

Article

Infrared Metrics for Fixation-Free Liver Tumor Detection

Zhaomin Chen, Ryan Butke, Barrie Miller, Charles L Hitchcock, Heather Cecile Allen, Stephen P. Povoski, Edward W. Martin, and James Vernon Coe

J. Phys. Chem. B, Just Accepted Manuscript • DOI: 10.1021/jp4073087 • Publication Date (Web): 20 Sep 2013

Downloaded from <http://pubs.acs.org> on October 2, 2013

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications

High quality. High impact.

The Journal of Physical Chemistry B is published by the American Chemical Society.
1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society.
However, no copyright claim is made to original U.S. Government works, or works
produced by employees of any Commonwealth realm Crown government in the course
of their duties.

1
2
3
4
5
6
7 Infrared Metrics for Fixation-Free Liver Tumor
8
9
10 Detection
11
12
13
14
15

16 *Zhaomin Chen^a, Ryan Butke^b, Barrie Miller^b, Charles L. Hitchcock^b, Heather C. Allen^{a,b*},*
17 *Stephen P. Povoski^c, Edward W. Martin^c Jr., and James V. Coe^{a*}*
18
19

20
21 ^aThe Ohio State University, Department of Chemistry and Biochemistry, 100 West 18th Avenue,
22
23 Columbus OH 43210-1173
24
25
26

27
28 ^bThe Ohio State University, Department of Pathology, 4132 Graves Hall, 333 W. 10th Ave,
29
30 Columbus, OH 43210
31
32

33
34 ^cThe Ohio State University, Department of Surgery, Division of Surgical Oncology, 410 W 10th
35
36 Avenue, Columbus, OH 43210
37
38

39
40 KEYWORDS. liver tumors, infrared metrics, k-means cluster analysis, human liver lipid, FTIR
41
42 imaging.
43
44

45 ABSTRACT. Infrared (IR) spectroscopic imaging of human liver tissue slices has been used to
46 identify and characterize liver tumors. Liver tissue, containing a liver metastasis of breast origin
47 (mucinous carcinoma) was surgically removed from a consenting patient and frozen without
48 formalin fixation or dehydration procedures, so that lipids and water remain in the tissues. A set
49 of IR metrics (ratios of various IR peaks) was determined for tumors in fixation-free liver tissues.
50
51 K-means cluster analysis was used to tell tumor from nontumor. In this case, there was a large
52
53
54
55
56
57
58
59
60

1
2
3 reduction in lipid content on going from nontumor to tumor tissue and a well resolved IR
4 spectrum of nontumor liver lipid has been obtained and analyzed. These IR metrics may
5 someday guide work on IR spectroscopic diagnostics on live patients in the operating room.
6
7 This work also suggests utility for these methods beyond the identification of liver tumors,
8 perhaps in the study of liver lipids.
9
10
11
12
13

14 INTRODUCTION 15 16 17 18 19

20 It is possible that infrared (IR) spectroscopy could prove useful as a real-time intraoperative
21 diagnostic tool^{1,2}, without the need of histologic fixation or staining of examined tissues, since IR
22 spectra can be recorded rapidly and with high signal-to-noise ratio. In this regard, one can
23 envision utilizing a fiber optic, attenuated total reflection (IR-ATR) probe³ to evaluate
24 potentially diseased organs or tissues *in situ* during surgical procedures. Ideally, a surgeon, who
25 is attempting to resect a tumor, would like to know whether the entire tumor has been
26 successfully removed and whether the surgical resection margins are negative for residual tumor
27 before leaving the operating room. However, the ability to obtain comprehensive real-time
28 information regarding completeness of surgical resection is generally highly labor-intensive,
29 time-consuming, and is fraught by a high level of false-negative results. This is secondary to the
30 intrinsic nature in which the evaluation of surgical resection margin is conducted by the
31 pathologist, and is primarily limited by the fact that the pathologist only evaluates a minute
32 fraction of the entire resected circumferential boundary of the tumor from within the operative
33 field. The aim of the current work is to identify the most important and characteristic IR spectral
34 features associated with tumor-bearing tissues within the liver, with the ultimate goal to develop
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

a method for real-time differentiation of specific cell types associated with diseased liver tissue and normal liver tissue that can be used to assist in the surgical resection of liver tumors.

Liver tissue, containing a liver metastasis of breast origin (mucinous carcinoma), was surgically removed from a consenting patient at the time of a planned liver resection, and was frozen without formalin fixation or dehydration procedures (i.e. lipids and water remain in the tissues). Fresh liver tissue was snap frozen in liquid nitrogen and cryostat sections obtained from which imaging IR spectra were recorded. A similar slice was also treated with a hematoxylin and eosin (H&E) stain⁴ and processed as a virtual Aperio slide for comparison to the IR results. A method based on k-means cluster analysis⁵⁻⁷ was used to identify a set of IR metrics which are important for identifying liver tumors. Some changes in chemical composition between the tumor and its surrounding tissue are presented.

In 2012, approximately 1.6 million new cancer cases were expected to be diagnosed in the United States, and approximately 570,000 Americans were expected to die of cancer¹⁰. Among these 1.6 million new cancer cases, there were approximately 29,000 new cases of primary liver cancer. Yet, more astonishingly, the incidence of secondary liver tumors (i.e., liver metastases) in the United States is estimated to be as many as 20-times greater than the incidence of primary liver cancer. Such liver metastases most commonly originate from colorectal, breast, and lung cancers cases. From a global perspective, the ratio of mortality to incidence of liver cancer is roughly 0.93 (www.who.int/mediacentre/factsheets/fs297/en, <http://globocan.iarc.fr/factsheets/cancers/liver.asp>). Clearly, tumors of the liver are an important, global problem and the motivation for the current study.

There is an extensive history of IR bioimaging⁸⁻¹³, including research on cancer with tissues other than liver¹⁴⁻¹⁹ and some IR spectral work on human liver²⁰⁻²⁶ and mice liver tissues²⁷.

Particularly notable to us was the methodology of Fernandez et al²⁸ who recently reported a table of IR metrics for prostate cancer. These metrics serve as a starting point for our work on IR metrics of liver tissue. Several groups^{19,29,30} have shown that the functional equivalent of an H&E stained image could be extracted from IR spectroscopic imaging information on tissues slices. In our current work, we evaluate the utility of the prostate tumor IR metrics of Bhargava and coworkers²⁸ for liver and augment these with metrics important when the tissue is not fixed. All of the potential metrics were merged into a large set and a quantitative determination revealed the most important for evaluation of tumors within the liver.

The paper continues with an Experimental section giving a description of the collection of human liver tissue, the recording of IR spectroscopic imaging data, and software that extracts the IR spectra from the commercial software enabling manipulation with statistical programs. The next section introduces the types of IR metrics under consideration, which is followed by a k-means cluster analysis using scaled IR metrics in order to identify the best IR metrics. Finally, the results of k-means cluster analysis with the unscaled subset of best IR metrics are described using both 5 and 25 groups. The paper concludes with a discussion of the k-means cluster results for liver diagnostics.

EXPERIMENTAL

The liver tissue was collected during surgery at the University Hospital (Ohio State University, Columbus, OH). The collection, reservation, and resecting of the tissue was approved by the Institutional Review Board (IRB, No. 2011C0085). Immediately after collection, the specimen was pretreated by snap freezing in liquid nitrogen, which allowed rapid lowering of the sample to a temperature below -70 °C. Snap freezing of the liver tissue provided specimen integrity for further analysis. A cryostat, set at -20 °C, was used to obtain 2-3 μm thick sections

of tissue. Normally, a pathologist will fix the tissue, i.e. soak the tissue in a neutral buffered formalin solution, and then dehydrate the sample with a sequence of graded ethanols, xylene, and finally parrafin. This preserves the tissue and works well with various staining procedures.

However, research presented here concerns the possibility of taking *in vivo* IR spectra, so the tissue was not fixed. It was sliced to a thickness of 2-3 μm and the slice was further cropped to fit onto an IR ZnSe window (8.0 mm diam., 1.0 mm thickness, from Crystran LTD. UK) within a home-designed, sealed sample holder. An optical microscope image of the sample is shown in Figure 1a where the intense yellow of the ZnSe window has been digitally reduced.

IR spectra were recorded with an imaging FTIR microscope (Perkin Elmer Spotlight 300) with a computer-controlled microscope stage and a liquid-nitrogen-cooled linear array of 16 mercury cadmium telluride (MCT) detectors. The data was recorded in seven windows as shown with dotted lines in Figure 1a. Since the imaging region is bigger than the microscope's field of view, the instrument patches together multiple optical views producing a rectangular tiling artifact which is apparently accentuated by the bright yellow color of ZnSe. In spite of the tiling artifact, this image can be overlaid, in a pixel-by-pixel fashion, with the k-means images of following sections. Each IR window region (shown in Figure 1a) was 2,200.0 μm vertically by 300.0 μm horizontally (352 pixels by 48 pixels). The instrument obtained a full IR spectrum at each image pixel (6.25 μm by 6.25 μm area) in each window (4 cm^{-1} resolution, 2 cm^{-1} steps in the range of $750\text{-}4000 \text{ cm}^{-1}$, and 16 scans per pixel). Each of the seven windows required about 3 hr of scanning time for a total of 21 hr of data collection.

Later, the data were merged with home-written Matlab (version R2013a) routines into one large window which was 2,200.0 μm by 2,100.0 μm including 118,272 separate pixels and/or IR spectra. All of the following tissue slice images are of this size. A Matlab routine from Perkin

1
2
3 Elmer called “fsm_load.m” by Ben Peterson facilitated the process
4
5

6 (<http://www.mathworks.it/matlabcentral/fileexchange/22736-perkinelmer-ir-data-file-import-tools>). Our home-written Matlab programs are described in the MS thesis of Z. Chen³¹. The
7 starting point for the analysis is a three dimensional matrix of spectra transmittances, data(i,j,k'),
8 where i is a pixel index for the image row, j is a pixel index for the image column, and k' is an
9 index stepping through the IR spectrum. After the IR spectroscopic imaging, the sample was
10 taken from the instrument and treated with an H&E stain⁴ (as shown in Figure 1b). The intense
11 yellow of the ZnSe window has been digitally reduced in the same manner as with Figure 1a.
12
13 The staining shows a tumor at the top half of the merged window, and so the bottom half was a
14 region just outside of the tumor.

15
16
17
18
19
20
21
22
23
24
25
26
27
28 **IR METRICS**
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

50 Considering that there are 1626 spectral steps (potential metrics from peak intensities) in each
51 IR spectrum and 118,272 spectra, there are more than 192 million pieces of information to be
52 analyzed. Clearly, there is utility in reducing the spectral information to something more
53 manageable and understandable, hence the utility of IR metrics. We have started with the IR
54 metrics (*b1-b36* in Table 1s of the supplementary materials) of Fernandez et al²⁸ (for prostate
55 cancer on fixed tissues) and added a set that might be more useful without fixation (*b37-b64* in
56 Table 1s of the supplementary materials). Matlab routines facilitate the calculation of a value of
57 each metric at each pixel in the image. The result is a three dimensional matrix of IR metrics,
58 $b(i,j,m)$, where i,j are indices over the rows and columns respectively of the image, and m is an
59 index over the metrics. The absorbance intensities at each peak were corrected with a two point
60 baseline specific to each IR peak eliminating the need for baseline offsetting and flattening

before analysis. The metrics reduce the original data, a spectrum of 1626 points at each image pixel, to 64 values at each pixel. The process of k-means clustering was used to evaluate the quality of the metrics leading to a reduced set of IR metrics that are highly appropriate for liver tissue work on fixation-free tissue with lipid.

K-MEANS CLUSTER ANALYSIS WITH SCALED METRICS

Variations apparent in the IR spectra at each image pixel encouraged an attempt to sort the pixels into groups using IR metrics. K-means clustering is one of the simplest clustering techniques, whereby clusters or groups are characterized by centroids (essentially average values of the members of the group). The technique varies the position of the centroids and the membership in the groups to minimize the sum of “distances” between group members and the group centroid. Unlike the normal three dimensional distances of Euclidean space, these distances are hyperdimensional having a component contribution from each of the metrics (64 in this part of the evaluation).

In order to use the metric results with Matlab routines for k-means clustering, the values of each metric, m , in the two dimensional plane of the image had to be mapped into a single column. Using Matlab’s linear mapping convention, the value of each biomarker at each pixel in the two-dimensional image was mapped into a one-dimensional array ($i,j \rightarrow n$), where n is a one dimensional index of the pixels [$n = (j - 1)i_{max} + i$]. In other words, the three-dimensional metric data, $b(i,j,m)$, was mapped into a two dimensional matrix, $X(n,m)$, where n is a single index over the pixels and m is an index over the metrics. Upon defining an index k over the groups, G_k , defined by k-means clustering, the centroids for each group can be denoted as $c(k,m)$. Then k-means clustering minimizes the quantity

$$\sum_k (\sum_{if\ n \in G_k} \sqrt{\sum_m [X(n,m) - c(k,m)]^2}), \quad (1)$$

by changing group membership and centroid position. The quantity $\sqrt{\sum_m [X(n, m) - c(k, m)]^2}$ is the hyperdimensional distance between each metric score, $X(n, m)$, for an image pixel in a group and its centroid, $c(k, m)$. This leads to a criterion for evaluating the metrics with regard to distinguishing groups. Noting that the distance between two groups, G_{k1} and G_{k2} , is

$$d_{G_{k1}, G_{k2}} = \sqrt{\sum_m [c(k1, m) - c(k2, m)]^2}, \quad (2)$$

the contribution of a metric to the separation between groups is

$$d_{G_{k1}, G_{k2}, m} = |c(k1, m) - c(k2, m)|. \quad (3)$$

So, the metrics can be evaluated by the size of $d_{G_{k1}, G_{k2}, m}$ which determines the importance of each metric to the separation of those groups. However, the metrics have various average values and different distributions which can bias a numerical comparison, so they were all scaled to the same average value (0.5) and the estimated standard deviation of each metric was used to set a range, i.e. the scaled value of zero was two standard deviations below the scaled average and the scaled value of one was set to two standard deviations above the average.

K-means cluster analysis was performed with 5 groups and scaled metrics (an image, Figure 1s, is given in the supplementary materials). There was one nontumor group at the bottom half of the image, three tumor groups, and a group of holes (some holes in the tissue slices were inadvertently created through the microtoming process). Matlab programs produce a black-and-white bitmap image of each metric, a color bitmap image of each of the groups, the centroid component of each metric for each group, the average x and y position of each group, and the IR spectrum of each group. Greyscale images of selected metrics (scaled) are given in Figure 2 showing that different IR metrics are sensitive to different aspects of the liver tissue. The contribution of each scaled metric to the difference between groups [equation (3)] was calculated for the nontumor group and two of the tumor groups. These contributions were added for each

of the 64 initial metrics and sorted from highest to lowest. The top twenty are given in Table 1 in order of importance.

Table 1. Top twenty IR metrics for fixation-free liver tumor tissues. Each metric is a baseline-corrected ratio of absorbance at the numerator (in cm^{-1}) to absorbance at denominator (in cm^{-1}). If a range is given, then the average absorbance over that range is used. The “*b*” labels are from the set of 64, while the “*L*” labels are for the new set of 20.

Name	Ratio	Name	Ratio
<i>L1(b43)</i>	1744/1548	<i>L11(b19)</i>	1016/1080
<i>L2(b37)</i>	1744/1244	<i>L12(b59)</i>	1252/1544
<i>L3(b64)</i>	1742/1256	<i>L13(b39)</i>	1024/1080
<i>L4(b53)</i>	1160/1548	<i>L14(b42)</i>	1080/1244
<i>L5(b28)</i>	1516/1236	<i>L15(b47)</i>	1516/1582
<i>L6(b54)</i>	2916/1548	<i>L16(b7)</i>	1744/1162
<i>L7(b45)</i>	1120/1020	<i>L17(b23)</i>	1080/3290
<i>L8(b46)</i>	2924/1544	<i>L18(b58)</i>	1556/1548
<i>L9(b48)</i>	1080/1548	<i>L19(b2)</i>	1012/1256
<i>L10(b38)</i>	1744/1162	<i>L20(b30)</i>	(1144-1182)/1544

RESULTS WITH UNSCALED IR METRICS

Using the top 20 metrics in Table 1 without scaling, a k-means cluster analysis was performed using 5 groups producing the image shown in Figure 3. The nontumor is green and the holes are yellow. Again there are three groups within the tumor (red, blue, and cyan). The IR spectrum of each group (baseline-flattened and offset to zero) is presented in the bottom part of Figure 3 with the same color coding for groups. The IR spectra are normalized by the number of pixels in each group, so the intensities are meaningfully compared. There are significant differences between the tumor and nontumor regions. Most notably, the peak at 1745 cm^{-1} which is assigned to ester-linked fats, is much smaller in this tumor.

Table 2. IR metric values for 5 groups from the k-means analysis with the best unscaled metrics.

The associated IR bands (in cm^{-1}) are given as a ratio in parentheses for each metric. If a range is given, then the metric involves an average over that range.

IR Metric	Nontumor (green)	Tumor 1 (red)	Tumor 2 (blue)	Tumor 3 (cyan)
<i>L1</i> (1744/1548)	0.679	0.132	0.288	0.092
<i>L2</i> (1744/1244)	0.986	0.375	0.616	0.216
<i>L3</i> (1742/1256)	0.904	0.425	0.643	0.258
<i>L4</i> (1160/1548)	0.353	0.107	0.155	0.167
<i>L5</i> (1516/1236)	1.025	1.741	1.416	1.505
<i>L6</i> (2916/1548)	0.625	0.284	0.373	0.322
<i>L7</i> (1120/1020)	0.755	1.816	1.088	1.231
<i>L8</i> (2924/1544)	0.786	0.357	0.471	0.391
<i>L9</i> (1080/1548)	0.383	0.166	0.223	0.204
<i>L10</i> (1744/1162)	1.951	1.166	1.858	0.566
<i>L11</i> (1016/1080)	0.892	0.255	0.462	0.488
<i>L12</i> (1252/1544)	0.771	0.355	0.485	0.412
<i>L13</i> (1024/1080)	0.962	0.440	0.588	0.631
<i>L14</i> (1080/1244)	0.557	0.450	0.471	0.466
<i>L15</i> (1516/1582)	1.277	1.120	1.165	1.241
<i>L16</i> (1744/1162)	0.372	0.110	0.163	0.172
<i>L17</i> (1080/3290)	0.359	0.179	0.231	0.235
<i>L18</i> (1556/1548)	0.832	0.896	0.878	0.907
<i>L19</i> (1012/1256)	0.433	0.110	0.216	0.251
<i>L20</i> [(1144-1182)/1544]	0.325	0.104	0.150	0.166
<i>pixels</i>	36,476	23,753	21,075	33,143

The metric scores for all groups except the holes are presented in Table 2 enabling comparisons between metrics and to other IR work. Many specific, chemical changes are archived in Table 2. As only one example, take the first new metric, *L1*, which is the ratio of the absorbance of ester-linked fat at 1744 cm^{-1} to that of the amide II band of protein at 1548 cm^{-1} . The value is 0.679 in the nontumor, but it falls to 0.288 in the blue tumor region which is closest to the nontumor part, to 0.132 in the red tumor group, and 0.092 in the cyan tumor group. There is a dramatic reduction of ester-linked fat from the nontumor upon going into the tumor.

The amide I (1656 cm^{-1}) and II (1548 cm^{-1}) protein bands are two of the strongest bands involving protein backbone motions which have many inflections indicating unresolved structure. There has been much IR work on the secondary structure of proteins³²⁻³⁵ and, given

the high signal-to-noise of the group spectra (they are averages of ~30,000 individual spectra),
2nd derivatives of the group IR spectra were determined for comparison upon going from tumor
to nontumor. The Perkin Elmer Spectrum program was used to get 5-point finite-difference, 2nd
derivatives that have been multiplied by a factor of -200 in order to be displayed on a scale
similar to the absorption spectra as shown in Figure 4. Second derivatives have large negative-
going features at the peak center which has been reversed and scaled for better comparison to the
absorption lineshape. Notice how the inflections in the absorption spectra correspond to the
peaks of the 2nd derivatives multiplied by -200. There are 7 or 8 discrete peak centers discernible
in the amide I band alone. Their peak centers are listed in Figure 4. The 2nd derivative peaks at
 1654 cm^{-1} are α -helix, while the peaks at 1637 and 1694 cm^{-1} (or possibly 1683 cm^{-1}) may be β -
sheet. The other peaks likely correspond to other helices and various turns in standard structures.
There is less α -helix in the tumor and more of the 1683 cm^{-1} amide I group in the tumor. The
biggest changes seem to be in the amide II band. The 2nd derivative peak at 1546.9 cm^{-1} is
largest in the nontumor region, while the two surrounding peaks at 1559.3 and 1537.9 cm^{-1}
become larger in the tumor. Clearly, there are discernible changes between proteins in and out of
the tumor.

It has been useful to make histograms of the IR metric values. In fact, all but one of the
metrics (L14) show two resolved distributions – one for the tumor and one for the nontumor. It
is even more useful to make two-dimensional histograms of one metric against another in order
to see if they work together to give even more separation of the tumor and nontumor groups.
Such a plot of L5 (baseline corrected absorption at 1516 cm^{-1} divided by that at 1236 cm^{-1}) vs L1
(baseline corrected absorption at 1744 cm^{-1} divided by that at 1548 cm^{-1}) is shown in Figure 5a.
Clearly, there are two separate distributions and the separation is greater than either metric by

1
2
3 itself (as is evident by the projections which reveal the one-dimensional histograms of each
4 metric). A plot of a different set of metrics, L14 vs L1, is presented in Figure 5b. A third
5 distribution is apparent, but there are no groups that identify this region.
6
7
8
9

10 One can address this issue by doing the k-means analysis with more groups, but this presents
11 challenges to graphical visualization. A k-means analysis was performed with 25 groups and the
12 results are given in Figure 6. The IR spectra of all 25 groups have been given in the
13 supplementary material. A matrix of the equation (2) distances between each of the 25 groups
14 was computed and an hierarchical dendrogram was constructed which supported classification of
15 the groups into categories of nontumor (green), tumor (purple), and holes (black or dark blue).
16 The images were plotted with shades of green for the nontumor, shades of purple for the tumor,
17 shades of dark blue for the holes, and yellow for the nontumor group closest to the tumor. This
18 exercise emphasizes that it is relatively easy to discern tumor from nontumor. Therefore, some
19 of the most important work ahead lies in trying to extract differences between the groups within
20 a classification.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

36 Noting that groups 3 and 13 are near each other and are neighbors in the dendrogram, the IR
37 spectrum of group 13 multiplied by 0.963 was subtracted from the IR spectrum of group 3 in
38 order to cancel the protein (by the amide I band) as best as possible. This leaves a remarkably
39 well resolved IR spectrum of the lipid from the nontumor region as shown in Figure 7. The
40 nontumor lipid peaks include 3002(shoulder), 2959(shoulder), 2925, 2854, 2249, 1746, 1469,
41 1443, 1369, 1255, 1157, 1107, 1083, 928, and 859 cm^{-1} . The regions from 1500-1700 cm^{-1} and
42 above 3100 cm^{-1} are dominated by changes in protein and are avoided for this discussion about
43 lipid. Also shown (in red) is a modeled spectrum using FDM's Very Large Bundle
44 (www.fdmsspectra.com) of IR libraries based on a weighted sum of 0.6 times glyceryl
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 triacetylricinoleate and 0.4 times ethyl cyanopolyacrylate (where the max peak in each was
4 scaled to one). A Euclidean IR search was performed on the nontumor, ester-linked lipid
5 difference spectrum (black trace in Figure 7) using the Perkin Elmer Spectrum Search program
6 and FDM's Very Large Bundle of IR libraries. The two top matches were glyceryl
7 triacetylricinoleate and ethyl cyanopolyacrylate. Glyceryl triacetylricinoleate is commonly used
8 in cosmetics. It is a triglyceride with C18:1 chains (18 carbons, one double bond at C9), each
9 containing one double bond and an extra ester linkage per fat chain (occurring below the double
10 bond towards the free end of the fat chain at C12), which is somewhat different than textbook
11 triglycerides. This species is an excellent match except for the peaks at 858 and 2248 cm⁻¹, for
12 which cyanopolyacrylate (super glue) does better. The super glue polymer does not match the fat
13 chains at all, but it is a good match in the fingerprint region and it does get the peaks at 858 and
14 2248 cm⁻¹. Taken together, this suggests that the average of the nontumor liver lipids (after
15 freezing, slicing, melting, IR exposure, and hours of time) is similar to a glyceryl triacetyl
16 ricinoleate, where polyacrylate polymerization has been occurring perhaps between different fat
17 chains. The average nontumor liver lipid is a triglyceride with roughly twice the ester groups of
18 textbook triglycerides, a prominent backbone vibration at 858 cm⁻¹ (not due to fat chains), and
19 some C≡N triple bond groups (less than one per fat chain). Note that the triple bond observation
20 at 2248 cm⁻¹ could also be a C≡C, and the only evidence at this point for C≡N is the match to
21 cyanopolyacrylate. Explanations for triple bonds in nontumor, human liver lipid need to
22 consider that the tissue is not fixed. Possible explanations include: 1) enzymes that destroy
23 triple bonds are no longer functional, 2) triple bonds exist as intermediates in normal lipid
24 degradation, 3) drug treatments of patients with tumors may damage the liver, and 4) triple bonds
25 are a chemical artifact of mixing cell parts as tissue warms after slicing. There are, after all,
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

good reasons why pathologists fix tissues extracting lipids. Further work is necessary to say more. In spite of this, there seems to be potential for analysis of liver lipids by these methods although caution must be exercised in extrapolating to the lipid conditions in the living liver. We believe that this merits further study.

With 25 k-means groups, it was informative to re-examine the L14 vs L1 metric-metric plot in order identify the new distribution from Figure 5b which was not identified by any groups of the 5 group analysis. The L14 metric is the ratio of absorbance at 1080 cm^{-1} to 1244 cm^{-1} (glycogen and/or lipid to phosphate) and the L1 is that at 1744 cm^{-1} to 1548 cm^{-1} (ester-linked lipid to protein). A plot of L14 vs L1 is given in Figure 8b with white numbers indicating the centroid values for these metrics from the 25 groups. Clearly, groups 7 and 23 (with L1, L14 scores of 0.107, 0.868 and 0.197, 0.845, respectively), now identify the third distribution and are imaged with dark purple colors in Figure 8a. These are groups further within the tumor. The lighter purple regions correspond to groups 24 and 18 (with L1, L14 scores of 0.071, 0.413 and 0.088, 0.367, respectively). By comparison to the standard H&E stains (Figure 1a inset), we believe there is a concentration of fibrous stroma cells (high protein, low lipid) in these light purple, tumor regions (see Figure 8a) which have H&E staining more like the nontumor region. In Figure 8b, it is also interesting that groups 5 and 13 occur in the region between the nontumor and tumor distributions. These groups have been colored light green (5) and yellow (13) as isolated in Figure 8a. We assert that these groups help to define the tumor's margin.

The IR spectra of selected groups are shown in Figure 9. Each group is normalized to the number of image pixels in the group, so intensities are meaningful. There are significant differences in all of the selected groups which have been identified with a color coded image inset in Figure 9. As mentioned above, there are many more subtle differences in the IR spectra

of these various groups. Therefore the IR spectra of all 25 groups (and identifying bitmaps of each group) have been provided in the supplemental materials.

DISCUSSION

A set of literature IR metrics were augmented by various lipid bands and evaluated giving a set of twenty that are good for detecting fixation-free liver tumors (see Table 1). A scaling of the metrics was introduced to make a level comparison. Only 6 out of 20 were from the previous literature work (on fixed prostate tissue) showing that the molecules removed by fixation are quite important. Ester-linked lipid provides for good tumor distinction, at least in this patient. The top IR metric, L1, is the ratio of absorbance at 1744 cm^{-1} to absorbance at 1548 cm^{-1} , i.e. a ratio of ester-linked lipid to protein. It was used as the ordinate in several IR metric vs IR metric plots (Figures 5 and 8) which show good distinction between tumor and nontumor. While this work concentrated on distinguishing tumor in one patient, work on differentiating cell types by the same methods is promising and continues. Work also continues on gathering data from other patients. It is interesting that the ester-linked lipids vary greatly between patients. The lipid-related peaks at 2248 cm^{-1} (likely a triple bond) and 858 cm^{-1} are very different than what is seen with IR spectra of fixed liver tissues. So far, these peaks have been seen in two other patients. Future work on more patients will provide a better perspective on the importance of this result.

Diagnostic tools have been presented that could one day be utilized within the operating room, providing vital real-time information on how the surgeon should proceed and for verifying completeness of surgical resection. Noting that fixation is a large perturbation from the tissue of living patients, the freezing and slicing of the current work is a lesser perturbation and a step closer to the ultimate goal of *in situ* IR spectral evaluation of the liver or on freshly excised liver

tissue from biopsy or resectional specimens. The sensitivity of IR spectra to liver lipids suggest utility beyond work on tumors of the liver. Steatosis or fatty liver disease may be readily assessed with the methods presented herein. Knowledge about the condition of liver might be important in a variety of health care decisions beyond liver tumors.

ASSOCIATED CONTENT

Supporting information includes a supplementary word file with a table of the original 64 IR metrics, an image of a 5 group k-means analysis using scaled IR metrics, and bitmap images of the groups from a 25 group k-means cluster analysis. Also included are the IR spectra of the groups from a 25 group, k-means analysis in an Excel file.

AUTHOR INFORMATION

Corresponding Authors

*James V. Coe, coe.1@osu.edu

*Heather C. Allen, allen.697@osu.edu

ACKNOWLEDGMENT

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

REFERENCES

- (1) Wu, J. G.; Xu, Y. Z.; Sun, C. W.; Soloway, R. D.; Xu, D. F.; Wu, Q. G.; Sun, K. H.; Weng, S. F.; Xu, G. X. Distinguishing malignant from normal oral tissues using FTIR fiber-optic techniques. *Biopolymers* **2001**, *62*, 185-192.

(2) Diem, M.; Miljkovic, M.; Bird, B.; Chernenko, T.; Schubert, J.; Marcsisin, E.; Mazur, A.; Kingston, E.; Zuser, E.; Papamarkakis, K.; Laver, N. Applications of Infrared and Raman Microspectroscopy of Cells and Tissue in Medical Diagnostics: Present Status and Future Promises. *Spectroscopy-an International Journal* **2012**, *27*, 463-496.

(3) Harrington, J. A. *Infrared Fibers and Their Applications*; SPIE Press: Bellingham, Washington USA, 2004.

(4) Prophet, E. B. *Laboratory Methods in Histotechnology*; American Registry of Pathology: Washington, DC, 1992.

(5) Lloyd, S. P. Least squares quantization in PCM. *IEEE Transactions on Information Theory* **1982**, *28*, 129–137.

(6) Kanungo, T.; Netanyahu, N. S.; Wu, A. Y. An Efficient k-Means Clustering Algorithm: Analysis and Implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2002**, *24*, 881-892.

(7) Seber, G. A. F. *Multivariate Observations*; John Wiley & Sons: Hoboken NJ, 1984.

(8) Coates, V. J.; Offner, A.; Siegler, E. H. Design and Performance of an Infrared Microscope Attachment. *Journal of the Optical Society of America* **1953**, *43*, 984-989.

(9) Petibois, C.; Deleris, G. Chemical mapping of tumor progression by FT-IR imaging: towards molecular histopathology. *Trends in Biotechnology* **2006**, *24*, 455-462.

- (10) Kohler, A.; Bertrand, D.; Martens, H.; Hannesson, K.; Kirschner, C.; Ofstad, R. Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information. *Analytical and Bioanalytical Chemistry* **2007**, 389, 1143-1153.
- (11) Noreen, R.; Moenner, M.; Hwu, Y.; Petibois, C. FTIR spectro-imaging of collagens for characterization and grading of gliomas. *Biotechnology Advances* **2012**, 30, 1432-1446.
- (12) Yang, T. T.; Weng, S. F.; Zheng, N.; Pan, Q. H.; Cao, H. L.; Liu, L.; Zhang, H. D.; Mu, D. W. Histopathology mapping of biochemical changes in myocardial infarction by Fourier transform infrared spectral imaging. *Forensic Science International* **2011**, 207, E34-E39.
- (13) Diem, M.; Papamarkakis, K.; Schubert, J.; Bird, B.; Romeo, M. J.; Miljkovic, M. The Infrared Spectral Signatures of Disease: Extracting the Distinguishing Spectral Features Between Normal and Diseased States. *Applied Spectroscopy* **2009**, 63, 307A-318A.
- (14) Maziak, D. E.; Do, M. T.; Shamji, F. M.; Sundaresan, S. R.; Perkins, D. G.; Wong, P. T. T. Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: An exploratory study. *Cancer Detection and Prevention* **2007**, 31, 244-253.
- (15) Bird, B.; Miljkovic, M.; Remiszewski, S.; Akalin, A.; Kon, M.; Diem, M. Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer. *Laboratory Investigation* **2012**, 92, 1358-1373.
- (16) Schubert, J. M.; Bird, B.; Papamarkakis, K.; Miljkovic, M.; Bedrossian, K.; Laver, N.; Diem, M. Spectral cytopathology of cervical samples: detecting cellular abnormalities in cytologically normal cells. *Laboratory Investigation* **2010**, 90, 1068-1077.

- 1
2
3 (17) Wood, B. R.; Chiriboga, L.; Yee, H.; Quinn, M. A.; McNaughton, D.; Diem, M.
4 Fourier transform infrared (FTIR) spectral mapping of the cervical transformation zone, and
5 dysplastic squamous epithelium. *Gynecologic Oncology* **2004**, *93*, 59-68.
6
7 (18) Romeo, M.; Matthäus, C.; Miljkovic, M.; Diem, M. Infrared microspectroscopy
8 of individual human cervical cancer (HeLa) cells. *Biopolymers* **2004**, *74*, 168-171.
9
10 (19) Pallua, J. D.; Pezzei, C.; Zelger, B.; Schaefer, G.; Bittner, L. K.; Huck-Pezzei, V.
11 A.; Schoenbichler, S. A.; Hahn, H.; Kloss-Brandstaetter, A.; Kloss, F.; Bonn, G. K.; Huck, C. W.
12 Fourier transform infrared imaging analysis in discrimination studies of squamous cell
13 carcinoma. *Analyst* **2012**, *137*, 3965-3974.
14
15 (20) Murakata, L. A.; Lewin-Smith, M. R.; Specht, C. S.; Kalasinsky, V. F.; McEvoy,
16 P. L.; Vinh, T. N.; Rabin, L. N.; Mullick, F. G. Characterization of acrylic polyamide plastic
17 embolization particles in vitro and in human tissue sections by light microscopy, infrared
18 microspectroscopy and scanning electron microscopy with energy dispersive X-ray analysis.
19 *Modern Pathology* **2006**, *19*, 922-930.
20
21 (21) Hara, A.; Taketomi, T. Isolation and Determination of Cholesterol Glucuronide
22 in Human Liver. *Lipids* **1982**, *17*, 515-518.
23
24 (22) Diem, M.; Chiriboga, L.; Lee, H. Infrared spectroscopy of human cells and tissue.
25 VIII. Strategies for analysis of infrared tissue mapping data and applications to liver tissue
26 *Biopolymers* **2000**, *57*, 282-290.
27
28 (23) Chiriboga, L.; Yee, M.; Diem, M. Infrared spectroscopy of human cells and
29 tissue. Part VI: A comparative study of histopathology and infrared microspectroscopy of
30 normal, cirrhotic, and cancerous liver tissue *Applied Spectroscopy* **2000**, *54*, 1-8.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- (24) Chiriboga, L.; Lee, H.; Diem, M. Infrared spectroscopy of human cells and tissue. Part VII: FT-IR microspectroscopy of DNase- and RNase-treated normal, cirrhotic, and neoplastic liver tissue. *Applied Spectroscopy* **2000**, *54*, 480-485.
- (25) Susie, B.-W.; Tamara, G.; Sandra, H.; Jill, B.; Max, D. Infrared spectroscopy of human tissue. V. Infrared spectroscopic studies of myeloid leukemia (ML-1) cells at different phases of the cell cycle. *Biospectroscopy* **1999**, *5*, 219-227.
- (26) diem, M.; Boydston-White, S.; Chiriboga, L. Infrared spectroscopy of cells and tissues: Shining light onto a novel subject *Applied Spectroscopy* **1999**, *53*, 148A-161A.
- (27) Wong, P. T. T.; Cadrin, M.; French, S. W. Distinctive Infrared Spectral Features in Liver-Tumor Tissue of Mice - Evidence of Structural Modifications at the Molecular Level. *Experimental and Molecular Pathology* **1991**, *55*, 269-284.
- (28) Fernandez, D. C.; Bhargava, R.; Hewitt, S. M.; Levin, I. W. Infrared spectroscopic imaging for histopathologic recognition. *Nat Biotech* **2005**, *23*, 469.
- (29) Walsh, M. J.; Reddy, R. K.; Bhargava, R. Label-Free Biomedical Imaging With Mid-IR Spectroscopy. *IEEE Journal of Selected Topics in Quantum Electronics* **2012**, *18*, 1502-1513.
- (30) Fernandez, D. C.; Bhargava, R.; Hewitt, S. M.; Levin, I. W. Infrared spectroscopic imaging for histopathologic recognition. *Nat. Biotechnol.* **2005**, *23*, 469-474.
- (31) Chen, Z. Imaging Infrared Microscope Analysis of Fixation-Free Liver Tumor Tissue, MS, The Ohio State University, 2013.
- (32) Tatulian, S. A. Structural Characterization of Membrane Proteins and Peptides by FTIR and ATR-FTIR Spectroscopy. In *Lipid-Protein Interactions: Methods and Protocol*; Kleinschmidt, J. H., Ed.; Springer Science+Business Media: New York, 2013; Vol. 974.

- 1
2
3 (33) Byler, D. M.; Susi, H. Examination of the Secondary Structure of Proteins by
4
5 Deconvolved FTIR Spectra. *Biopolymers* **1986**, *25*, 469-487.
6
7 (34) Dong, A.; Huang, P.; Caughey, W. S. Protein Secondary Structures in Water from
8 Second-Derivative Amide I Infrared Spectra. *Biochemistry* **1990**, *29*, 3303-3308.
9
10 (35) Barth, A. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta* **2007**,
11
12 1767, 1073-1101.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1. Optical microscope images of a slice of liver tissue with a tumor on ZnSe for IR spectroscopic imaging (scale on left). (a) Seven windows (top) were merged into one window (2.1000 mm wide by 2.2000 mm tall) for IR spectroscopic imaging. An Aperio H&E stain for an optical microscope of a nearby slice at approximately the same place and size is inset at right. The bottom (b) shows the same tissue as in (a) after H&E staining on ZnSe. The tumor is lighter and at the top half of the images.

Figure 2. Greyscale bitmap images of selected scaled metrics – L1, L7, L14, and L18 from top left to bottom right. The tumor is largely in the top half of the images.

Figure 3. K-means cluster analysis with 5 groups using the *unscaled*, top-twenty metrics (top). The nontumor portion is in green and the holes are yellow. The tumor has red, blue, and cyan groups. The IR spectrum of each group is given at the bottom with the same color coding.

Figure 4. Second derivatives of the IR spectra for each group in Figure 3 (excluding holes) with the same color coding. The dotted traces are absorption spectra, while the solid traces are 2nd derivatives multiplied by -200 in order to put the derivatives on a scale similar to the absorption spectra. Bottom (green) is nontumor and the top has the three tumor groups.

Figure 5. a) Two-dimensional histogram (top) of IR metrics L5 vs L1. L1 is a ratio of ester-linked fat to protein, while L5 is the side of the amide II protein band to phosphate. A contour diagram of the same information is given at the bottom. The centroid values for the k-means groups are given with cross symbols identifying the distributions. There is greater separation of the tumor and nontumor values with two metrics than with either metric by itself (see the projections in the top part). b) Two-dimensional histogram (top) of the values of IR metric L14 vs L1. L1 is absorbance at 1744 cm⁻¹ divided by that at 1458 cm⁻¹, while L14 is absorbance at 1080 cm⁻¹ divided by that at 1244 cm⁻¹. A contour diagram of the same information is given at

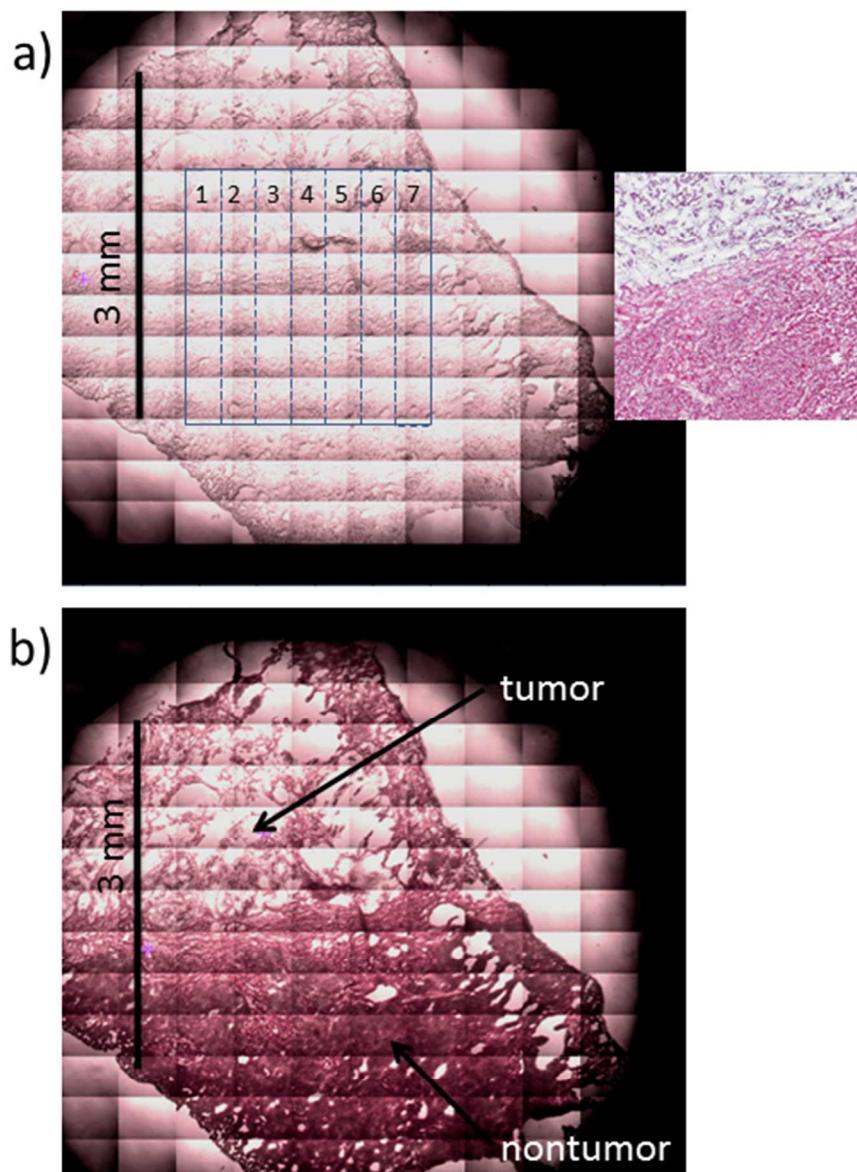
1
2
3 the bottom. A third distribution is evident that was not identified by the 5 group k-means
4 analysis suggesting that more k-means groups would be worthwhile.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Figure 6. A 25 group k-means analysis with nontumor groups in shades of green (top left), tumor groups in shades of purple (top middle), and holes in dark blue-grey (top right). A dendrogram of the distances between groups (bottom) dictated the color scheme and the full image is given as an inset to the dendrogram. Yellow indicates the group closest to the interface between tumor and nontumor.

20
21
22
23
24
25
26
27
28
29
30
31
32
33
Figure 7. IR difference spectrum of lipid in the nontumor region (black trace, green dots) obtained by subtracting 0.963 times the IR spectrum of group 13 (yellow) from the IR spectrum of group 3 (light green), where the colors refer to the inset image. The spectrum was arbitrarily scaled to make the max peak equal to one for comparison to IR search results. The red curve is a composite from the FDM library with 60% glyceryl triacetylrincinoleate and 40% ethyl cyanopolyacrylate.

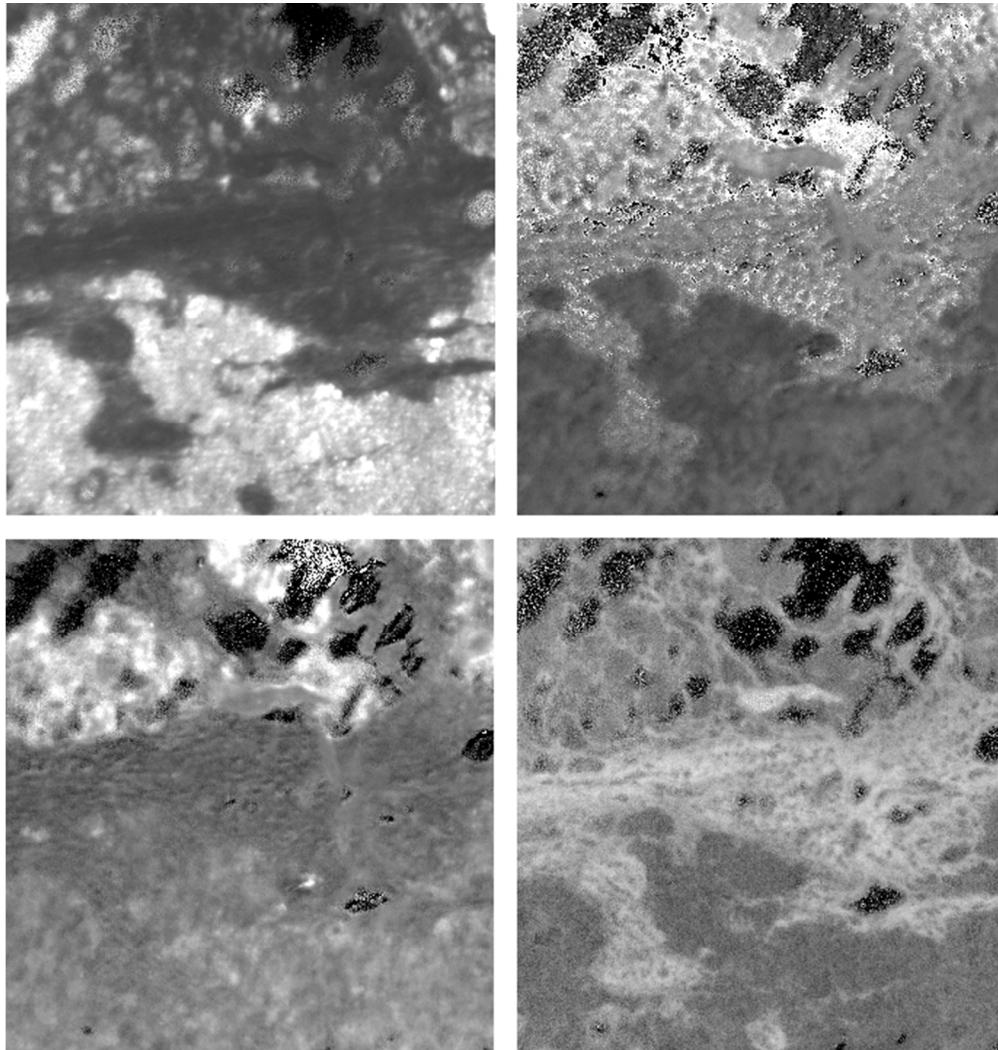
34
35
36
37
38
39
40
41
42
43
44
Figure 8. a) Color coded image of tumor groups 7 and 23 (darker shades of purple) which comprise inner parts of the tumor, groups 24 and 18 which have fibrous stroma cells (lighter shades of purple). Groups 13 (yellow) and 5 (light green) help to indicate the margin. b) Contour diagram of L14 IR metric vs L1 IR metric. A third distribution is identified groups 7 and 23, which was not identified by a 5 group k-means analysis.

45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Figure 9. IR spectra of selected groups from the 25 group k-means analysis. The dark (19) and medium green (9) groups are nontumor, the yellow (13) and light green (5) groups are near the margin, the light purple (24) indicates a region with fibrous stroma cells, and purple (7) indicates an inner tumor region.



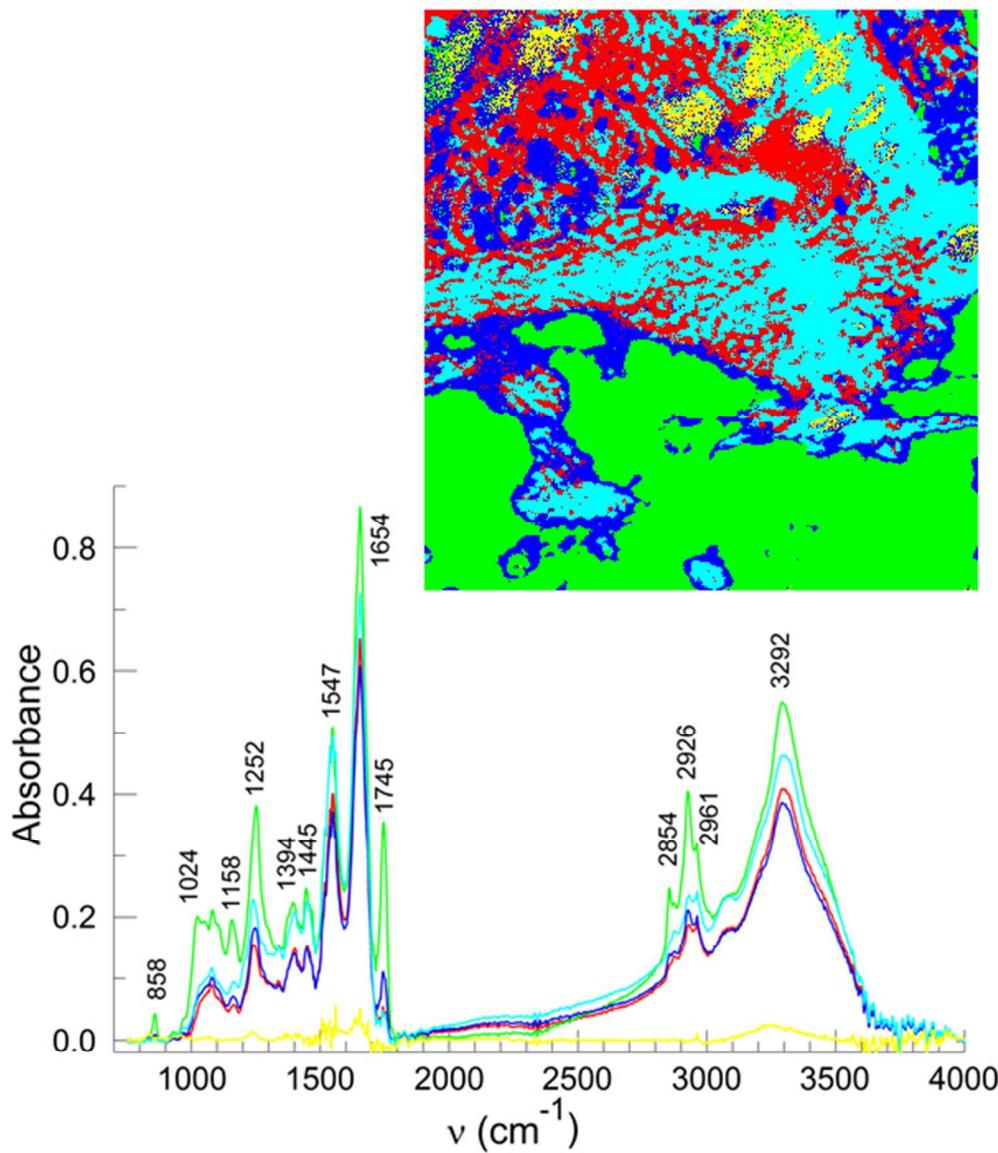
Optical microscope images of a slice of liver tissue with a tumor on ZnSe for IR spectroscopic imaging (scale on left). (a) Seven windows (top) were merged into one window (2.1000 mm wide by 2.2000 mm tall) for IR spectroscopic imaging. An Aperio H&E stain for an optical microscope of a nearby slice at approximately the same place and size is inset at right. The bottom (b) shows the same tissue as in (a) after H&E staining on ZnSe. The tumor is lighter and at the top half of the images.

137x186mm (96 x 96 DPI)



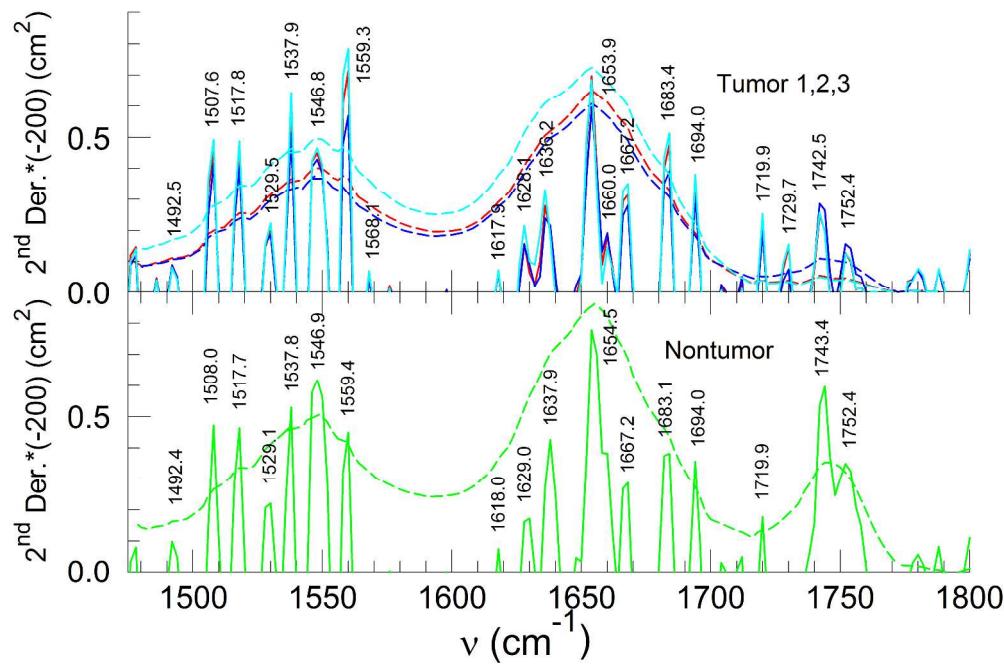
Greyscale bitmap images of selected scaled metrics – L1, L7, L14, and L18 from top left to bottom right. The tumor is largely in the top half of the images.

182x190mm (96 x 96 DPI)



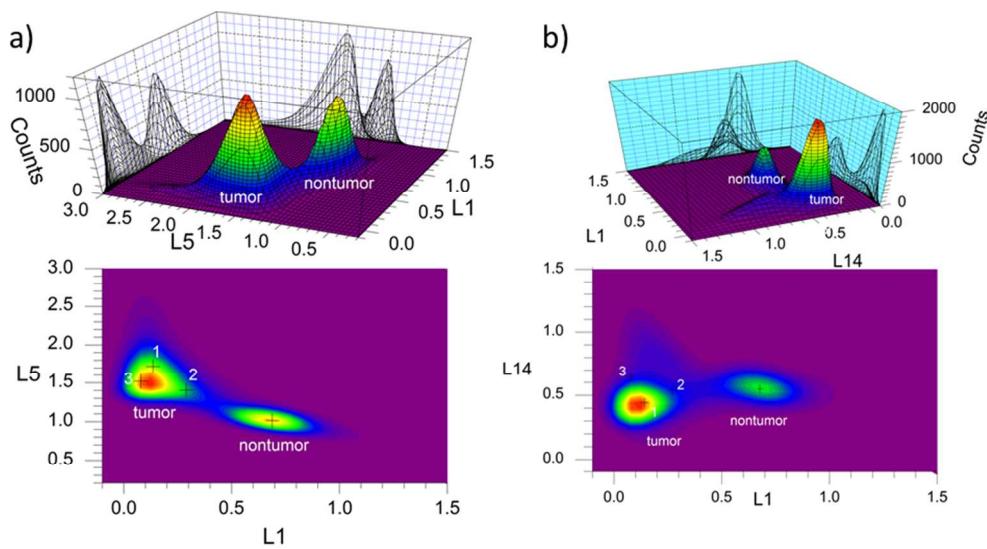
K-means cluster analysis with 5 groups using the unscaled, top-twenty metrics (top). The nontumor portion is in green and the holes are yellow. The tumor has red, blue, and cyan groups. The IR spectrum of each group is given at the bottom with the same color coding.

160x187mm (96 x 96 DPI)



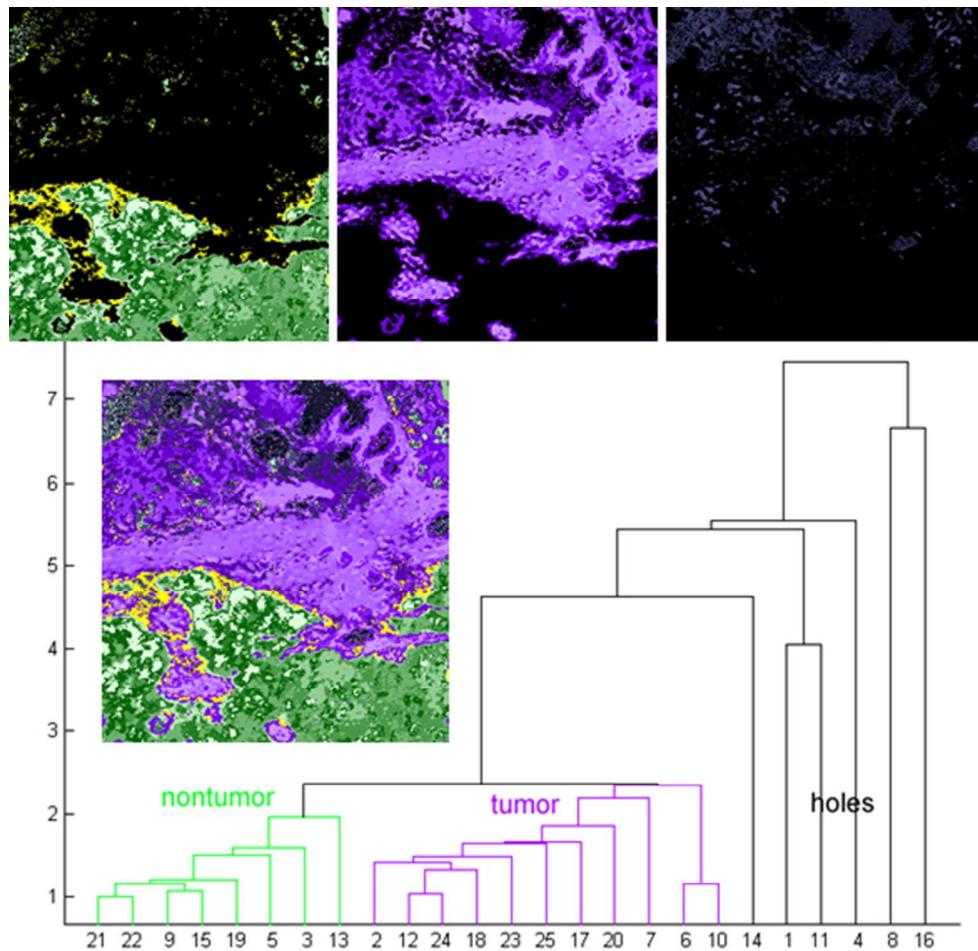
Second derivatives of the IR spectra for each group in Figure 3 (excluding holes) with the same color coding. The dotted traces are absorption spectra, while the solid traces are 2nd derivatives multiplied by -200 in order to put the derivatives on a scale similar to the absorption spectra. Bottom (green) is nontumor and the top has the three tumor groups.

1416x948mm (96 x 96 DPI)



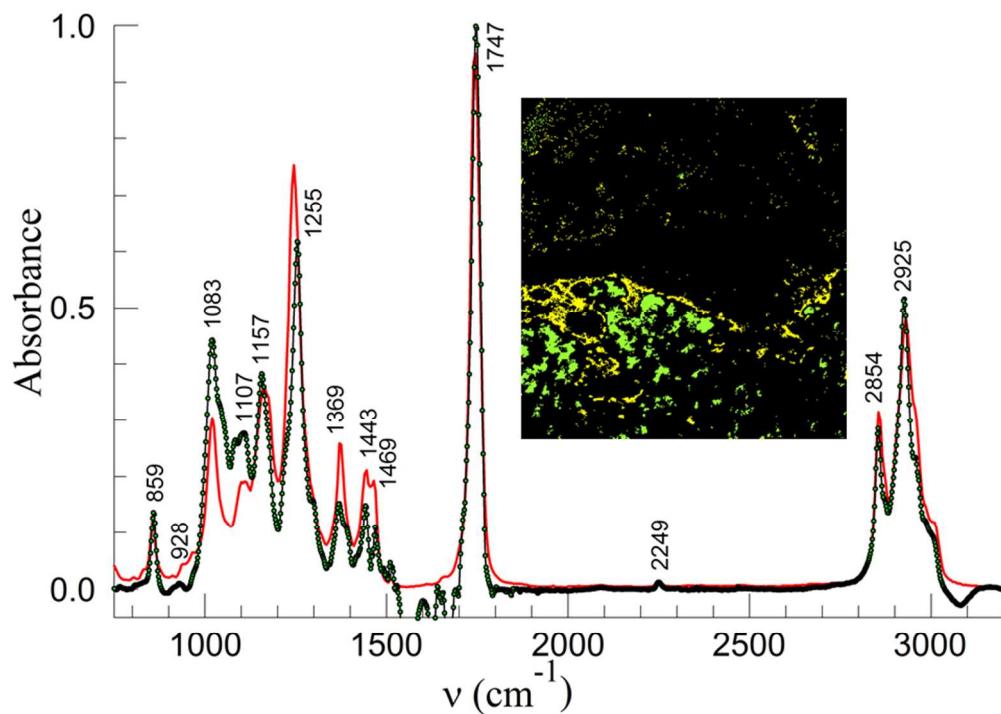
a) Two-dimensional histogram (top) of IR metrics L5 vs L1. L1 is a ratio of ester-linked fat to protein, while L5 is the side of the amide II protein band to phosphate. A contour diagram of the same information is given at the bottom. The centroid values for the k-means groups are given with cross symbols identifying the distributions. There is greater separation of the tumor and nontumor values with two metrics than with either metric by itself (see the projections in the top part). b) Two-dimensional histogram (top) of the values of IR metric L14 vs L1. L1 is absorbance at 1744 cm^{-1} divided by that at 1458 cm^{-1} , while L14 is absorbance at 1080 cm^{-1} divided by that at 1244 cm^{-1} . A contour diagram of the same information is given at the bottom. A third distribution is evident that was not identified by the 5 group k-means analysis suggesting that more k-means groups would be worthwhile.

254x138mm (96 x 96 DPI)



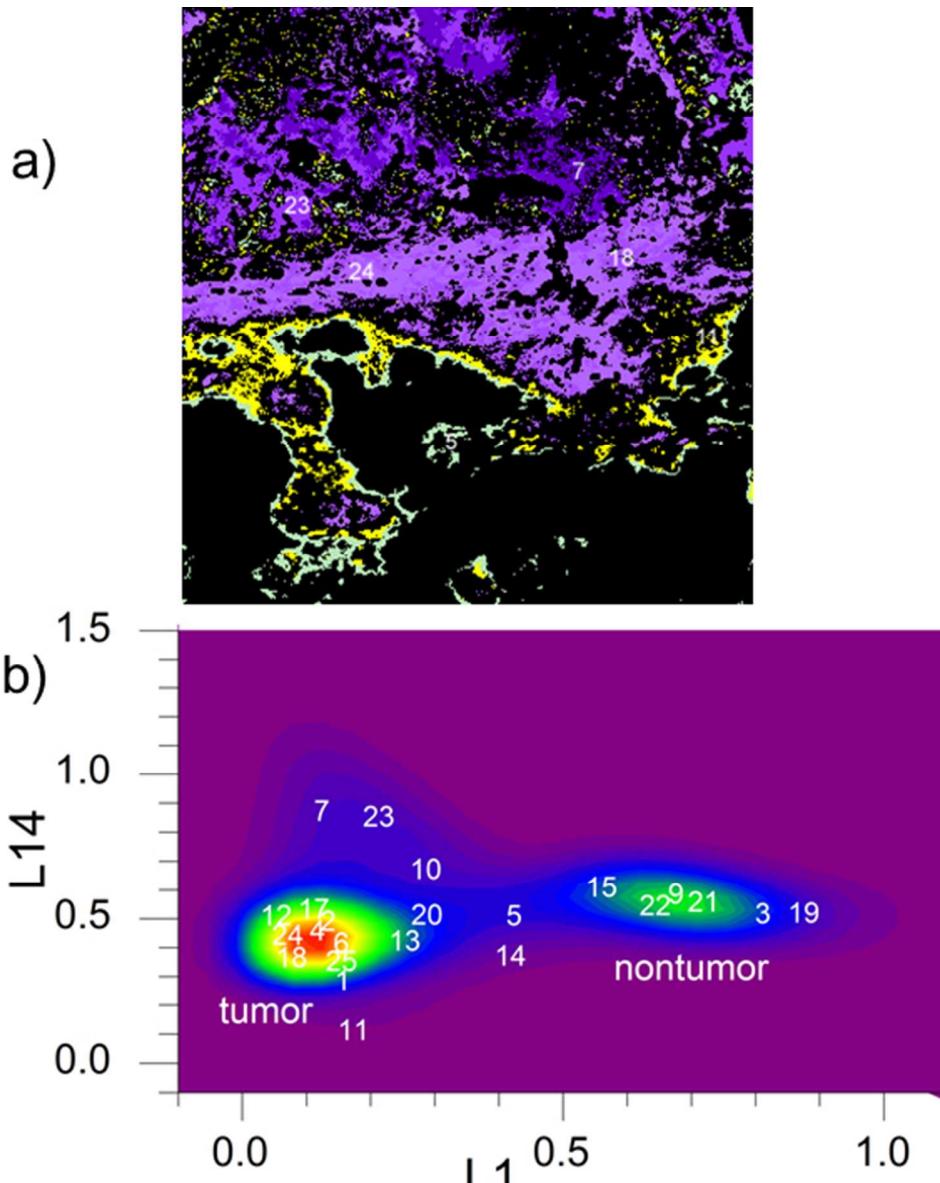
A 25 group k-means analysis with nontumor groups in shades of green (top left), tumor groups in shades of purple (top middle), and holes in dark blue-grey (top right). A dendrogram of the distances between groups (bottom) dictated the color scheme and the full image is given as an inset to the dendrogram. Yellow indicates the group closest to the interface between tumor and nontumor.

158x154mm (96 x 96 DPI)



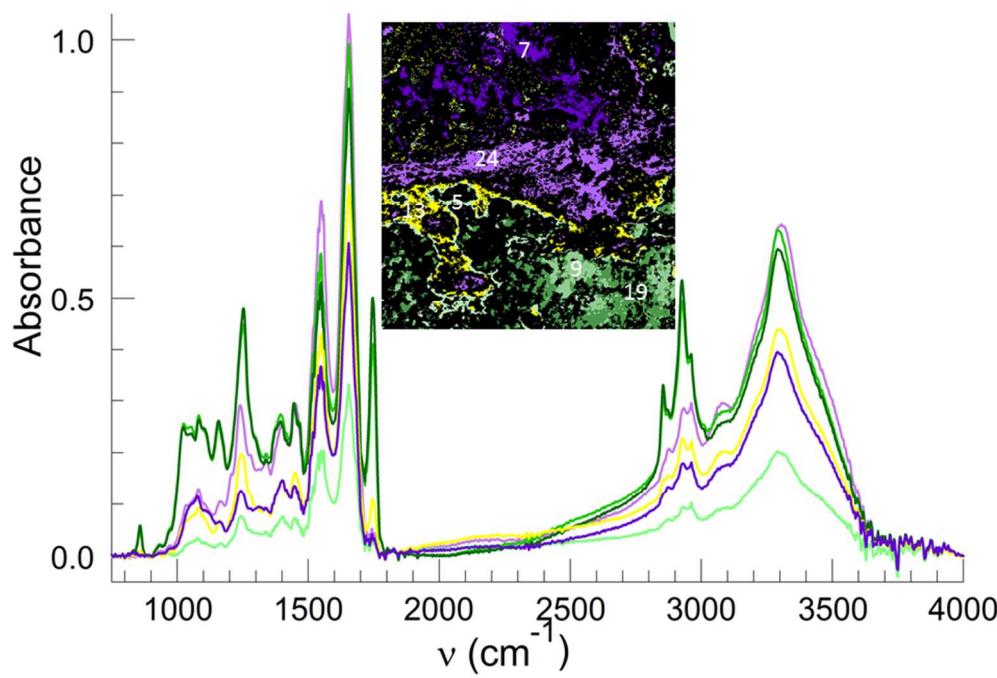
IR difference spectrum of lipid in the nontumor region (black trace, green dots) obtained by subtracting 0.963 times the IR spectrum of group 13 (yellow) from the IR spectrum of group 3 (light green), where the colors refer to the inset image. The spectrum was arbitrarily scaled to make the max peak equal to one for comparison to IR search results. The red curve is a composite from the FDM library with 60% glyceryl triacetylrincinoleate and 40% ethyl cyanopolycrylate.

254x190mm (96 x 96 DPI)



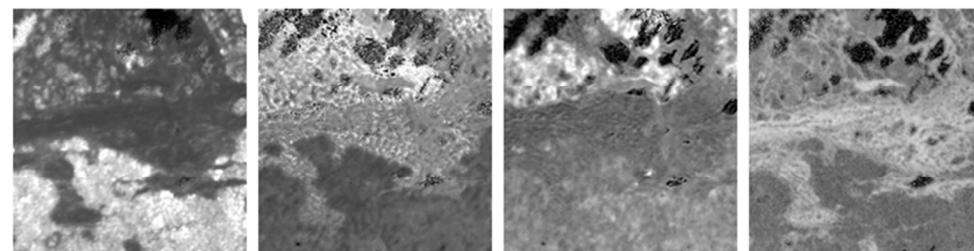
a) Color coded image of tumor groups 7 and 23 (darker shades of purple) which comprise inner parts of the tumor, groups 24 and 18 which have fibrous stroma cells (lighter shades of purple). Groups 13 (yellow) and 5 (light green) help to indicate the margin. b) Contour diagram of L14 IR metric vs L1 IR metric. A third distribution is identified groups 7 and 23, which was not identified by a 5 group k-means analysis.

153x190mm (96 x 96 DPI)



IR spectra of selected groups from the 25 group k-means analysis. The dark (19) and medium green (9) groups are nontumor, the yellow (13) and light green (5) groups are near the margin, the light purple (24) indicates a region with fibrous stroma cells, and purple (7) indicates an inner tumor region.

254x170mm (96 x 96 DPI)



Different IR biomarkers showing sensitivity to different regions of liver tissue with a tumor.
181x49mm (96 x 96 DPI)