sample. When the complex refractive index has frequency (or wavelength) dependence, then its real and imaginary parts are coupled through the Kramers–Kronig (KK) relations.20 The H&E stain of histology specimens changes the absorption coefficient \( \alpha(w) \) of clinical histology specimens, resulting in \( \kappa \) being modified through the relation \( \kappa(w) = \alpha(w)/(2\pi w) \), where \( c \) is the speed of light in vacuum, and \( w \) denotes the frequency dependence. This change in the imaginary part is in turn reflected as a change in the refractive index through the KK relations. It can also be understood in terms of the binding of hematoxylin and eosin molecules to DNA and proteins resulting in alteration of the measured refractive index of the cell nucleus.

It is common that routine clinical histology specimens have variations in the amount of H&E stains in the cell nucleus and cytoplasm. Recognizing that inherent stain variation could affect the accuracy of SL-QPM-derived refractive index measurement, we aim to develop a simple and clinically applicable refractive index correction method to remove the stain-induced variation in refractive index.

Theoretically, the well-established KK relations provide the ideal basis for computing variations in refractive index of H&E stained samples. However, their accuracy is limited by the narrow bandwidth of the measured absorption spectrum. This limitation either requires absorption data extrapolation beyond the measured spectral range or the assumption that the narrow bandwidth data is sufficient.12–24 The former assumption requires a valid extrapolation model, while the latter puts strict restrictions on the behavior of the absorption spectrum outside of the known range. Consequently, both are not practical for clinical samples.

We, therefore, employ a different strategy. We develop a simple model that derives from the empirically well-established linear relationship between the change in refractive index of a protein or nucleic acid solution and their concentration or mass density.25–28 This relationship was further validated by the work of Lee et al.29 and Popescu et al.30 among others.
Mathematically, it is described as

$$\Delta n = \alpha_r C,$$

(6)

where $\Delta n$ is the change in refractive index within the cell nucleus due to nucleoprotein or nucleic acid concentration $C$, and $\alpha_r$ is a constant of proportionality, referred to as the specific refraction increment. The specific refraction increment measures the increase in refractive index of the nucleic acid/nucleoprotein solution for a 1% increase in nuclear concentration.

The binding of H&E stain with nucleic acid/nucleoprotein (hematoxylin being the dominant stain) results in the absorption of light that can be quantified using Beer–Lambert law. We, therefore, interpret Eq. (6) in terms of Beer–Lambert law. Specifically, we divide and multiply the right-hand side of Eq. (6) by the unknown, but constant, molar extinction coefficient $\varepsilon$ of H&E stains to get

$$\Delta n = \left(\frac{\alpha_r}{\varepsilon}\right) (\varepsilon C) = \beta \alpha r.$$

(7)

This operation does not fundamentally change Eq. (6), but it allows us to express $\Delta n$, which we will refer to as stain-induced change in refractive index, in terms of the absorption coefficient $\alpha = \varepsilon C$, and a constant of proportionality factor $\beta$ that we refer to as the stain-modified specific refraction increment. The absorption coefficient $\alpha$ can be easily obtained by taking the measurement in the transmission-mode. According to Beer–Lambert law

$$A = \log_{10}(I_0/I) = \varepsilon LC,$$

(8)

where $A$ is absorbance, $I_0$ is the incident light intensity, and $I$ is the transmitted light intensity. The sample thickness $L$ is experimentally controlled using microtome sectioning. The absorption coefficient is then given by $\alpha = A/L$. Note that the absorption coefficient is shown without wavelength dependence because we calculate it only for the central wavelength ($\lambda_0 = 550$ nm) of the source.

We compute the absorption coefficient for each pixel $(x,y)$ of the image, resulting in an absorption coefficient image $\alpha(x, y)$. We, therefore, re-write Eq. (7) as

$$\Delta n(x, y) = \beta(x, y)\alpha(x, y).$$

(9)

The stain-induced change in refractive index, $\Delta n$, can be expressed as $\Delta n(x, y) = n_r(x, y) - n_0(x, y)$, where $n_r(x, y)$ is the observed refractive index, and $n_0(x, y)$ is the refractive index after correcting for stain-induced variation. Substituting this expression in Eq. (9) we get the final form of our correction model

$$n_r(x, y) = n_0(x, y) + \beta(x, y)\alpha(x, y).$$

(10)

To determine the stain-modified specific refractive increment constant $\beta$, we perform a calibration experiment using a sample set of various amounts of H&E stains described in Sec. 3.

### 3 Results and Discussion

#### 3.1 Experimental Determination of Stain-Modified Specific Refractive Increment Constant

We prepare a calibration sample set which consists of a series of standard histology slides with different amounts of H&E stains representing the range of variation that could be encountered in routine clinical specimens. Specifically, we use a piece of small intestine tissue removed from a healthy C57BL mouse sacrificed at the age of 6 weeks. The small intestine is washed with phosphate buffered saline before processing. A segment of the washed small intestine is cut and processed with standard tissue histology processing protocol: 10% formalin fixation, paraffin embedded and cut into 15 (4-μm thick) serial sections using a microtome. Each tissue section is mounted onto a glass slide,
deparaffinized, stained with H&E, and coverslipped. The serial sectioning of the same healthy tissue sample ensures that the variation is due to variation in stain uptake, rather than pathological difference. Although we focus on the analysis of cell nucleus, in which hematoxylin is the dominant nuclear stain, the potential effect of eosin—as the counter stain for cytoplasm—on the nucleus cannot be discarded. We prepare a total of 15 histology slides with different stain amounts of H&E, by varying stain concentrations and staining time of hematoxylin and eosin. There are 5 different hematoxylin levels (very light, light, normal, dark, very dark) and 3 different eosin levels (none, light, normal). Figure 1 shows the bright-field images of all 15 histology slides. The wide variations in hematoxylin and eosin are visually discernible.

For each of the 15 slides, we take both reflection- and transmission-mode measurements. As explained in Sec. 2, the reflection-mode measurement allows us to calculate the observed refractive index, while the transmission-mode measurement allows the calculation of the absorption coefficient. Both are calculated with respect to the central wavelength of 550 nm.

Given that all cells are from a serial section of the same tissue piece from a healthy mouse, we assume that the nuclear refractive index without the effect of stains, or stain-corrected refractive index \( n_{\text{nc}} \), has the same value. According to Eq. (10), the stain-modified specific refraction increment constant \( \alpha \) can be determined by the slope of the linear relationship between the observed refractive index \( n_o \) and the absorption coefficient \( \alpha \). However, since computing a constant pixel-wise \( \beta \) image to correct the observed nuclear refractive index of pixel-wise co-registered transmission and reflectance-mode image of every cell is very time-consuming and not practical, we simplify Eq. (10) by correcting for the average nuclear refractive index. The simplified form is

\[
\langle n_o(x, y) \rangle = \langle n_{\text{nc}}(x, y) \rangle + \hat{\beta} \langle \alpha(x, y) \rangle,
\]

where \( \langle \rangle \) denotes the averaging over the entire cell nucleus. This is a reasonable simplification as in most cases we are interested in the average nuclear refractive index of a cell. We stress that this simplification is not a limitation of our method, but a consequence of practical consideration. Figure 2 shows the observed refractive index plotted as a function of the absorption coefficient, along with the linear regression fit. For a significance level of 0.05, the \( p \)-values for the model parameters \( \hat{\beta} \) and \( \langle n_{\text{nc}}(x, y) \rangle \) are 0.0223 and 2.1 \times 10^{-27}, respectively, confirming the statistical significance of the linear relationship between \( \langle n_{\text{nc}}(x, y) \rangle \) and \( \langle \alpha(x, y) \rangle \). We estimate the model parameters \( \hat{\beta} \) and \( \langle n_{\text{nc}}(x, y) \rangle \) using simple linear regression to be 5.795 \times 10^{-8} and 1.540, respectively.

We now have a simple refractive index correction method to remove the stain variation. For a given histology specimen, we obtain the stain-induced refractive index change \( \Delta n \) by calculating the average absorption coefficient \( \alpha \) from the transmission-mode measurement, and multiplying it with the model parameter \( \hat{\beta} \). The final corrected refractive index \( n_c \) is obtained by subtracting this stain-induced refractive index correction factor from the observed nuclear refractive index.

### 3.2 Independent Validation of the Correction Model

Before showing the efficacy of our correction model, we first validate our method using an independent validation set of histology slides obtained from small intestine tissue from a different healthy C57BL mouse at 6 weeks from what was used in the above experiment. The validation set of histology slides are stained with unknown variations in H&E stain. Figure 3 visually depicts three stained samples from the validation set with the corresponding average absorption coefficients that corroborate the stain variations: darkly stained samples have a higher absorption coefficient, while the lightly stained samples have a lower absorption coefficient. Figure 4 shows the measured nuclear refractive indices as a function of the absorption coefficients before applying our correction model. The variations in the nuclear refractive index are clearly visible. Since all of these histology slides are from the serial sections of the same piece of healthy small intestine tissue, we assume...
the same value of nuclear refractive index. The variation is primarily due to the effect of variation in staining. Figure 4 shows the corresponding corrected nuclear refractive indices after applying the correction model. As the stain-induced variations have been removed, the nuclear refractive index remains stable for varying absorption coefficients in the validation histology sample set. Indeed, as shown in Fig. 4, the linear regression fit for the corrected nuclear refractive index as a function of absorption coefficient gives a very small slope of $-1.2 \times 10^{-9}$, indicating the validity of our correction model.

### 3.3 Experiments With an Animal Model of Carcinogenesis and Human Breast Tissue

To evaluate the efficacy of our correction model in studying various tissue research and clinical histology specimens, we investigate the SL-QPM-derived nuclear refractive index in an animal model of carcinogenesis and human breast tissue prepared with the standard clinical protocols.

We use a mouse model of intestinal carcinogenesis, the multiple intestinal neoplasia (Min) mouse model, which has a germ line adenomatous polyposis coli (APC) gene mutation that causes the mouse to spontaneously develop multiple intestinal adenomas (tumors). It is a well-established animal model to study colorectal carcinogenesis. Three sets of aged-match mice are used: 6-week wild-type mice to mimic the healthy condition, and 6-week and 4.5-month APCMin mice to mimic the early stage and advanced stage of carcinogenesis. The H&E-stained histology slides from small intestine epithelial tissue of the mice are prepared following the same protocol described in Sec. 3.1. Figure 5(a) shows the histology images of the normal tissue from the wild-type mice and dysplastic tissue (marked by an expert gastrointestinal pathologist) from Min mice at 6 weeks and 4.5 months. Due to stain variations, the cell nuclei of three groups have different colors. The corresponding values of nuclear refractive index are shown in Fig. 5(b). The values of nuclear refractive index clearly correlate with stain levels.

![Fig. 5](https://example.com/fig5.png)

(a) Histology images of the normal tissue from the wild-type mice and dysplastic tissue (marked by an expert gastrointestinal pathologist) from Min mice at 6 weeks and 4.5 months. (b) Average nuclear refractive index before applying the stain-induced correction model, and (c) average nuclear refractive index after correcting for the H&E stain-induced correction model. The average is taken over 60 cells from each mouse.

After applying our stain-induced refractive index correction model, we observe [see Fig. 5(c)] a progressive change in these three groups ($P$-value < 0.05). Such trend is consistent with the development of carcinogenesis. The increased nuclear refractive index in cancer tissue has also been previously shown by other investigators, in agreement with our findings.

We also analyze the clinical histology specimens of human breast tissue from a total of 30 patients, categorized into 3 groups: normal tissue from 10 healthy patients who underwent reduction mammoplasty, tissue from benign nonproliferative breast lesions (10 patients), and malignant tissue from 10 patients with invasive breast cancer. Figures 6(a) and 6(b) show the histology images of breast tissue of each of these three groups, and the corresponding average nuclear refractive index before applying the stain-induced correction model. The nuclear stains exhibit significant color variation, from nearly transparent to dark blue. Similar to our findings in the animal model, the values of nuclear refractive index correlate with stain levels.

![Fig. 6](https://example.com/fig6.png)

(a) Bright-field images of histology specimens from breast tissue biopsies of a normal healthy patient, a patient with a benign lesion, and a cancer patient with malignant lesions. (b) Average nuclear refractive index before correction, and (c) average nuclear refractive index after correcting for H&E stain-induced variation. Each group has 10 patients.
corresponding to the highest refractive index. Both normal and benign cells have similar low risk of cancer development, but their nuclear refractive indices are unexpectedly different due to stain variations. Similarly, the nuclear refractive index of malignant cells is unexpectedly less than that of benign cells. However, after correcting the stain-induced refractive index variation, as shown in Fig. 6(c), we found that nuclear refractive index from normal and benign cells has similar value ($P = 0.9$), in agreement with the similar cancer risk of these two pathological entities. Furthermore, the nuclear refractive index from malignant cells is significantly higher than both normal and benign cells ($P < 0.05$). Such trend is in good agreement with breast tumorigenesis.

4 Conclusion
We have developed a simple and effective method to correct for H&E stain-induced refractive index variations of the cell nucleus. The method can work easily with our SL-QPM system. Most importantly, this approach allows our SL-QPM system to be applied on routine histology specimens prepared with standard clinical protocol. Such ability integrates well with the existing workflow of clinical diagnostic pathology. We have validated the performance of our model with an independent set of histology specimens, and shown its efficacy with an animal model of carcinogenesis and clinical breast tissue samples in different pathological states (normal, benign, and malignant). We clearly show that the importance of applying the stain-induced refractive index correction model to properly characterize the properties of nuclear refractive index in various biological systems for basic research and clinical pathology diagnosis. Although our initial pilot clinical studies show certain promise, the efficacy of this technique needs to be further validated in a large patient population and different tumor types.

Acknowledgments
We acknowledge the funding support from National Institute of Health (Grant No. R21CA152935), Wallace H. Coulter foundation, and University of Pittsburgh Medical Center.

References
29. H. Lee, V. Richards, and C. Klauser, “Dry mass and cell area changes in ehrlich mouse ascites carcinoma cells after complement-fixation
reaction measured by interference microscopy,” Cancer Res. 20(10), 1415–1421 (1960).