Published in final edited form as:

Technology (Singap World Sci). 2015 March ; 3(1): 27–31. doi:10.1142/S2339547815200010.

Stain-less staining for computed histopathology

David Mayerich^{1,2}, Michael J. Walsh^{2,3,4}, Andre Kadjacsy-Balla⁵, Partha S. Ray⁶, Stephen M. Hewitt⁷, and Rohit Bhargava^{2,3,8}

²Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

³Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 and Carle Foundation Hospital, Urbana, IL 61801, USA.

⁵Department of Pathology and University of Illinois Cancer Center, University of Illinois at Chicago, Chicago, IL 60612, USA

⁶Carle Cancer Center, Carle Foundation Hospital, Urbana, IL and Department of Surgery, and University of Illinois Cancer Center, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

⁷Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA.

⁸Departments of Mechanical Science and Engineering, Electrical and Computer Engineering, Chemical and Biomolecular Engineering and Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

Abstract

Dyes such as hematoxylin and eosin (H&E) and immunohistochemical stains have been increasingly used to visualize tissue composition in research and clinical practice. We present an alternative approach to obtain the same information using stain-free chemical imaging. Relying on Fourier transform infrared (FT-IR) spectroscopic imaging and computation, stainless computed histopathology can enable a rapid, digital, quantitative and non-perturbing visualization of morphology and multiple molecular epitopes simultaneously in a variety of research and clinical pathology applications.

INNOVATION

Examination of the molecular and microscopic structure in tissue, or histopathology, is essential to understanding disease in modern biomedical science and performing accurate

© World Scientific Publishing Co./Imperial College Press

Correspondence should be addressed to R.B. (rbx@illinois.edu).

¹Present address: Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004, USA.

⁴Present address: Department of Pathology, University of Illinois at Chicago, Chicago, IL 60612, USA.

AUTHOR CONTRIBUTIONS

RB conceived the idea; MJW, DM, AB, SMH and RB designed the study; MJW recorded data and performed spectroscopic analysis; DM performed mathematical and image analyses; AB and SMH provided histologic guidance; PR assisted in the study design and result evaluation; RB, MJW and DM wrote the first draft; All authors wrote the final manuscript.

clinical diagnoses. Histo-pathology involves the use of dyes and stains to illustrate morphology and molecular content via optical microscopy. We show that the same is possible without dyes using computational analysis of spectral data from infrared (IR) microscopy. By acquiring a single IR spectral image, we are able to reproduce staining patterns for multiple epitopes that competes with and, in some cases, outperforms the quality of staining achieved with present reagents. The work should lead a facile “molecular dial-in” to image multiple markers from a single tissue slice that can simplify and improve molecular information in pathology, leading directly to digital molecular pathology.

NARRATIVE

Clinical decision-making and biomedical research rely significantly on imaging the architecture and morphology of tissues. Tissues have little contrast in brightfield optical imaging (**Fig. 1a**), however, and require the use of stains or dyes to provide contrast. Contrast agents include those that highlight morphology as well as those that highlight specific molecular species¹, usually using immunohistochemical (IHC) techniques². The use of staining is especially critical in histopathologic analyses that are the gold standard for the diagnoses of many diseases, including cancer, and for almost all tissue-related research. Usually, the pathologist examines tissue architecture and histology (cell types) to provide an initial diagnosis. In some cases, confirmatory IHC staining may be employed to improve diagnostic accuracy or to guide appropriate therapy. While tremendous advances have been made in precise measurements using quantitative techniques^{3,4}, the time required for staining and the expense of obtaining multiple stains may be a limiting factor in some settings, while inconsistent staining due to a variety of technological and tissue factors may be a practical limitation⁵. Often, staining patterns need to be interpreted in the context of multiple stains or appropriate morphologic visualization in order to be effective, compounding the need for multiplex marker staining and further interpretation.

As opposed to imaging with dyes or stains, a new paradigm is emerging in the form of chemical imaging. Here, imaging is combined with spectroscopy to provide both the morphologic detail of microscopy and the molecular selectivity of spectroscopy. In particular, the optical frequencies of the mid-infrared (IR) region of the spectrum are in the range of molecular vibrational frequencies. Hence, if a molecular species is present, light is absorbed in a pattern consistent with its molecular constitution. While a microscope provides straightforward imaging capability⁶, the absorption spectrum can be used as a pattern of composition as well as a readout of metabolic activity⁷. Combined with numerical algorithms, the data can be used to recognize a range of structures important for research and clinical applications, including bacteria, isolated aberrant cells, tissue structures and disease⁸. The contrast mechanism between different components in a sample is straightforward, instrumentation is readily available and underlying principles are largely understood⁶. While the data have been used by numerous groups to relate spectral properties to cell type and disease state⁹, the approach has largely been that this information is in addition to or complementary to conventional laboratory and clinical analyses¹⁰. Color coded information in the IR¹¹, Raman and mass spectrometry, among others, has provided excellent visualization but has often not related to images used clinically. Spectrometric technologies are considered different from clinical methods. Here we show that

spectroscopic data can also be used to generate the same information as classical dyes and various molecular stains.

RESULTS

The brightfield optical image of tissue (**Fig. 1a**) has little contrast. Conventionally, the use of H&E stain allows a visualization of tissue morphology using optical microscopy (**Fig. 1b**). In our approach, infrared spectroscopic imaging data is recorded from the same unstained tissue in the form of two spatial dimensions and a spectral dimension (**Fig. 1c**). The spectrum at every pixel contains specific features that are indicative of the molecular content of the sample, morphology and optical system^{12,13}. These features are then used as input for statistical pattern recognition algorithms¹⁴, in this case an artificial neural network (ANN) model¹⁵ that relates the biochemical input to molecular or dye parameters. The ANN transforms the recorded spectroscopic data at every pixel into color values or staining intensity found in histological stains imaged using bright-field microscopy (**Fig. 1d**). Finally, the predicted stain or dye values are used to generate a computed stain image (**Fig. 1e**). Details of the approach are provided in the Methods section. The computed stain image faithfully reproduces the staining pattern of H&E images that are important for recognition. Though useful, generation of H&E images from optical data has been proposed earlier but can simply and cheaply be achieved in practice. Significantly, H&E stains lack specificity in terms of the functional or molecular content of the tissue.

Visualizing molecular content is significantly more expensive, time-consuming and difficult; yet, this is precisely the origin of the contrast mechanism in chemical imaging. Hence, we evaluated additional molecularly-specific stains that are indicative of tissue function and integrity. IHC stains for cell types are often used in diagnostic imaging or for specific research purposes, for example, high molecular weight (HMW) cytokeratin (epithelial-type cell), vimentin (fibroblast-like cell), smooth muscle alpha actin (myo-like cell), P63 (myoepithelial cells), CD31 (endothelial cells) and Masson's trichrome stain (collagen and keratin fibers), are commonly employed. A comparison of the physically stained and computationally stained images can be seen in **Fig. 2**. Since the process is correlative, we were exceptionally careful in training and validation. We employed high throughput tissue microarrays, in which histopathologic, clinical and patient diversity is carefully built-in. Depending on the stain type and cellular abundance, we trained our model on 600,000 to 2 million spectra from 96 patients and validate the approach in an independent set of 98 patients. Details of the tissue diversity and statistical methods are provided in the Methods section.

Using a combination of the staining results it is possible to subsequently deduce the cell types and/or molecular transformations present. The chemical content in spectroscopy, recorded once, can be used multiple times to relate to expression levels in tissue. Hence, the approach allows us to compute multiple stains from a single spectroscopic image for the same sample. Such a capability would be exceptionally useful when the tissue available is limited, e.g. in fine needle biopsies or spheroids in 3D cell culture. This capability provides a quantitative method for histology, eliminating staining variance between sample replicates and providing a means of quickly extracting a large range of histological information from a

single tissue sample. **Figure 3** demonstrates this potential by examining typical fields of view from a needle biopsy (**Fig. 3a–e**). Eliminating the need to stain can lead to faster availability of multiple stain-derived results in time-starved settings or for activities such as response to drug screening. Precious or limited samples can be imaged without perturbation and exploratory molecular staining is easily enabled.

In summary, an approach is presented that can considerably enhance the ability of researchers and clinicians to make molecularly-informed decisions in the microscopic analyses of tissue. Several future directions become apparent from this work. Dispensing with the need for stains can imply that completely non-perturbative monitoring of specific molecular or chemical species via this emerging approach to microscopy. This capability can enable in many situations, accelerate and reduce costs for multiplexed molecular analyses. From a technology perspective, the extension of these methods to high-definition IR imaging, especially with rapid imaging systems using discrete frequency lasers, can lead to higher quality images and more sensitive molecular data. Extension to other spectroscopic contrast mechanisms, such as from Raman or mass spectrometry, can be similarly accomplished. More sophisticated molecular data will likely be targeted, by improved chemical imaging technology, using multimodal approaches as well as using improved analytical methods. The stainless staining approach bridges the fields of label-free and molecularly-labeled imaging technologies, offering opportunities for hybrid approaches that make optimal use of each technology's strengths. Finally, the all-digital nature of this approach also considerably extends the emerging field of “digital pathology” to “molecular digital pathology”.

METHODS

Tissue microarrays

Tissue microarrays (TMAs) representing normal breast tissue, pre-cancer, non-malignant and malignant breast cancer were acquired from the cooperative human tissue network via the Tissue Array Research Program, NCI and from US Biomax (Rockville, MD). TMAs are advantageous for analyzing large patient sets since they provide a wide variety of cell types and disease states to ensure robust classifiers, and they have been shown to retain predictive power in translation to tissue sections for staining purposes¹⁶. The training and validation arrays consisted of 1.5 mm diameter needle core biopsies. One section was placed on an IR transparent substrate, barium fluoride (BaF₂), for mid-infrared spectroscopic imaging while serial sections were placed on glass slides and stained with H&E, Masson's trichrome and a panel of immunohistochemical stains commonly used in breast cancer. Staining was performed at the University of Illinois at Chicago using the same methods employed for routine clinical samples.

The tissues on BaF₂ had their paraffin removed by emersion in hexane at 40°C for 48 hours. IR imaging was performed in transmission-mode using a Perkin Elmer Spotlight 400 Fourier Transform IR Imaging system in transmission mode at 6.25 µm spatial resolution and 4 cm⁻¹ spectral resolution from 750 cm⁻¹ to 4000 cm⁻¹. Spectral data was processed to remove background absorbance and each pixel was independently baseline corrected. Stained sections were imaged using a Hamamatsu NanoZoomer 2.0 series scanner with 20X

objective (0.46 $\mu\text{m}/\text{pixel}$). Bright-field images were filtered and downsampled to match the resolution of the IR image and individual cores were manually aligned to construct the training set. Since the infrared and bright-field images were from adjacent sections, cell-level alignment was not possible. In order to compensate for shifted positions in high-frequency features, a Gaussian blur with 5-pixel standard deviation was applied to the bright-field image.

Surgical specimens

Formalin-fixed paraffin-embedded (FFPE) surgical specimens were sectioned and placed on MirrIR low-emissivity microscope slides (Kevley Technologies) and imaged using a Perkin Elmer Spotlight 400 Fourier Transform IR Imaging system in reflection mode at 6.25 μm pixel size and 8 cm^{-1} spectral resolution from 750 cm^{-1} to 4000 cm^{-1} . Spectral data was processed to remove background absorbance and each pixel was independently baseline corrected. The sections were then stained with Haematoxylin and Eosin (H&E) or Masson's trichrome. Stained sections were then imaged using a Zeiss Axiovert 200M Microscope with a 2.5X Plan-Neofluar objective and Zeiss AxioCam MRm color camera (2.39 $\mu\text{m}/\text{pixel}$). The bright-field images were then filtered and resampled to match the resolution of the IR spectroscopic images. Since the same tissue was imaged in both IR and bright-field, an affine transformation was then applied to bring all pixels into alignment. Transformation and alignment was performed using the GNU Image Manipulation Program (GIMP).

Tissue preparation

Tissue samples were collected from anonymized, formalin-fixed, paraffin embedded breast biopsies taken with 0.6 mm and 1.6 mm needles. Clinical samples were obtained from the University of Chicago Department of Pathology and Biomax, Inc. Approximately 5 μm sections prepared for infrared imaging were placed on low-emissivity MirrIR coated glass slides from Kevley Technologies. Neighboring sections were placed on standard glass slides and stained using a BioGenex i6000 Automated Staining System. All proper protocols were followed to ensure patient anonymity.

Imaging

Unstained sections placed on low-emissivity glass were imaged using a Perkin-Elmer Spotlight 400 Mid-Infrared imaging system with a 16-element linear array detector. Imaging was performed in reflection mode at 4 cm^{-1} spectral resolution using a 15X, 0.5 NA objective, providing a spatial resolution of 6.25 μm . After imaging, the sections were stained and imaged using a Zeiss Axiovert 200M fluorescence light microscope with a 2X Plan-Neofluar 0.3NA objective providing a spatial resolution of 1 μm . Color images were collected using an AxioCam HRC color CCD camera. Neighboring histological sections used for qualitative validation were imaged using a Hamamatsu NanoZoomer.

Neural network training and simulation

We created a mapping from spectral metrics to RGB color values using a neural network model. We constructed a 2-layer feed-forward neural network. The output layer uses a linear transfer function and the input layer uses a sigmoid transfer function. The number of input

nodes and internal nodes is specified by the user. The number of input nodes corresponds to the number of metrics used in the regression. Increasing these values generally reduces error on the training data but can reduce performance during validation. One of the major advantages of our spectra/RGB mapping of the training data is that it provides a large number of training samples (several million spectra). This allows us to increase the number of spectral features while enforcing model generality and minimizing over-fitting for each stain we uniformly sample a tissue array, constructing a training set of 600,000 to 2 million spectrum/RGB pairs. We divide this set into two groups: 75% training and 25% validation. Training is performed using Levenberg-Marquardt minimization while the validation set is repeatedly checked to prevent over-training. If the mean-squared error of the validation set begins to increase, training is terminated.

For each stain, a slice of the data cube taken at the Amide-I peak (1650 cm^{-1}) was taken and saved as a single two-dimensional image. Individual cores from the color images were aligned to the Amide-I mosaic and resampled to match the resolution of the IR image. Since both tissue samples were identical, only affine transformations were necessary to achieve alignment between the IR and brightfield images. For each stain, 1 M sample spectra were randomly selected across 80 cores. These samples, along with their corresponding brightfield color values, were used to fit the following nonlinear neural network model:

$$\mathbf{W}_1^T \tanh \left[\mathbf{W}_0^T \left[\mathbf{P}^T (\bar{s} - \bar{\mu}) \right] + \bar{b}_0 \right] + \bar{b}_1 = \bar{C}_0$$

where the hyperbolic tangent is an element-wise operation on the vector x :

$$\tanh \bar{x} = \begin{bmatrix} \tanh(x_0) \\ \tanh(x_1) \\ \tanh(x_2) \\ \dots \\ \tanh(x_n) \end{bmatrix}$$

This equation represents a neural network with a single hidden layer and three output values corresponding to the color values of the matching brightfield image. The tensors $\mathbf{W}$ and $\mathbf{b}$ contain the weight and bias values for the hidden (0) and output (1) layers. The matrix $\mathbf{P}$ is the spectral feature basis and $\bar{\mu}$ is the mean spectrum. We used 151 spectral metrics as input features ($\mathbf{W}_0 = [151 \times 10]$). The weight and bias values were computed using Levenberg-Marquardt (LM) backpropagation, which is an iterative nonlinear fitting algorithm with a dynamic step size. The mean-squared error (MSE) between the computed output values and the corresponding brightfield color values was used as a performance metric for LM minimization. Over fitting was minimized by using 25% of the original training pairs as a validation set. The validation set was simulated after each iteration of LM. If the MSE of this validation set began to diverge, training was terminated. Qualitative validation was performed using cores independent from the training set.

ACKNOWLEDGEMENTS

The work was supported by NIH grants 1R01CA138882, 4R00LM011390 and 2R01EB009745.

REFERENCES

1. Bratthauer GL, et al. Combined E-cadherin and high molecular weight cytokeratin immunoprofile differentiates lobular, ductal, and hybrid mammary intraepithelial neoplasias. *Hum. Pathol.* 2002; 33:620–627. [PubMed: 12152161]
2. Dabbs DJ. *Diagnostic Immunohistochemistry: Theranostic and Genomic Applications — Expert Consult* (Elsevier Health Sciences. 2010
3. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* 2001; 23:291–299. [PubMed: 11531144]
4. Taylor CR, Levenson RM. Quantification of immunohistochemistry — issues concerning methods, utility and semiquantitative assessment II. *Histopathology.* 2006; 49:411–424. [PubMed: 16978205]
5. Goldstein NS, Hewitt SM, Taylor CR, Yaziji H, Hicks DG. Recommendations for improved standardization of immunohistochemistry. *Appl. Immunohistochem. Mol. Morphol.* 2007; 15:124–133. [PubMed: 17525622]
6. Bhargava R. Infrared spectroscopic imaging: The next generation. *Appl. Spectrosc.* 2012; 66:1091–1120. [PubMed: 23031693]
7. Ellis DI, Goodacre R. Metabolic fingerprinting in disease diagnosis: Biomedical applications of infrared and Raman spectroscopy. *Analyst.* 2006; 131:875–885. [PubMed: 17028718]
8. Diem M, et al. A decade of vibrational micro-spectroscopy of human cells and tissue (1994–2004). *Analyst.* 2004; 129:880–885. [PubMed: 15457314]
9. Baker MJ, et al. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* 2014; 9:1771–1791. [PubMed: 24992094]
10. Lasch, P.; Kneipp, J. *Biomedical Vibrational Spectroscopy.* John Wiley & Sons; 2008.
11. Fernandez DC, Bhargava R, Hewitt SM, Levin IW. Infrared spectroscopic imaging for histopathologic recognition. *Nat. Biotech.* 2005; 23:469–474.
12. Kazarian SG, Chan KA. ATR-FTIR spectroscopic imaging: Recent advances and applications to biological systems. *Analyst.* 2013; 138:1940–1951. [PubMed: 23400222]
13. Mayerich D, et al. On the importance of image formation optics in the design of infrared spectroscopic imaging systems. *Analyst.* 2014; 139:4031–4036. [PubMed: 24936526]
14. Bhargava R. Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology. *Anal. Bioanal. Chem.* 2007; 389:1155–1169. [PubMed: 17786414]
15. Haykin, SS. *Neural Networks: A Comprehensive Foundation.* Prentice Hall; 2007.
16. Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. *Lab. Invest.* 2000; 80:1943–1949. [PubMed: 11140706]

Figure 1.

Microscopic analysis of tissue — present status and proposed approach. **(a)** Brightfield microscopic imaging of an unstained tissue section, provides little contrast within the sample. **(b)** Contrast is generated by the application of stains, most commonly using hematoxylin and eosin (H&E) as shown here, to visualize tissue morphology. **(c)** In contrast to present practice, we record spectroscopic imaging data from unstained tissue. The data can be conceptualized as a spectrum that reports the overall chemical composition of every spatial location or a stack of chemically-specific images. **(d)** The underlying chemical composition can be converted to conventional pathology images in an objective and automated manner using computer algorithms, such as artificial neural networks. This approach is used in **(e)** to reproduce the H&E stained image (similar to **(b)**) without staining.

Figure 2.

Molecular imaging (three sample panel on the left) can be reproduced by chemical imaging (right panel). In addition to H&E stained images (**a**), we extend the concept of stainless staining to molecularly-specific stains, including (**b**) Masson's trichrome stain (collagen and keratin fibers) (**b**) high molecular weight (HMW) cytokeratin (epithelial-type cell), (**c**) smooth muscle alpha actin (myo-like cell) and (**d**) vimentin (fibroblast-like cell). Each spot is 1.4 mm in diameter.

Figure 3.

The same sample can be stained with many computational stains as well as provide consistently uniform, high-quality results. Sample dimensions that may be common for a biopsy needle show (a) a physically-stained H&E image of a series of samples and the corresponding images derived from an unstained sample — (b) chemical absorbance image at the Amide I vibrational mode, (c) computed H&E image, (d) computed Masson's trichrome, and (e) computed cytokeratin-stained image. The scale bar represents 800 μm . For fields of view typical under a microscope, the digital stains (f, bottom) help prevent

artifacts sometimes observed in conventional H&E staining (**f**, top). The scale bar represents 700 μm . Comprehensive staining is possible for tumor resections (**g**) where multiple stained images can be generated from the sample for different regions and at different scales. Images can be overlaid, merged or multiply highlighted. The scale bar represents 2.6 μm .