

# Imaging infrared spectroscopy for fixation-free liver tumor detection

James V. Coe<sup>a,\*</sup>, Zhaomin Chen<sup>a</sup>, Ran Li<sup>a</sup>, Ryan Butke<sup>b</sup>, Barrie Miller<sup>b</sup>, Charles L. Hitchcock<sup>b</sup>, Heather C. Allen<sup>a,b,\*</sup>, Stephen P. Povoski<sup>c</sup>, and Edward W. Martin Jr.<sup>c</sup>

<sup>a</sup>The Ohio State University, Department of Chemistry and Biochemistry, 100 West 18th Avenue, Columbus OH 43210-1173; <sup>b</sup>The Ohio State University, Department of Pathology, 4132 Graves Hall, 333 W. 10th Ave, Columbus, OH 43210; <sup>c</sup>The Ohio State University, Department of Surgery, Division of Surgical Oncology, 410 W 10th Avenue, Columbus, OH 43210

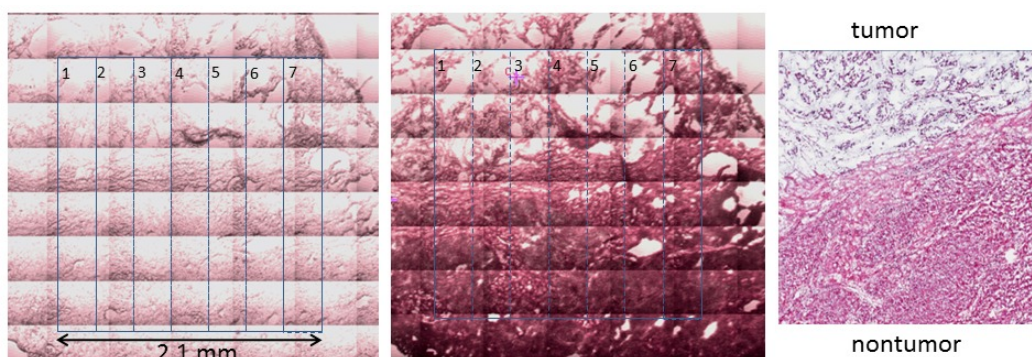
## ABSTRACT

Infrared (IR) imaging spectroscopy of human liver tissue slices has been used to identify and characterize a liver metastasis of breast origin (mucinous carcinoma) 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. Previously, a set of IR metrics was determined for tumors in fixation-free liver tissues facilitating a k-means cluster analysis differentiating tumor from nontumor. Different and more in depth aspects of these results are examined in this work including three metric color imaging, differencing for lipid identification, and a new technique to simultaneously fit band lineshapes and their 2<sup>nd</sup> derivatives in order to better characterize protein changes.

**Keywords:** liver tumors, infrared metrics, k-means cluster analysis, human liver lipid, FTIR imaging.

## 1. IMAGING FTIR AND A LIVER TUMOR

Incisive observations about molecular changes in tissue associated with tumors can be extracted from the large amount of data generated by Fourier transform infrared (FTIR) imaging spectroscopy<sup>1-4</sup>. This work involves recording an FTIR spectrum at every 6.25  $\mu\text{m}$  by 6.25  $\mu\text{m}$  pixel within a 2,100.0  $\mu\text{m}$  by 2,200.0  $\mu\text{m}$  region of a  $\sim 3$   $\mu\text{m}$  thick slice of liver tissue containing a tumor as shown in Figure 1. The tissue contains a liver metastasis of breast origin (mucinous carcinoma) which was surgically removed from a consenting patient (IRB # 2011C0085) at the time of a planned liver



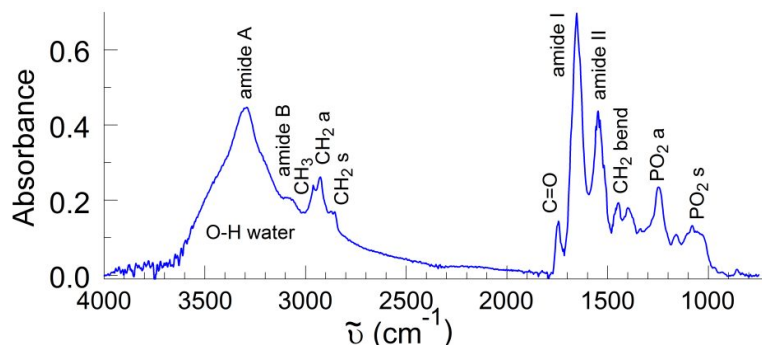
**Figure 1.** Optical images of liver tissue with a tumor. The left panel is an optical microscope image of a slice of liver tissue with a tumor on a ZnSe window for IR spectroscopic imaging. Seven windows were imaged separately and then merged into one 2.1000 mm x 2.2000 mm window (shown as a box) for analysis. The middle panel shows the same tissue on ZnSe as the first panel after H&E staining. Note that the bottom is darker indicating nontumor tissue, while the top is lighter indicating the tumor. The third panel displays a segment cutout from a standard Aperio H&E stain of a neighboring slice at approximately the same place definitively showing that the tumor is at the top.

resection at the University Hospital (Ohio State University, Columbus, OH). Most importantly, the tissue was snap frozen in liquid nitrogen without formalin fixation or dehydration procedures. Cryostat sections of  $\sim 2$ -3  $\mu\text{m}$  thickness  
\*coe.1@osu.edu; phone (614)292-9489; fax (614)292-1685

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<sup>1</sup>, 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<sup>-1</sup> resolution, 750-4000 cm<sup>-1</sup> range, 16 scans per pixel, 6.25 μm pixel size) on seven adjacent windows (each 2,200.0 μm vertically by 300.0 μm horizontally) of the tissue slice as shown on the left side of Figure 1. Each window took 3 hr of scanning for a total of 21 hr of scanning time. After recording the IR data, the same slice of tissue was stained with hematoxylin and eosin (H&E)<sup>5</sup> as shown in the middle panel of Figure 1. The tumor is clearly at the top of the imaged area. A subsequent cryostat slice was treated with H&E stain and processed as a virtual Aperio slide. A subsection was extracted, oriented, and scaled (right hand side of Figure 1) to match, as best as possible, the IR imaging region. Note that the tumor is more clearly delineated. The commercial software does not have k-means cluster analysis, so MATLAB® routines were used to extract and manipulate the spectral data.

## 2. K-MEANS CLUSTER ANALYSIS, IR METRICS, AND MATLAB

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 seven windows and to concatenate them into one large three dimensional matrix of spectra transmittances,  $data(i,j,k')$ , where  $i$  is a pixel index for the image row,  $j$  is a pixel index for the image column, and  $k'$  is an index stepping through the IR spectrum. The resulting imaging area has 352 rows of pixels and 336 columns of pixels yielding 118,272 pixels in the image. Each image pixel is associated with a distinct IR spectrum and the average IR spectrum from all 118,272 pixels is given in Figure 2. There exist definitive vibrational bands for proteins and ester-linked fats, but the rest is a tangle of many types of molecules. With such extensive averaging, the signal-to-noise ratios are good and differencing may reveal features not currently apparent.



**Figure 2.** Average IR spectrum of all 118,272 image pixels for a liver tissue slice that is roughly half tumor.

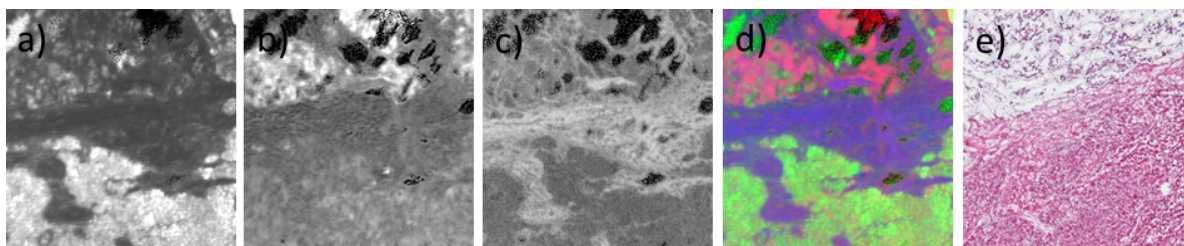
Considering that there are 1626 wavenumber steps in the spectrum at each of 118,272 pixels, there are more than 192 million measured transmittances to evaluate. IR metrics are devised to reduce the information to a more manageable and meaningful set. We have previously considered a set of 64 metrics with a determination of the 20 most important ones in this specific case<sup>1</sup> as shown in Table 1. The IR metrics are mostly ratios of absorbance at one band ratio-ed

**Table 1.** IR metrics for tumor vs nontumor. The metric is the ratio of absorbance at the cm<sup>-1</sup> values given in the ratio.

| Name | Ratio     | Name | Ratio     | Name | Ratio     | Name | Ratio            |
|------|-----------|------|-----------|------|-----------|------|------------------|
| L1   | 1744/1548 | L6   | 2916/1548 | L11  | 1016/1080 | L16  | 1744/1162        |
| L2   | 1744/1244 | L7   | 1120/1020 | L12  | 1252/1544 | L17  | 1080/3290        |
| L3   | 1742/1256 | L8   | 2924/1544 | L13  | 1024/1080 | L18  | 1556/1548        |
| L4   | 1160/1548 | L9   | 1080/1548 | L14  | 1080/1244 | L19  | 1012/1256        |
| L5   | 1516/1236 | L10  | 1744/1162 | L15  | 1516/1582 | L20  | (1144-1182)/1544 |

to absorbance at another with corrections for baseline effects. Metric values were calculated at each pixel and scaled into a 0 to 255 range for presentation as a black-and-white bitmap image. Images for metrics L1, L14, and L15 are shown in Figure 3a-c. These were chosen to highlight different regions of tissue. A composite image was created using L14 as red, L1 as green, and L18 as blue as shown in Figure 3d. It is compared to an H&E stain of a subsequent

slice in Figure 3e. Both the IR composite and the H&E stain define tumor (top) and nontumor (bottom). The transition region between tumor and nontumor exhibits more contrast in the IR composite image than in the H&E stain due to the *L18* metric which is the ratio of absorbance at  $1556\text{ cm}^{-1}$  to absorbance at  $1548\text{ cm}^{-1}$ . Both of these wavenumbers lie within the amide II band profile which is a band of the protein backbone. The IR spectrum is detecting protein changes in the critical region of transition between tumor and nontumor – ones that are not as readily evident in the H&E stain. Careful examination of the H&E stain image in the regions that are purple in the IR color composite, finds a texture associated with fibrous stroma cells that are differentiated in the H&E image, but not with the stain colors. The IR work (which is unstained) is clearly capable of distinguishing a tumor and appears useful in characterizing bulk molecular changes in the transition from nontumor to tumor.



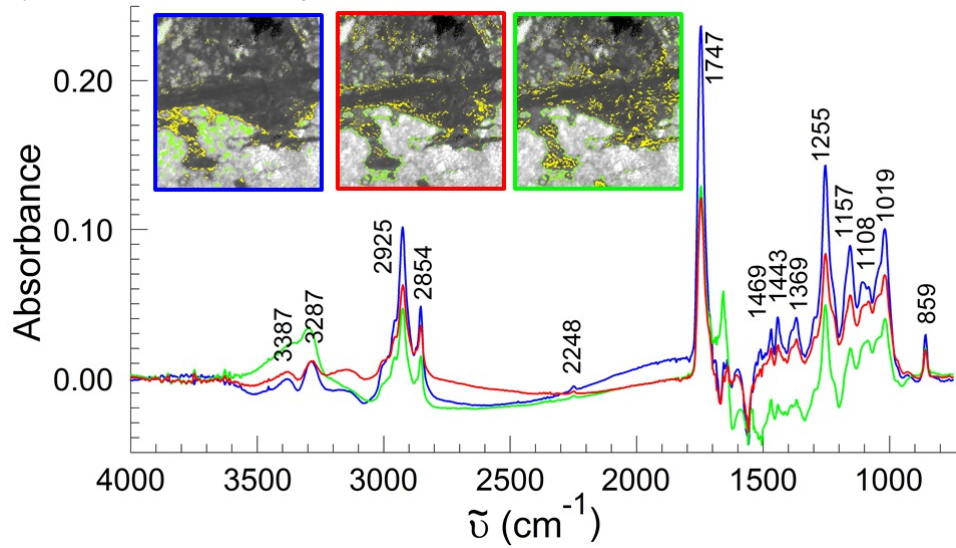
**Figure 3.** Three metric imaging. a) image from metric *L1* ( $A_{1744\text{cm}^{-1}}/A_{1548\text{cm}^{-1}}$ ), b) image from metric *L14* ( $A_{1080\text{cm}^{-1}}/A_{1244\text{cm}^{-1}}$ ), c) image from metric *L18* ( $A_{1556\text{cm}^{-1}}/A_{1548\text{cm}^{-1}}$ ), d) color image using *L14* as red, *L1* as green, and *L18* as blue. e) H&E stain of a neighboring slice.

K-means cluster analysis was previously employed<sup>1</sup> to distinguish 25 groups based on the 20 IR metrics of Table 1, i.e. the 1626 IR spectral points at each image pixel have been reduced to 20 metric values. K-means cluster analysis uses metric scores at each image pixel to calculate a similarity (or a “distance”) between a particular image pixel and the average metric scores of the group, i.e. a group centroid. The routine changes membership of image pixels in groups to minimize the sum of the sum of “distances” for each group. The output is an assignment of each image pixel to a group with similar metrics. After the K-means analysis is finished, groups can be organized by “distances” between group centroids (hierarchical analysis) as shown in our previous work<sup>1</sup>. MATLAB routines were written to obtain the average IR spectrum of each group. These group spectra are themselves averages of ~5000 individual spectra each and so potentially offer good signal-to-noise upon differencing group IR spectra.

### 3. LIPIDS

The largest peak in the IR spectrum of liver tissue is the amide I band of protein backbone, so we became curious as to whether it could be subtracted away revealing the next most prominent material after protein. The metric *L1* ( $A_{1744\text{cm}^{-1}}/A_{1548\text{cm}^{-1}}$  or the ratio of ester linked fat to protein) does a good job of defining the nontumor region as the bright region at the bottom of Figure 3a. Clearly the nontumor region has more ester linked fat than the tumor region. There is great contrast between the nontumor region and the middle of this image, so we looked for pairs of K-means groups that spanned that interface. These groups are shown in green (nontumor) and yellow (tumor) and have been overlaid with the greyscale *L1* metric image in the inset of Figure 4. The IR spectra of each of these groups were differenced (green-yellow, i.e. nontumor-tumor) with a scaling (within 10% of 1) that best subtracted the amide I protein band. The resulting difference spectra are plotted with color coding (color of the outside of the images matches the color of the corresponding traces of the difference spectra). The protein does not subtract away perfectly revealing that there are important changes going on with the proteins. In fact these changes have a much smaller bandwidth than the amide I band, but this is the subject of the next section. The remaining upward-going spectra are remarkably narrow and well defined compared to the overall tissue spectrum in Figure 2. These spectra correspond largely to triglycerides. The prominent  $\text{CH}_2$  stretches ( $2925$  and  $2854\text{ cm}^{-1}$ ) provide evidence of long hydrocarbon chains. The  $\text{C=O}$  band at  $1744\text{ cm}^{-1}$  suggests prominence of triglyceride rather than, for instance, phospholipid which must also be present. Most interestingly, the  $\text{C=O}$  band is twice as intense as the  $\text{CH}_2$  stretches. This suggests that there may be twice as many  $\text{C=O}$  groups per fat chain as, for instance, triglycerides in textbooks. Since these triglycerides would be removed under normal tissue fixation procedures, the fact that such distinct spectra are revealed in this work is interesting. There are other notable details including the peak at  $859\text{ cm}^{-1}$  whose intensity is linearly correlated with the intensity of large peak at  $1744\text{ cm}^{-1}$ . It is narrow and distinct after protein subtraction. There is another peak that is linearly correlated with both the  $1744$  and  $859\text{ cm}^{-1}$  peak intensities that is not observed in the overall spectrum in Figure 2. Upon differencing, however, a peak at  $2248\text{ cm}^{-1}$  is small but evident in each of the three difference spectra being considered. This is the region of  $\text{C}\equiv\text{C}$  and  $\text{C}\equiv\text{N}$  triple bonds. These are not at all expected – they are hardly mentioned in biochemistry textbooks. There is not much information about triglycerides in snap frozen tissues, so further studies on other

patients are underway to examine this surprising issue. Finally, we note that these methods might be diagnostic in other issues, such as steatosis or fatty liver disease. Protein changes are discerned in the next section.



**Figure 4.** IR difference spectra of three different pairs of K-means groups as shown with green and yellow in the inset images. These spectra are dominated by triglyceride as protein has been subtracted away. The groups are overlaid with the greyscale image of the metric  $LI$  ( $A_{1744\text{cm}^{-1}}/A_{1548\text{cm}^{-1}}$ ) to show that groups in the transitional region (dark) are being subtracted from nontumor spectra with high triglyceride (light). The image of each pair of groups has been placed in a colored box which matches the trace of the corresponding difference spectrum.

#### 4. PROTEINS

The amide I ( $1656\text{ cm}^{-1}$ ) and II ( $1548\text{ cm}^{-1}$ ) bands involve protein backbone motions and are two of the strongest bands in the IR spectra of liver tissue. Each amino acid in the protein chain gives rise to an oscillator that shifts due to its local environment and the oscillators sum to broad bands with many inflections. The inflections of the band lineshapes for two specific groups are revealed by calculating 9-point 2<sup>nd</sup> derivatives,  $d_i$ , based on spectral points  $\{\tilde{\nu}_i, A_i\}$ , where  $\tilde{\nu}_i$  is the wavenumber and  $A_i$  is the absorbance, as

$$d_i = \frac{-9A_{i+4} + 128A_{i+3} - 1008A_{i+2} + 8064A_{i+1} - 14350A_i + 8064A_{i-1} - 1008A_{i-2} + 128A_{i-3} - 9A_{i-4}}{5049(\tilde{\nu}_2 - \tilde{\nu}_1)^2}. \quad (1)$$

The lineshapes and 2<sup>nd</sup> derivatives of the amide I, amide II, and C=O bands are shown for a K-means group in the nontumor region and one in the region transitioning to the tumor in Figure 5. Structure in the 2<sup>nd</sup> derivatives shows that the oscillators are not uniformly distributed throughout the band, instead they pile up at specific subbands. There has been much IR work on the secondary structure of proteins<sup>6-9</sup> showing that the oscillators for  $\alpha$ -helix occur at different wavenumbers than those for  $\beta$ -sheet. Since it was very difficult to perfectly subtract away protein peaks in the previous section, many protein changes are evident with the K-means groups of IR spectra. In order to characterize these changes, a process of simultaneously fitting the lineshapes and 2<sup>nd</sup> derivatives was developed whereby the error contributions of the lineshapes and the 2<sup>nd</sup> derivatives are constrained to be equal. The fitted lineshape is modeled as a sum of Lorentzian functions

$$f_i = \sum_{j=1}^{n_{peaks}} p_{3(j-1)+1} / \left[ 1 + \left( \frac{x_i - p_{3(j-1)+2}}{p_{3(j-1)+3}} \right)^2 \right], \quad (2)$$

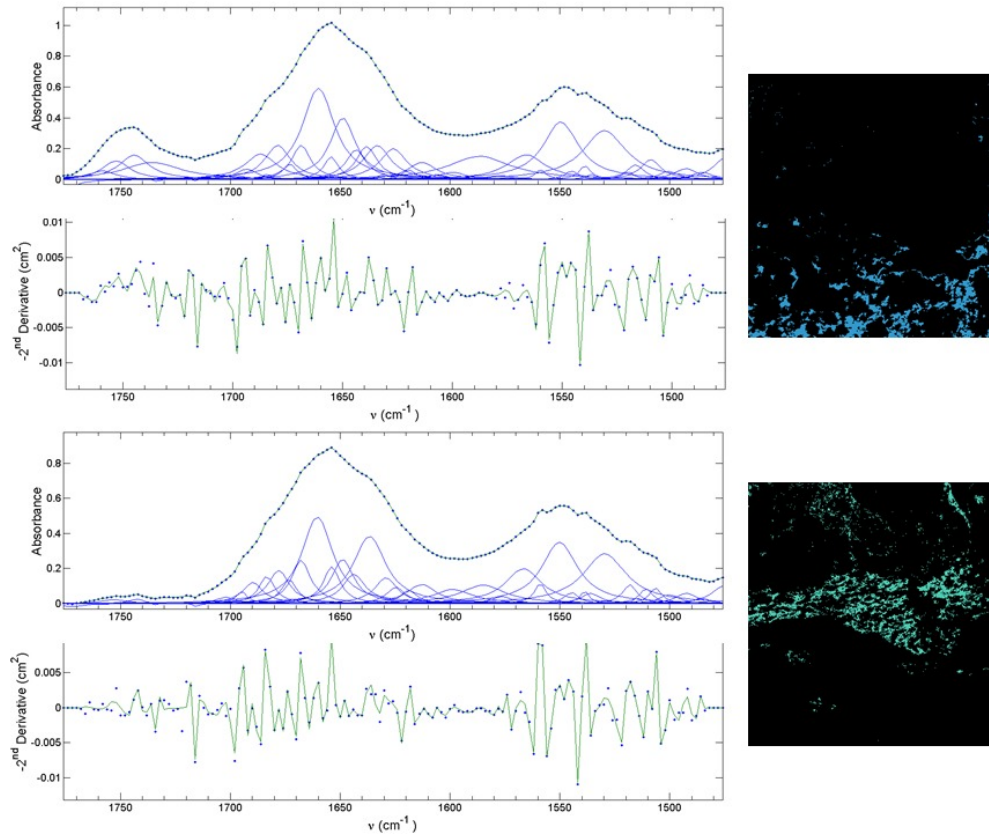
where  $j$  is an index over the Lorentzian peaks. There are three parameters ( $p$ ) for each peak which are organized in a one dimensional array,  $p_1, p_2, p_3, p_4, p_5, p_6, \dots, p_{n_{par}-2}, p_{n_{par}-1}, p_{n_{par}}$ , in order to use a standardized MATLAB optimization function (*fminsearch*). The parameters come in sets of three: the first of the set is the intensity (absorbance units), the second is the band position ( $\text{cm}^{-1}$ ), and the third is the half-width-at-half-max ( $\text{cm}^{-1}$ ). Given an initial set of parameters, one calculates the fit function ( $Af_i$ ) and then its 2<sup>nd</sup> derivative ( $d_{fi}$ ) [with the same eq. (1) formula, but using the fitted values  $Af_i$  rather than the measured values,  $A_i$ ]. So, for each value of  $\tilde{\nu}_i$ , we have four functions: the experimental spectrum ( $A_i$ ), the experimental 2<sup>nd</sup> derivative ( $d_i$ ), the fitted spectrum ( $Af_i$ ), and the fitted



2<sup>nd</sup> derivative ( $df_i$ ). The cost function ( $\sigma_{cost}$ ) for fitting both the lineshape and the 2<sup>nd</sup> derivative (scaled by  $|s|$ ) simultaneously is

$$\sigma_{cost} = \sqrt{\frac{\sum_i^n [A_i - A_{fi}]^2}{n - n_{par}}} + |s| \sqrt{\frac{\sum_i^n [d_i - df_i]^2}{n - n_{par}}} = \sigma_{lineshape} + |s| \sigma_{2nd\ derivative} \quad (3)$$

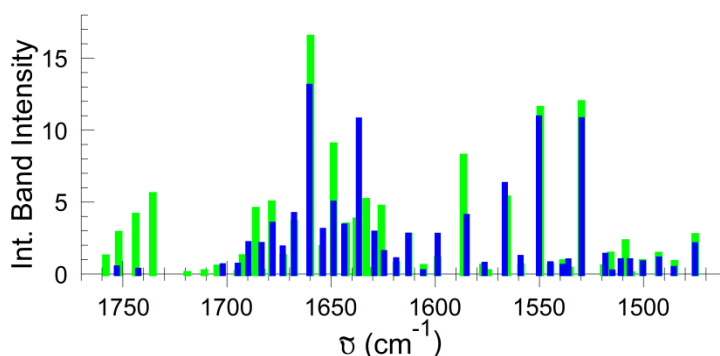
Our goal was to minimize the value of the cost function ( $\sigma_{cost}$ ) under the constraint of equal contributions from the lineshape and 2<sup>nd</sup> derivative. Initially, the proper value of  $s$  is unknown, so the process was started by choosing a value too big in magnitude, i.e. one that overestimates the importance of the 2<sup>nd</sup> derivative, while also locking-in



**Figure 5.** Simultaneous fits of the amide I, amide II, and C=O bands and their 2<sup>nd</sup> derivatives for two specific K-means groups that are imaged at the right. The top set corresponds to a nontumor group with  $s = -0.98 \text{ cm}^{-2}$ ,  $\sigma_{lineshape} = 0.00196$ , and  $\sigma_{2nd\ derivative} = 0.00201 \text{ cm}^2$ . The bottom set corresponds to a group in the transition region between nontumor and tumor with  $s = -1.26 \text{ cm}^{-2}$ ,  $\sigma_{lineshape} = 0.00214$ , and  $\sigma_{2nd\ derivative} = 0.00168 \text{ cm}^2$ .

the fit positions of the Lorentzian bands. Given an initial guess, a MATLAB program using the *fminsearch* function minimizes the cost function for the fixed value of  $s$  producing a new set of fit parameters,  $p$ . An ideal fit with equal contributions has  $\sigma_{lineshape} = |s| \sigma_{2nd\ derivative}$ , and therefore  $s_{ideal} = \sigma_{lineshape} / \sigma_{2nd\ derivative}$ . So, the ratio  $\sigma_{lineshape} / \sigma_{2nd\ derivative}$  is calculated at the end of an optimization and compared to the initially set value of  $s$ . If the optimization has helped at all, it will produce a ratio smaller in magnitude than the set value of  $s$ . In which case,  $s$  is set to the value of this ratio and the optimization program is run again using the new set of parameters,  $p$ , as a new and better initial guess. The process is iterated until the set value of  $s$  equals the resulting ratio  $\sigma_{lineshape} / \sigma_{2nd\ derivative}$ . The fitted lines are green in Figure 5. The fitting procedure remarkably finds ~17 subbands in the amide I range (1600-1700  $\text{cm}^{-1}$ ). Since the subbands are extensively overlapped, the integrated band intensities [ $\pi(\text{width})(\text{height})$ ] are plotted against subband position in Figure 6 for the two K-means groups considered in Figure 5. The nontumor group (green in Figure 6) has more ester-linked fat at ~1744  $\text{cm}^{-1}$  (subbands at 1735.6, 1743.8, 1752.0, and 1758.1  $\text{cm}^{-1}$ ). It has its most intense amide I subband at 1659.9  $\text{cm}^{-1}$  (often attributed to  $\alpha$ -helix) which is greatly reduced in

the transition to the tumor region. The 2<sup>nd</sup> largest amide I subband is at 1648.8 cm<sup>-1</sup> in the nontumor, but 1636.6 cm<sup>-1</sup> in the transition group. The nontumor amide II subband at 1586.4 cm<sup>-1</sup> is greatly reduced in intensity and shifted to 1585.0 cm<sup>-1</sup> in the transition group. In future work, we intend to find the subbands in the amide I region which are highly correlated with those in the amide II region enabling assignments to be extended into this region. Since there were no constraints to subband positions, it is notable how similar the subbands are in the two different groups. The possibility of addressing protein changes is exciting, but much more work is needed to get a handle on subband assignments. There are many more subtle differences associated with many other protein motifs (including random coil, triple helix, and various turn regions) and motifs are likely to exhibit more than one subband each. However, there are discernible changes between proteins in and out of the tumor which is encouraging.



**Figure 6.** Bar plot of the integrated band intensity vs subband position for the two groups considered in Figure 5: green is from a nontumor region, while blue is for the transitional group in the region changing from nontumor to tumor.

## ACKNOWLEDGMENT

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

## REFERENCES

- [1] Chen, Z., Butke, R., Miller, B. *et al.*, "Infrared Metrics for Fixation-Free Liver Tumor Detection," *The Journal of Physical Chemistry B*, 117(41), 12442-12450 (2013).
- [2] Bhargava, R., "Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology," *Anal. Bioanal. Chem.*, 389(4), 1155-1169 (2007).
- [3] Fernandez, D. C., Bhargava, R., Hewitt, S. M. *et al.*, "Infrared spectroscopic imaging for histopathologic recognition," *Nat. Biotechnol.*, 23(4), 469-474 (2005).
- [4] Diem, M., Papamarkakis, K., Schubert, J. *et al.*, "The Infrared Spectral Signatures of Disease: Extracting the Distinguishing Spectral Features Between Normal and Diseased States," *Appl. Spectrosc.*, 63(11), 307A-318A (2009).
- [5] Prophet, E. B., [Laboratory Methods in Histotechnology] American Registry of Pathology, Washington, DC(1992).
- [6] Tatulian, S. A., [Structural Characterization of Membrane Proteins and Peptides by FTIR and ATR-FTIR Spectroscopy] Springer Science+Business Media, New York, 9 (2013).
- [7] Byler, D. M., and Susi, H., "Examination of the Secondary Structure of Proteins by Deconvolved FTIR Spectra," *Biopolymers*, 25, 469-487 (1986).
- [8] Dong, A., Huang, P., and Caughey, W. S., "Protein Secondary Structures in Water from Second-Derivative Amide I Infrared Spectra," *Biochemistry (Mosc.)*, 29, 3303-3308 (1990).
- [9] Barth, A., "Infrared spectroscopy of proteins," *Biochim. Biophys. Acta*, 1767, 1073-1101 (2007).