Molecular constituents of colorectal cancer metastatic to the liver by imaging infrared spectroscopy

James V. Coe[a]*, Zhaomin Chen[a], Ran Li[a], Steven V. Nystrom[a], Ryan Butke[b], Barrie Miller[b], Charles L. Hitchcock[b], Heather C. Allen[a,b]*, Stephen P. Povoski[c], and Edward W. Martin Jr.[c]
[a]The Ohio State University, Department of Chemistry and Biochemistry, 100 West 18th Avenue, Columbus OH 43210-1173; [b]The Ohio State University, Department of Pathology, 4132 Graves Hall, 333 W. 10th Ave, Columbus, OH 43210; [c]The Ohio State University, Department of Surgery, Division of Surgical Oncology, 410 W 10th Avenue, Columbus, OH 43210

ABSTRACT

Infrared (IR) imaging spectroscopy of human liver tissue slices has been used to identify and characterize liver metastasis of colorectal origin which was surgically removed from a consenting patient and frozen without formalin fixation or dehydration procedures, so that lipids and water remain in the tissues. First, a k-means clustering analysis, using metrics from the IR spectra, identified groups within the image. The groups were identified as tumor or nontumor regions by comparing to an H&E stain of the same sample after IR imaging. Then, calibrant IR spectra of protein, several fats, glycogen, and polyvinyl alcohol were isolated by differencing spectra from different regions or groups in the image space. Finally, inner products (or scores) of the IR spectra at each pixel in the image with each of the various calibrants were calculated showing how the calibrant molecules vary in tumor and nontumor regions. In this particular case, glycogen and protein changes enable separation of tumor and nontumor regions as shown with a contour plot of the glycogen scores versus the protein scores.

Keywords: liver tumors from colorectal cancer, bulk molecular components, FTIR imaging

1. IDENTIFYING A LIVER TUMOR IN COLORECTAL CANCER

The liver is a frequent site for metastases from cancers in other organs since the liver serves as an interface between the digestive and circulatory systems. It is the most common site for metastatic disease from colorectal cancer[1]. The liver has a high regenerative capacity and resection of liver tumors is potentially curative. One critical issue with resection is the histopathological determination of whether all of a tumor has been removed. Since infrared (IR) spectroscopy is sensitive to molecular level biochemical changes, nondestructive, and involves no labeling or staining, it has potential to someday be used as a real-time intraoperative diagnostic tool[2, 3] . IR spectroscopy on tumor samples has the potential to rapidly objectify and inform histopathological judgments. Someday surgical oncologists may employ attenuated total reflection (IR-ATR) probes to assay tissues in situ during surgical procedures. This paper examines the changes in bulk molecular constituents associated with a metastatic liver tumor of colorectal origin in order to identify IR signatures for identifying liver tumors.

Incisive observations about molecular changes in tissue associated with tumors can be extracted from the large amount of data generated by Fourier transform IR (FTIR) imaging spectroscopy[4-8]. This work involves recording an FTIR spectrum at every 6.25 μm by 6.25 μm pixel of liver tissue containing tumor within an area of 2,200 μm by 1,200 μm. The tissue contains a liver metastasis of colorectal origin which was surgically removed from a consenting patient (IRB # 2011C0085) at the time of a planned liver resection at the University Hospital (Ohio State University, Columbus, OH). The tissue was snap frozen in liquid nitrogen without formalin fixation or dehydration procedures. Cryostat sections of ~2-3 μm thickness were obtained at -20 °C. They notably still contain lipid and water and therefore have less perturbed biomolecules than with fixation. The acquisition of data has been previously described[4, 9], so only a brief overview is given herein. Imaging FTIR data sets were recorded (Perkin Elmer Spotlight 400, 16 element MCT array detector, 4 cm^{-1} resolution, 750-4000 cm^{-1} range, 16 scans per pixel, 6.25 μm

*coe.1@osu.edu; phone (614)292-9489; fax (614)292-1685

Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XIII, edited by Daniel L. Farkas, Dan V. Nicolau, Robert C. Leif, Proc. of SPIE Vol. 9328, 93280R · © 2015 SPIE
CCC code: 1605-7422/15/$18 · doi: 10.1117/12.2079884

pixel size) on four adjacent windows (each 2,200.0 μm by 300.0 μm). Each window took ~3 hr of scanning. A MATLAB function from Perkin Elmer called "fsm_load.m" by Ben Peterson was used within a home-written MATLAB routine to download the spectral data for each of the four windows and to catenate them into one large data block. Analysis starts with the construction of a matrix wherein each pixel has a row containing a corresponding IR absorption spectrum. In this case there were 352x192=67,584 image pixels, each having an IR spectrum with 1626 values.

After recording the IR data, the same slice of tissue was stained with hematoxylin and eosin (H&E)[10] as shown in Figure 1a. The tumor resides in the bottom half of the image as indicated by the glandular patterning of dark and light regions. The blue-ish regions in the top half are lymphocyte-rich areas of inflammatory response to cancer. The lighter pink areas in the top half are nontumor regions. The IR spectra at each pixel in the image was used to generate a set or vector of 20 IR metrics at each image pixel. The specific metrics are described in our previous

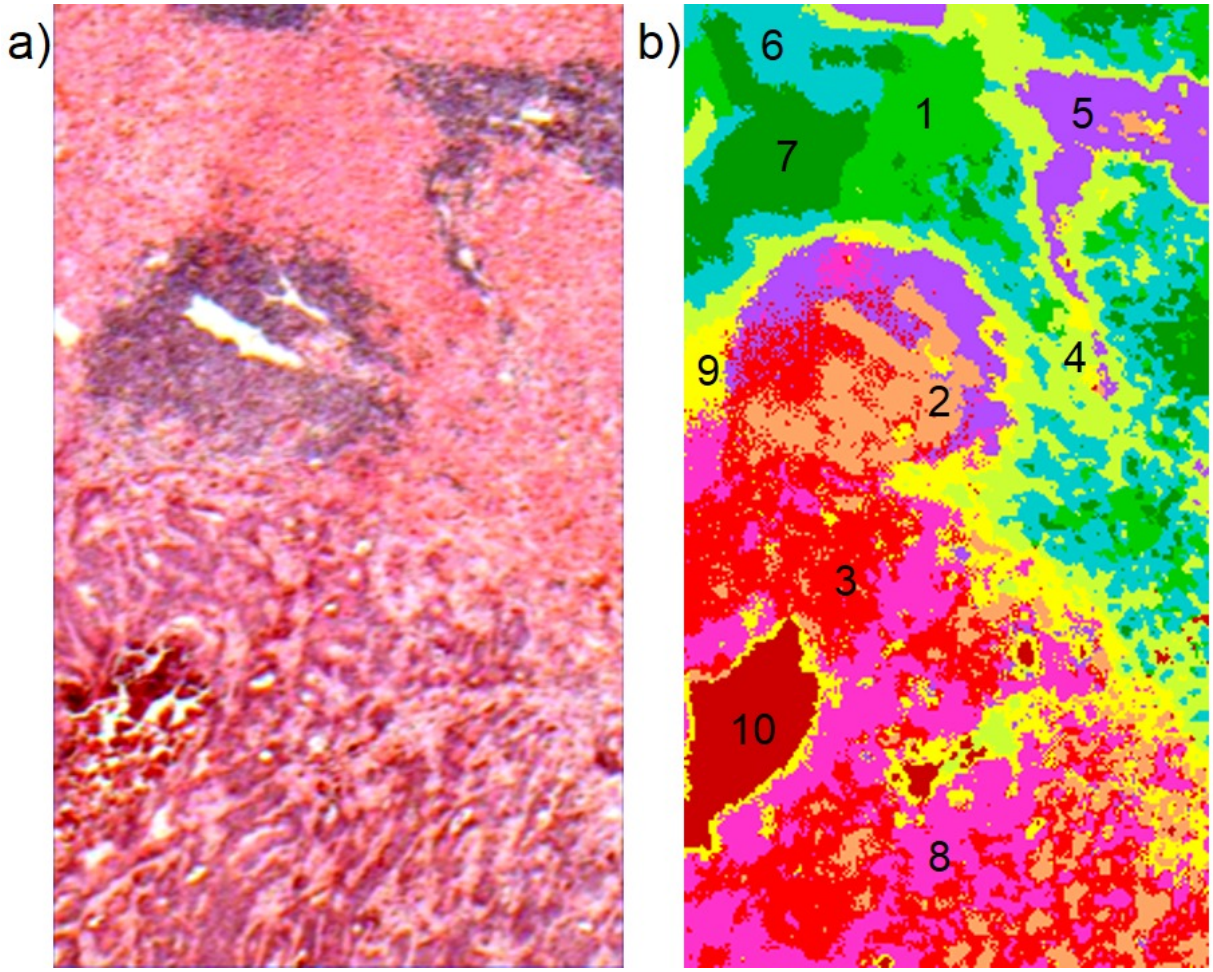

Figure 1. a) An H&E stained optical image of liver tissue (2,200 μm by 1,200 μm) with a tumor towards the bottom half as indicated by glandular patterns. The bluer regions in the top half are lymphocyte-rich regions exhibiting an inflammatory response to the cancer, while the lighter pink regions in the top half are nontumor. b) An image of the same region as in (a) by k-means clustering analysis with 10 groups using metrics from IR spectra. The tumor regions have been colored with hot colors (pink, red, orange, and yellow), the nontumor regions are colored with cool colors (teal, Kelly green, and dark green), while the transition regions are colored with yellow-green and purple. The dark red region of group 10 is one associated with blood clumping as evident by the deep red color in the corresponding region of the H&E stain.

work[4]. An image of the same region as in Figure 1a was obtained with k-means clustering analysis using 10 groups as shown in Figure 1b. The coloring of the k-means groups was accomplished by comparing to the H&E stain in Figure 1a. Hot colors were used for tumor groups, while cool colors were used for nontumor groups. A further

reduction in the number of groups is useful for tumor/nontumor decisions. Figure 2a shows a classification of the k-means groups into a smaller number of groups. Groups 2, 3, 8, and 9 were considered tumor (red), groups 1, 6, and 7 were considered nontumor (green), groups 4 and 5 were considered as transitioning between tumor and nontumor (yellow), while group 10 was clumping of erythrocytes in blood (dark red). The IR spectra for each of these groups excepting the blood are shown in Figure 2b with plot traces sharing the colors of Figure 2a. The nontumor region exhibits a slightly more intense IR spectrum (on average per pixel) than the tumor region with the most prominent differences lying in the range from 1000-1200 cm^{-1} which is likely attributable to glycogen[11-13]. A healthy liver stores energy with glycogen which is a polysaccharide of glucose. As liver cells transform to tumor, they may lose to some extent and in some regions the normal liver function of making and storing glycogen.

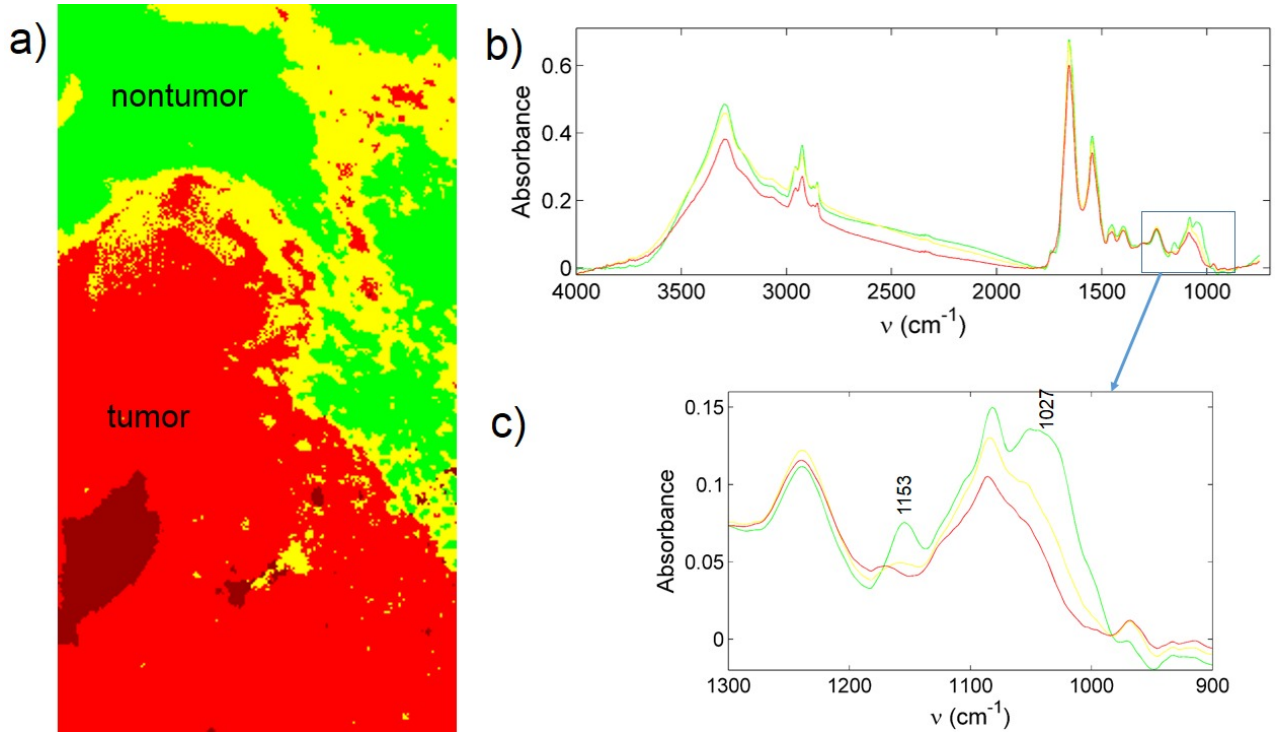

Figure 2. a) Classification of the k-means groups of Figure 1b into tumor (red), nontumor (green), transitioning (yellow), and blood clumping (dark red). b) IR spectra of the tumor (red trace), transitioning (yellow trace), and nontumor (green trace) groups. The nontumor group has a slightly stronger spectrum in general with the largest differences in the range from 1000 to 1200 cm^{-1}. c) The box in Figure 2b has been expanded to better observe the region of strong glycogen bands.

2. CALIBRANT IR SPECTRA

A number of IR spectra of bulk constituents of liver tissue have been isolated for use as molecular calibrants as shown in Figure 3. They arise primarily by differencing the IR spectra of various k-means groups or by extracting the IR spectra associated with specific regions of score plots. The "protein" spectrum shown at the bottom of Figure 3 arises from the first principal component of a principal component analysis (PCA) of all IR spectra from our liver tumor studies. This work will be presented elsewhere at a later date. The first principal component spectrum had a small ester-linked fat peak (at ~1744 cm^{-1}) which was subtracted away using ester-linked fat spectra like the ones in Figure 3. This spectrum was matched to albumin using the Perkin-Elmer Spectrum Search program and FDM's Very Large Bundle of IR libraries. Albumin has an α-helix dominated structure and constitutes ~80% of the protein in normal liver[14], so the "protein" calibrant is dominated by albumin-like protein. The strategy for isolating the spectrum of "fat 1" involved subtracting the protein bands of the IR spectra of adjacent k-means groups as described in our previous work[4]. The "fat 2" spectrum was isolated by comparing score plots of "fat 1" versus "protein" in the high fat regions. Both "fat 1" and "fat 2" show prominent C=O bands typical of ester-linked fats, each shows a CH stretch of the H-C=C group at ~3001 cm^{-1} indicating an abundance of C18 chains with one C=C double bond per

chain, and both show evidence of phosphate bands (~1250 and ~1165 cm^{-1}). The primary difference between the two "fat" calibrants involves the relative intensity of the C=O band (~1744 cm^{-1}) to the CH_2 antisymmetric (~2922 cm^{-1}) and symmetric (~2852 cm^{-1}) stretches of the hydrocarbon chains. The C=O band in "fat 1" is considerably more intense and probably contains about two C=O groups per C18 fat chain (like the glyceryl triacetyl ricinoleate used in cosmetics). The C=O band in "fat 2" is considerably less intense than the CH_2 stretches indicating about one C=O bond per C18 fat chain which is typical of phospholipids and triglycerides in biochemical textbooks. The spectrum of glycogen was isolated by subtracting group spectra as shown in Figure 2c. The subtraction of protein was imperfect leaving two artifact peaks in the amide I and II regions. The last calibrant was isolated by k-means groups near holes in the tissue. It was matched and assigned using the Perkin-Elmer Spectrum Search program and FDM's Very Large Bundle of IR libraries.

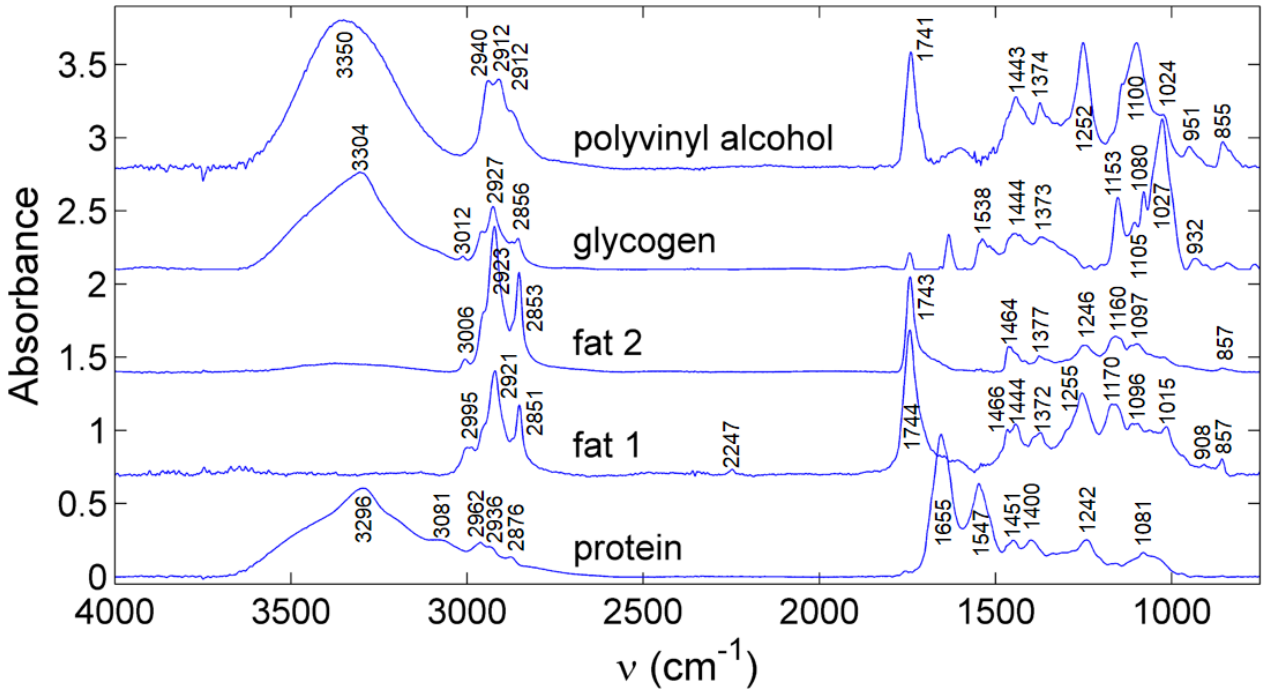

Figure 3. Calibrant IR spectra of important bulk constituents of liver tissue. The "protein" spectra was obtained from the first principal component of a PCA analysis of all of liver tissue cases. The others were obtained by finding groups rich in the desired calibrant and subtracting the protein bands of the IR spectra of a similar groups with less of the desired calibrant.

3. SCORES OF CALIBRANTS WITH LIVER SPECTRA AT IMAGE PIXELS

The process to be described is similar to the process in PCA of plotting scores of one principle component against another. Consider that the scores in a PCA analysis of IR spectra can be obtained by taking the inner product of a principle component in spectral form with the IR spectrum at a pixel in the tissue image. In the following, scores are calculated by taking the inner product of the calibrant IR spectra in Figure 3 with the IR spectra at pixels in the tissue image. This proceeds more meaningfully if both the calibrant spectra and the pixel spectra are normalized, i.e. all of the spectra are scaled so that the inner product of any spectrum with itself is one. The inner product scores of normalized IR spectra of all 67,584 image pixels with each of the normalized calibrant spectra have been presented as histograms in Figure 4. The results fall between zero and one due to the normalization. This process puts a number on how similar a calibrant's IR spectrum is to the liver tissue spectrum. The "protein" calibrant has the highest score as the protein amide I band (~1654 cm^{-1}) is the dominant band in almost every liver tissue spectrum. If a particular calibrant is to be useful for distinguishing tumor and nontumor, then one would like to see structure or multiple distributions in these histograms. The calibrant histograms with the most structure appear to be "protein", "fat 1", and "glycogen". The next step is to plot the scores of one calibrant against another (just as is done with PCA). While one would like to examine all pairwise combinations of calibrants, space precludes a complete examination. Instead, the most interesting pair, "glycogen" versus "protein" is examined in more detail in the following section.

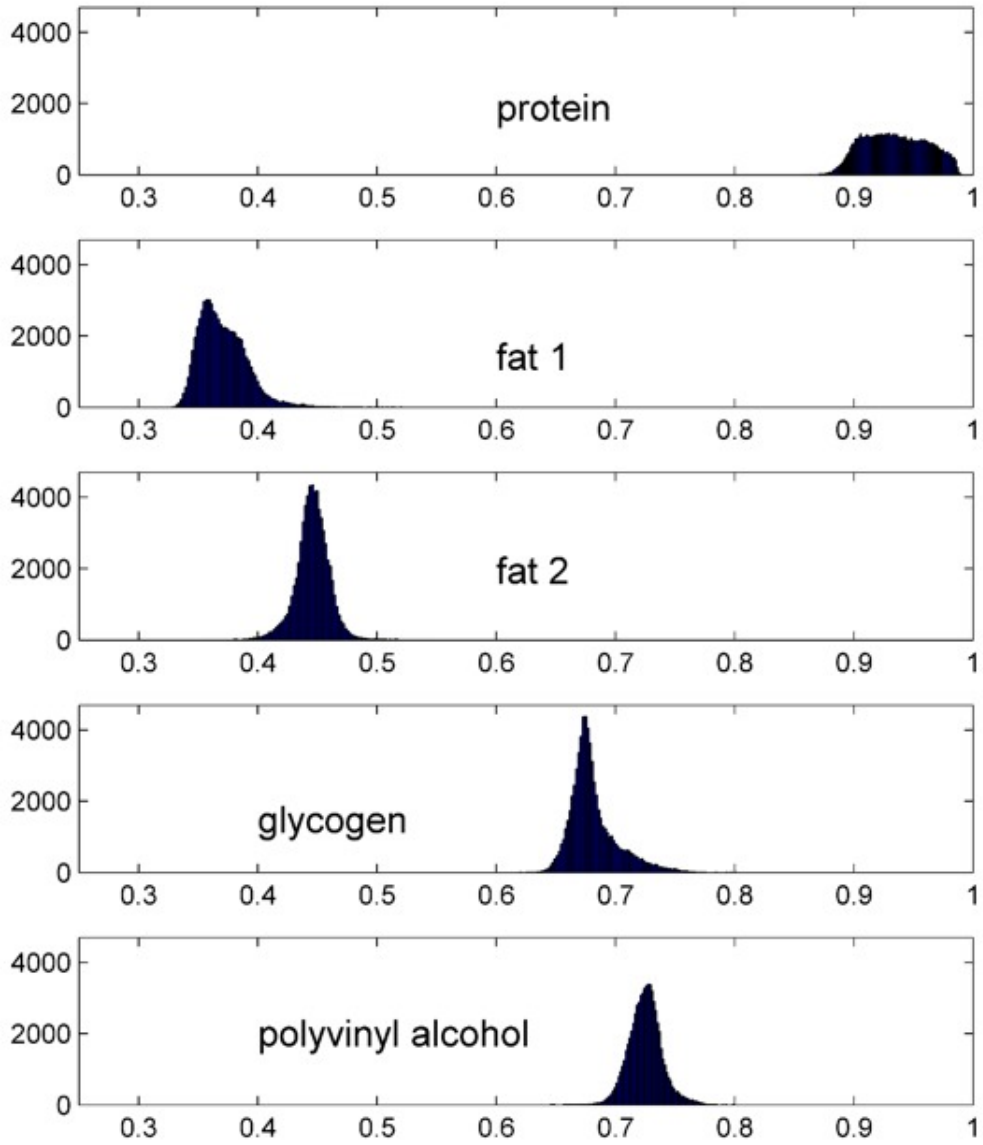

Figure 4. Histograms of inner product scores of normalized IR spectra at pixels in the tissue image with the normalized IR spectra of calibrants from Figure 3.

4. PROTEIN AND GLYCOGEN

A contour plot of the scores of "glycogen" versus the scores of "protein" is shown in Figure 5a. It overlays a dot plot of the scores which is useful in the low density regions, but overcrowded in high density regions due to the plotting of 67,584 image pixel values. The contour plot is constructed by dividing the score space into two dimensional bins and counting how many pixel scores (dots) fall into each bin. Structure within the contour plot, i.e. multiple maxima, reveals utility for differentiating tissue features. The highly structured plots in Figure 5a show the utility of these calibrants for analyzing liver tissues. Matlab routines have been constructed which allow the investigator to draw an ellipse on the contour plot. The programs determine both the average IR spectrum of all image pixels that fall within the ellipse and produce color coded bitmap images of these pixels. The colored ellipses in

Figure 5a correspond to the same color pixels in Figure 5b. The yellow ellipse falls between the tumor and nontumor regions, and the yellow pixels pick out some of the transitioning regions in Figure 2a. The red tumor and green nontumor regions differ very little in their "glycogen" scores. However, the tumor has a low glycogen region isolated with an orange ellipse, while the nontumor has a high glycogen region isolated with a blue ellipse. Furthermore, the IR spectra of Figure 2b show that there is on average more protein in the nontumor regions. However the contour plot of Figure 5a, which employs normalized spectra, shows that the nontumor region (green) has a lower "protein" score than the tumor region (red). Together, these observations suggest that the nontumor region has more protein, but that protein is a smaller fraction of all of the molecules present, i.e. there are also more molecules that are not protein in the nontumor. The favorable comparison of the image in Figure 5b to the image in Figure 2a (in both red indicates tumor, while green indicates nontumor), clearly reveals the utility of these calibrant spectra for distinguishing tumor and nontumor regions. The use of multiple calibrants can help to sort out the anatomical origins of these molecular changes as a tumor grows within the liver tissue.

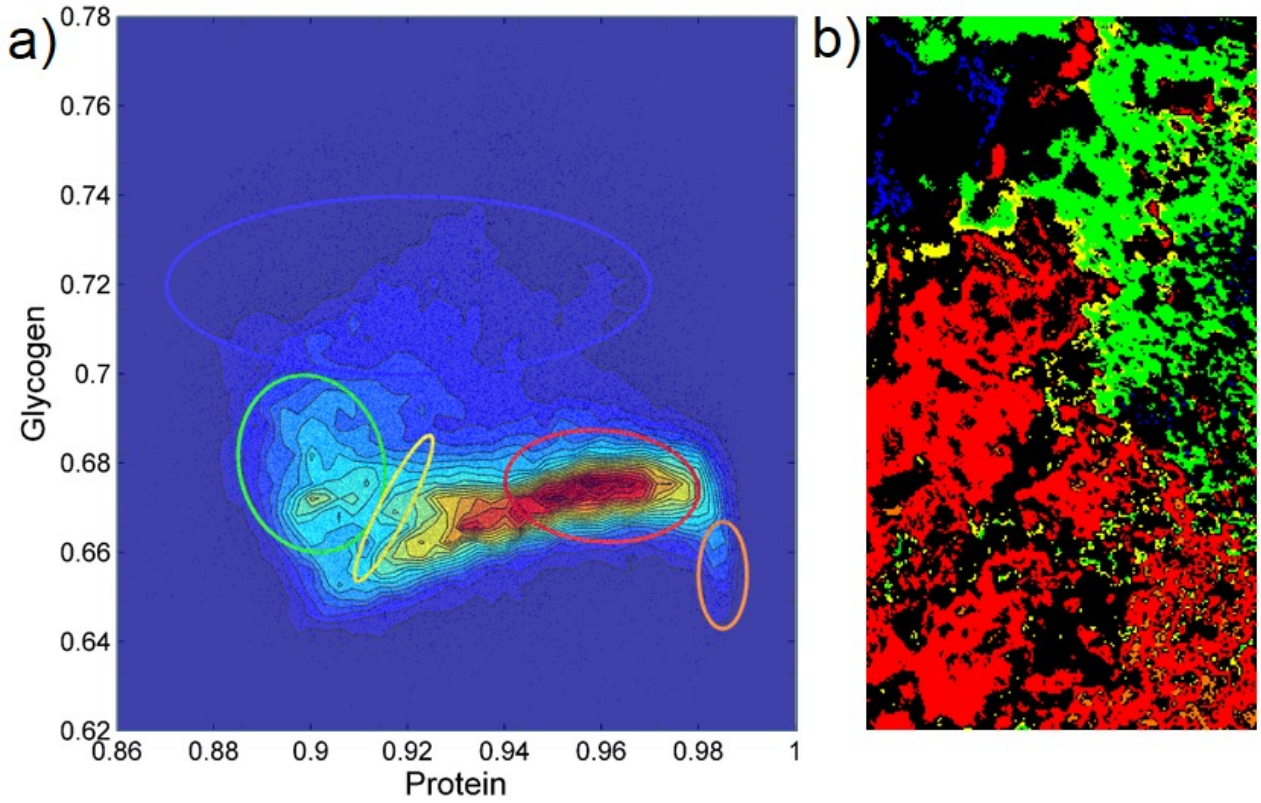

Figure 5. a) Contour plot of the normalized calibrant-pixel spectra dot product scores of "glycogen" versus "protein" which differentiate into tumor (red) and nontumor regions (green). b) The colored pixels correspond to pixels inside the matched colored ellipses on the contour plot of part (a). There is a good correspondence between the tumor and nontumor regions of this plot and Figure 2a showing the utility of these IR calibrants for differentiating tumors.

5. CONCLUSIONS

Meaningful scores for calibrants are obtained by taking the inner product of normalized calibrant spectra with normalized IR spectra at pixels in an image of liver tissue. The scores are some measure of the similarity of the calibrants with liver tissue IR spectra. The plots of the score of one calibrant against another (in a manner similar to that done with principle components) gives rise to richly structured contour plots that distinguish tumor and nontumor regions. Normalization (Euclidian) is the key to getting well-behaved contour plots that readily distinguish tumor and nontumor regions. The technique could actually be employed with the IR spectra of any calibrant, although this work featured calibrants that were extracted largely by differencing IR spectra of liver tissues. It is likely that many more

calibrant IR spectra exist that are yet to be discovered. They may prove useful in differentiating tissue samples, just as there are a great many stains that are useful in histopathology with optical microscopes. This work shows that common bulk molecular constituents of tissues can be readily assayed in the complex mixtures of tissue samples using IR imaging without stains or labels. The bulk molecular constituents reveal important changes that characterize the presence of tumors, as well as unexplored changes in bulk molecular chemistry associated with cancer.

ACKNOWLEDGMENT

We thank the National Cancer Institute and the National Institutes for Health for grant NIH R21 CA167403.

REFERENCES

[1] [Cancer Medicine, 4th Edition, Vols. I and II] Williams & Wilkins, Baltimore (1997).

[2] J. G. Wu, Y. Z. Xu, C. W. Sun *et al.*, "Distinguishing malignant from normal oral tissues using FTIR fiber-optic techniques," Biopolymers, 62(4), 185-192 (2001).

[3] M. Diem, M. Miljkovic, B. Bird *et al.*, "Applications of Infrared and Raman Microspectroscopy of Cells and Tissue in Medical Diagnostics: Present Status and Future Promises," Spectroscopy-an International Journal, 27(5-6), 463-496 (2012).

[4] Z. Chen, R. Butke, B. Miller *et al.*, "Infrared Metrics for Fixation-Free Liver Tumor Detection," The Journal of Physical Chemistry B, 117(41), 12442-12450 (2013).

[5] R. Bhargava, "Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology," Analytical and Bioanalytical Chemistry, 389(4), 1155-1169 (2007).

[6] D. C. Fernandez, R. Bhargava, S. M. Hewitt *et al.*, "Infrared spectroscopic imaging for histopathologic recognition," Nature Biotechnology, 23(4), 469-474 (2005).

[7] M. Diem, K. Papamarkakis, J. Schubert *et al.*, "The Infrared Spectral Signatures of Disease: Extracting the Distinguishing Spectral Features Between Normal and Diseased States," Applied Spectroscopy, 63(11), 307A-318A (2009).

[8] P. Lasch, W. Haensch, D. Naumann *et al.*, "Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis," Biochim Biophys Acta, 1688(2), 176-86 (2004).

[9] J. V. Coe, Z. Chen, R. Li *et al.*, "Imaging infrared spectroscopy for fixation-free liver tumor detection," Proc. SPIE, 8947(Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XII), 89470B/1-89470B/6 (2014).

[10] E. B. Prophet, [Laboratory Methods in Histotechnology] American Registry of Pathology, Washington, DC(1992).

[11] R. Gautam, B. Chandrasekar, M. Deobagkar-Lele *et al.*, "Identification of early biomarkers during acetaminophen-induced hepatotoxicity by Fourier Transform Infrared microspectroscopy," PLoS One, 7(9), e45521 (2012).

[12] J. R. Mourant, J. Dominguez, S. Carpenter *et al.*, "Comparison of vibrational spectroscopy to biochemical and flow cytometry methods for analysis of the basic biochemical composition of mammalian cells," J. Biomed. Opt., 11(6), 064024/1-064024/11 (2006).

[13] K. E. Skinner, "Estimation of yeast glycogen content from the mid-infrared spectra of yeast," J. Am. Soc. Brew. Chem., 54(2), 71-5 (1996).

[14] "First Insight into the Human Liver Proteome from PROTEOMESKY-LIVERHu 1.0, a Publicly Available Database," J. Proteome Res., 9(1), 79-94 (2010).