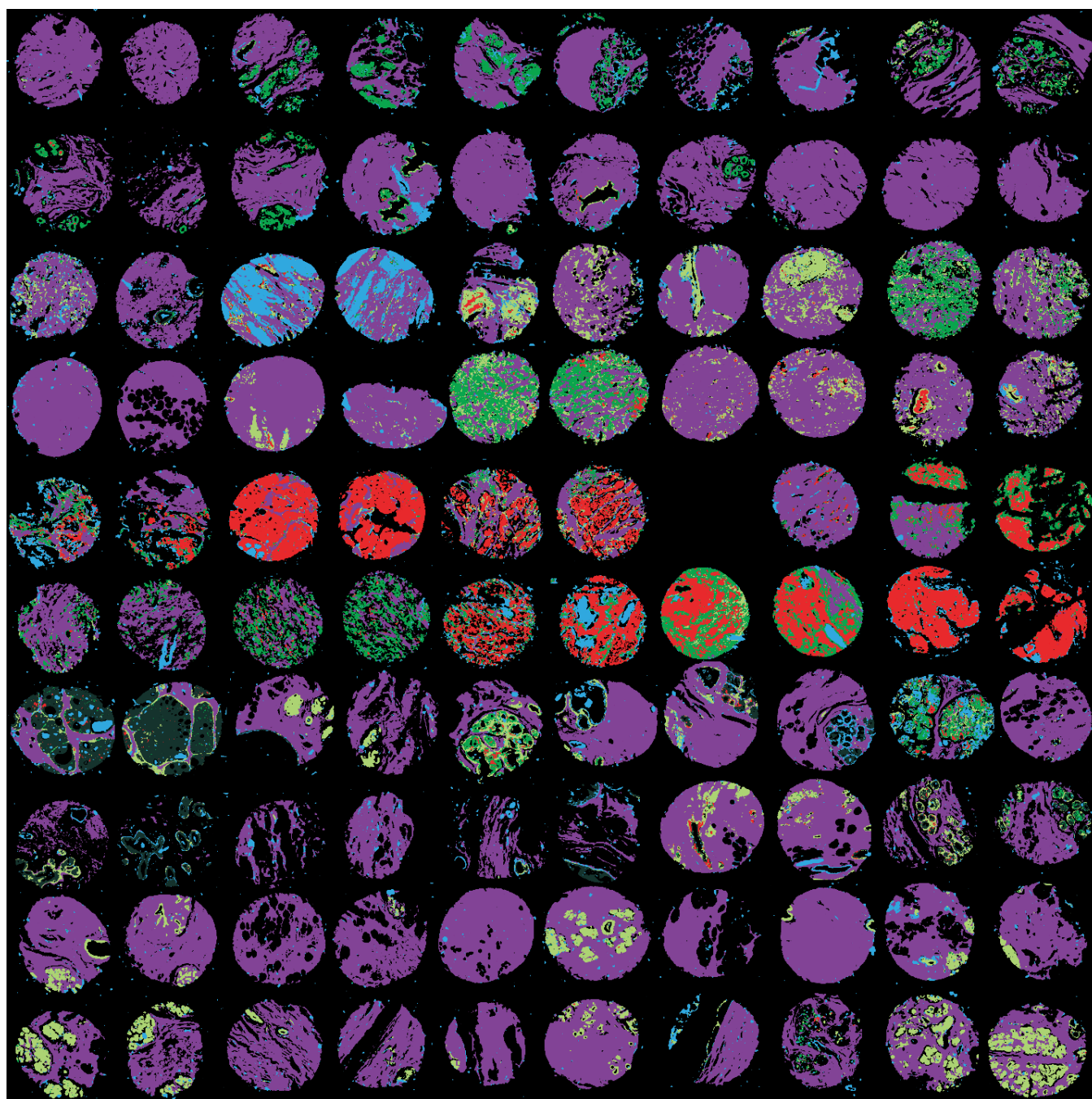
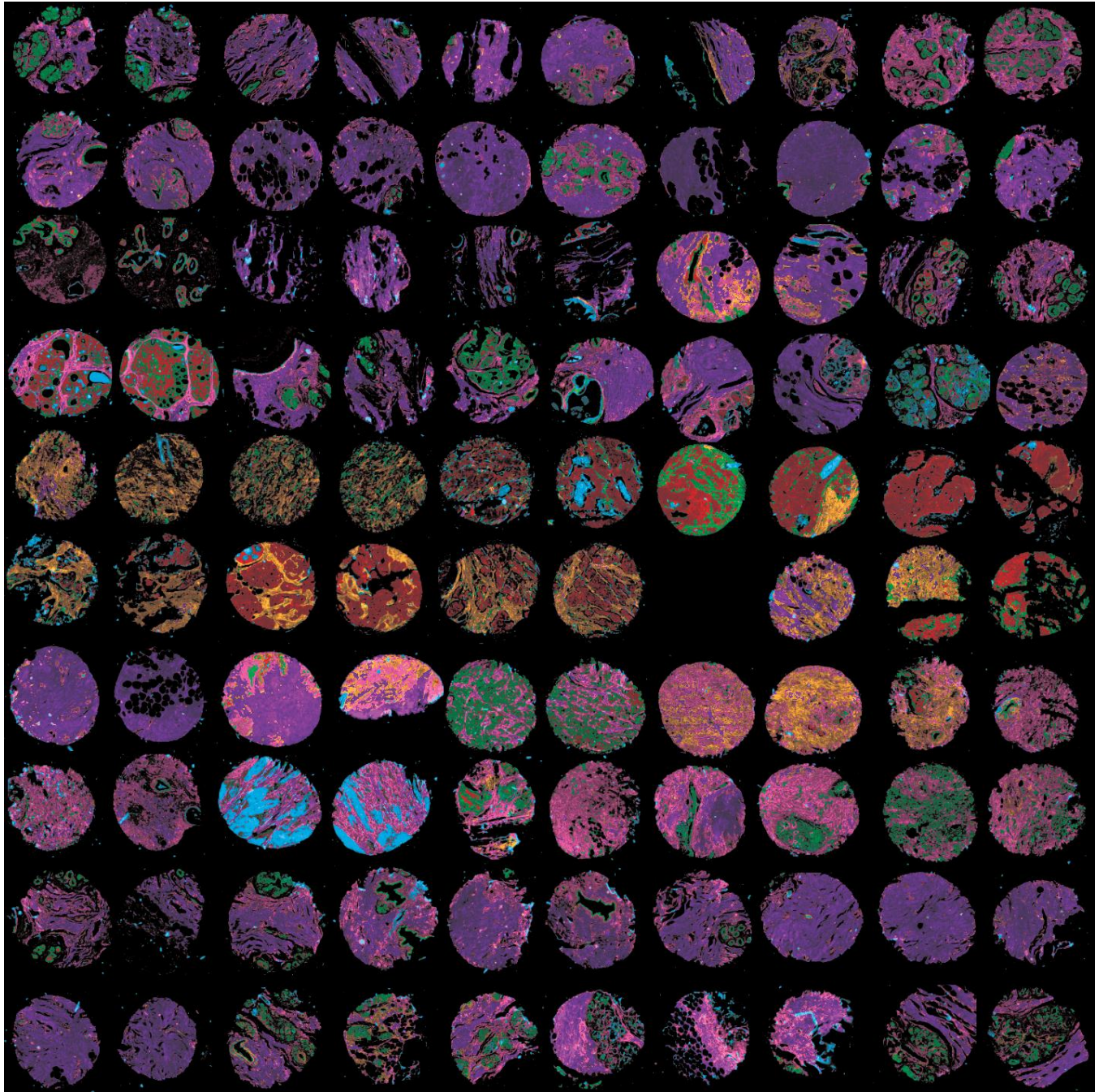


## Supporting Information



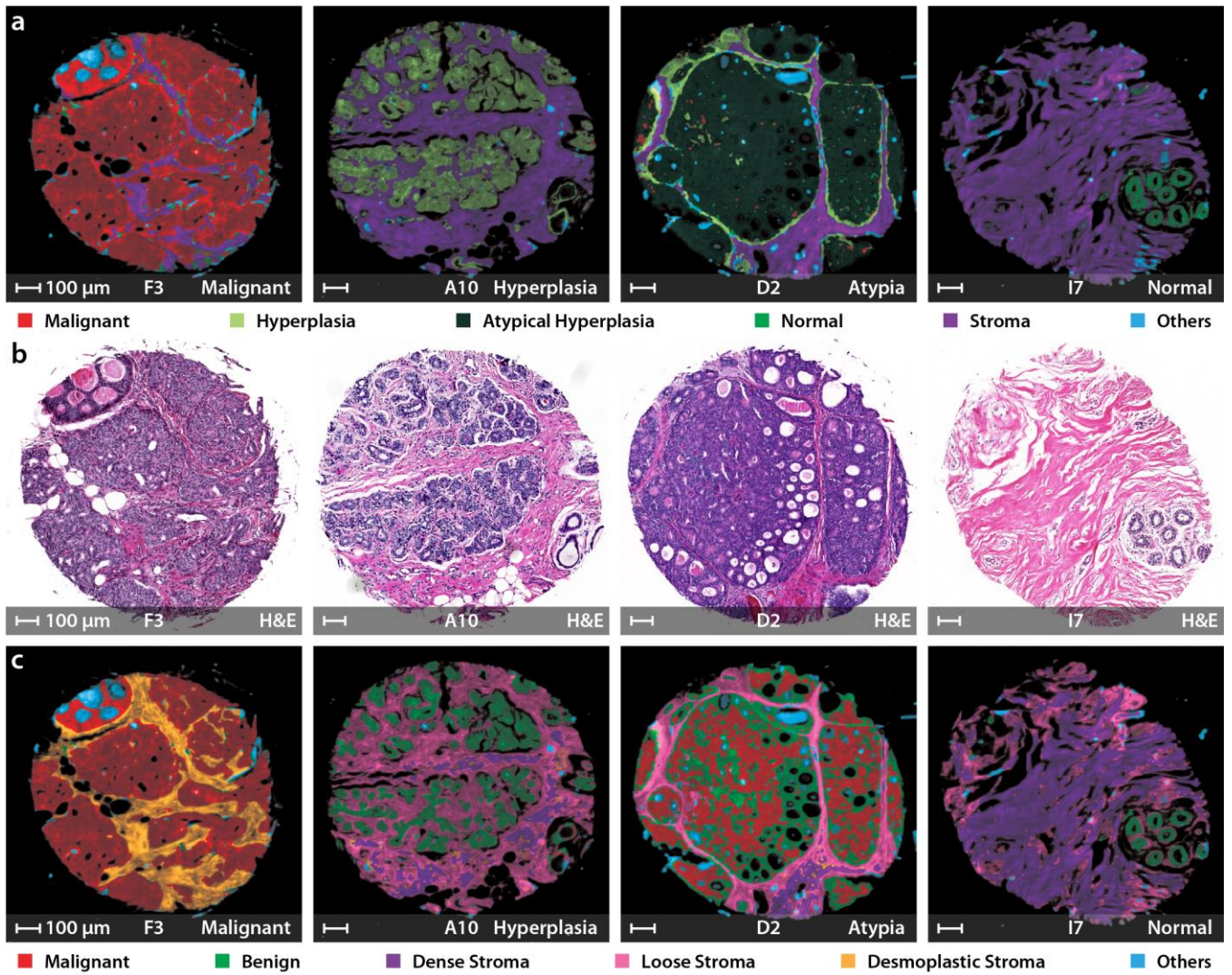
**Fig. S1. Epithelial tumor classification and microenvironment models for breast cancer characterization using FT-IR imaging of the BR1003 microarray.**



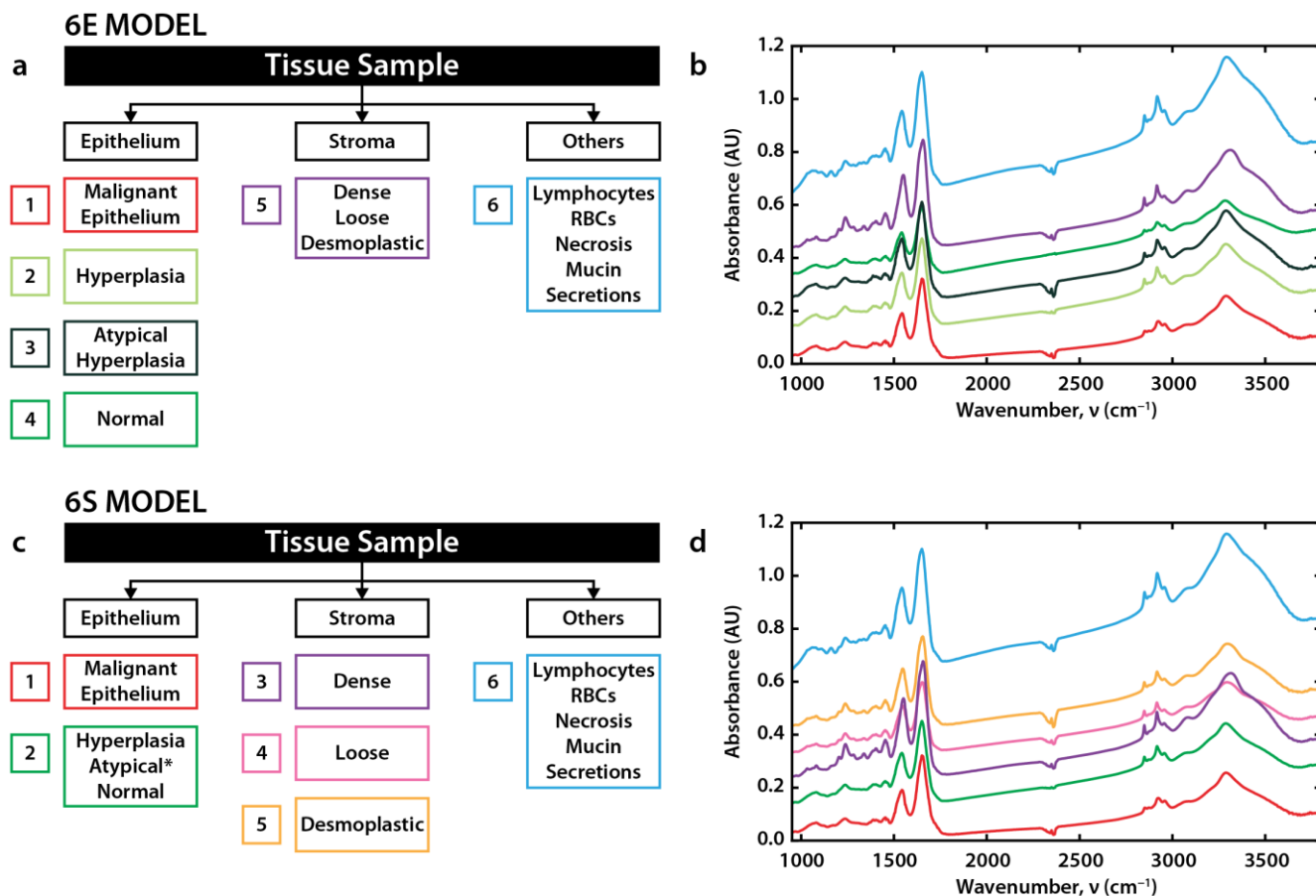


**Fig. S2. Stromal tumor classification and microenvironment models for breast cancer characterization using FT-IR imaging of the BR1003 microarray.**





**Fig. S3.** Select cores acquired using FTIR and classified using the epithelial (a) and stromal models (b) FT-IR imaging compared to H&E stained tissue (c).



**Fig. S4: Detailed description of the classification models based on breast tissue histopathology. a) Different components of the epithelial characterization model ( 6E model), the first class being malignant epithelium ( tumor regions), 2-4 being the various types of benign lesions, 5th as stroma and the last class as others (individual components listed in the chart). b) Average spectra of the six classes of the 6E model with the bottommost spectrum being of class 1 and the topmost of class 6. c) Comprehensive listing of the components of the 6S model with the benign lesions combined and the stromal compartments segregated into different states. \* In the benign lesions, atypical pixels are not used for training, as they are known to have malignant signatures. d) Average spectra of the individual classes of the 6S model.**

**Table S1. Improved classification accuracy (tested on a subset of patient cases) after applying the minimum noise fraction (MNF) technique.**

<b>Confusion Matrix of 6S Model without MNF</b>						
<b>Classes</b>	<b>Malignant Epithelium</b>	<b>Benign Epithelium</b>	<b>Dense Stroma</b>	<b>Loose Stroma</b>	<b>Desmoplastic Stroma</b>	<b>Others</b>
<b>Malignant Epithelium</b>	<b>0.95</b>	0.03	0.00	0.00	0.01	0.01
<b>Benign Epithelium</b>	0.04	<b>0.92</b>	0.00	0.01	0.01	0.02
<b>Dense Stroma</b>	0.00	0.00	<b>0.86</b>	0.13	0.00	0.00
<b>Loose Stroma</b>	0.01	0.03	0.08	<b>0.78</b>	0.03	0.07
<b>Desmoplastic Stroma</b>	0.02	0.01	0.00	0.02	<b>0.92</b>	0.04
<b>Others</b>	0.03	0.04	0.00	0.05	0.05	<b>0.83</b>

<b>Confusion Matrix of 6S Model with MNF</b>						
<b>Classes</b>	<b>Malignant Epithelium</b>	<b>Benign Epithelium</b>	<b>Dense Stroma</b>	<b>Loose Stroma</b>	<b>Desmoplastic Stroma</b>	<b>Others</b>
<b>Malignant Epithelium</b>	<b>1.00</b>	0.00	0.00	0.00	0.00	0.00
<b>Benign Epithelium</b>	0.00	<b>0.99</b>	0.00	0.00	0.00	0.00
<b>Dense Stroma</b>	0.00	0.00	<b>0.99</b>	0.01	0.00	0.00
<b>Loose Stroma</b>	0.00	0.00	0.02	<b>0.97</b>	0.00	0.00
<b>Desmoplastic Stroma</b>	0.00	0.00	0.00	0.00	<b>0.99</b>	0.00
<b>Others</b>	0.00	0.00	0.00	0.02	0.02	<b>0.95</b>

**Table S2: 134 metric definitions used in all the high definition FT-IR classification models. Abbreviations: H: Peak Height; A: Peak Area at the indicated the wavenumber. For instance, the H<sub>1545</sub>: H<sub>1653</sub> indicates the peak height ratio of 1545 cm<sup>-1</sup> and 1653 cm<sup>-1</sup> wavenumbers.**

Peak Height Ratios					
H <sub>1032</sub> : H <sub>1082</sub>	H <sub>1032</sub> : H <sub>1238</sub>	H <sub>1032</sub> : H <sub>1653</sub>	H <sub>1032</sub> : H <sub>2848</sub>	H <sub>1032</sub> : H <sub>3288</sub>	
H <sub>1082</sub> : H <sub>1340</sub>	H <sub>1082</sub> : H <sub>1454</sub>	H <sub>1082</sub> : H <sub>2848</sub>	H <sub>1082</sub> : H <sub>2873</sub>	H <sub>1082</sub> : H <sub>2916</sub>	
H <sub>1238</sub> : H <sub>1340</sub>					
H <sub>1340</sub> : H <sub>1454</sub>	H <sub>1340</sub> : H <sub>1545</sub>	H <sub>1340</sub> : H <sub>2848</sub>	H <sub>1340</sub> : H <sub>2873</sub>	H <sub>1340</sub> : H <sub>2916</sub>	H <sub>1340</sub> : H <sub>3288</sub>
H <sub>1454</sub> : H <sub>2848</sub>	H <sub>1454</sub> : H <sub>2956</sub>	H <sub>1454</sub> : H <sub>3288</sub>			
H <sub>1545</sub> : H <sub>1653</sub>	H <sub>1545</sub> : H <sub>2873</sub>	H <sub>1545</sub> : H <sub>3288</sub>			
H <sub>1653</sub> : H <sub>3288</sub>					
H <sub>2873</sub> : H <sub>2956</sub>					
H <sub>2916</sub> : H <sub>2956</sub>	H <sub>2916</sub> : H <sub>3288</sub>				
H <sub>3080</sub> : H <sub>3288</sub>					

Peak Area to Height Ratios					
A <sub>984-1045</sub> : H <sub>1082</sub>	A <sub>984-1045</sub> : H <sub>1340</sub>	A <sub>984-1045</sub> : H <sub>1545</sub>	A <sub>984-1045</sub> : H <sub>2916</sub>		
A <sub>1070-1142</sub> : H <sub>1032</sub>	A <sub>1070-1142</sub> : H <sub>1238</sub>	A <sub>1070-1142</sub> : H <sub>1454</sub>	A <sub>1070-1142</sub> : H <sub>1653</sub>	A <sub>1070-1142</sub> : H <sub>3288</sub>	
A <sub>1182-1217</sub> : H <sub>1032</sub>	A <sub>1182-1217</sub> : H <sub>1340</sub>	A <sub>1182-1217</sub> : H <sub>2873</sub>	A <sub>1182-1217</sub> : H <sub>2916</sub>	A <sub>1182-1217</sub> : H <sub>3080</sub>	
A <sub>1217-1271</sub> : H <sub>1032</sub>	A <sub>1217-1271</sub> : H <sub>1454</sub>	A <sub>1217-1271</sub> : H <sub>2916</sub>	A <sub>1217-1271</sub> : H <sub>2956</sub>		
A <sub>1321-1354</sub> : H <sub>1545</sub>	A <sub>1321-1354</sub> : H <sub>1653</sub>	A <sub>1321-1354</sub> : H <sub>2873</sub>	A <sub>1321-1354</sub> : H <sub>2916</sub>		
A <sub>1373-1394</sub> : H <sub>1082</sub>	A <sub>1373-1394</sub> : H <sub>1340</sub>	A <sub>1373-1394</sub> : H <sub>1653</sub>	A <sub>1373-1394</sub> : H <sub>2873</sub>		
A <sub>1427-1483</sub> : H <sub>1032</sub>	A <sub>1427-1483</sub> : H <sub>2848</sub>	A <sub>1427-1483</sub> : H <sub>2873</sub>	A <sub>1427-1483</sub> : H <sub>2916</sub>	A <sub>1427-1483</sub> : H <sub>2956</sub>	A <sub>1427-1483</sub> : H <sub>3080</sub>
A <sub>1483-1587</sub> : H <sub>1032</sub>	A <sub>1483-1587</sub> : H <sub>1238</sub>	A <sub>1483-1587</sub> : H <sub>1454</sub>	A <sub>1483-1587</sub> : H <sub>2916</sub>	A <sub>1483-1587</sub> : H <sub>3288</sub>	
A <sub>1587-1766</sub> : H <sub>1082</sub>	A <sub>1587-1766</sub> : H <sub>1238</sub>	A <sub>1587-1766</sub> : H <sub>1545</sub>	A <sub>1587-1766</sub> : H <sub>2873</sub>	A <sub>1587-1766</sub> : H <sub>2956</sub>	A <sub>1587-1766</sub> : H <sub>3080</sub>
A <sub>2816-2862</sub> : H <sub>1238</sub>	A <sub>2816-2862</sub> : H <sub>1340</sub>	A <sub>2816-2862</sub> : H <sub>1653</sub>	A <sub>2816-2862</sub> : H <sub>2873</sub>		
A <sub>2862-2893</sub> : H <sub>1032</sub>	A <sub>2862-2893</sub> : H <sub>1238</sub>	A <sub>2862-2893</sub> : H <sub>1454</sub>	A <sub>2862-2893</sub> : H <sub>2916</sub>	A <sub>2862-2893</sub> : H <sub>3080</sub>	
A <sub>2893-2947</sub> : H <sub>1032</sub>	A <sub>2893-2947</sub> : H <sub>1082</sub>	A <sub>2893-2947</sub> : H <sub>2848</sub>			
A <sub>2947-3000</sub> : H <sub>1082</sub>	A <sub>2947-3000</sub> : H <sub>1238</sub>	A <sub>2947-3000</sub> : H <sub>1545</sub>	A <sub>2947-3000</sub> : H <sub>1653</sub>		
A <sub>3000-3118</sub> : H <sub>1032</sub>	A <sub>3000-3118</sub> : H <sub>1454</sub>	A <sub>3000-3118</sub> : H <sub>2873</sub>			
A <sub>3118-3674</sub> : H <sub>1082</sub>	A <sub>3118-3674</sub> : H <sub>1238</sub>	A <sub>3118-3674</sub> : H <sub>1545</sub>	A <sub>3118-3674</sub> : H <sub>2873</sub>	A <sub>3118-3674</sub> : H <sub>2916</sub>	A <sub>3118-3674</sub> : H <sub>2856</sub> A <sub>3118-3674</sub> : H <sub>3080</sub>

Peak Area to Area Ratios				
A <sub>984-1045</sub> : A <sub>1217-1271</sub>	A <sub>984-1045</sub> : A <sub>1587-1766</sub>	A <sub>984-1045</sub> : A <sub>2893-2947</sub>	A <sub>984-1045</sub> : A <sub>2947-3000</sub>	A <sub>984-1045</sub> : A <sub>3000-3118</sub>
A <sub>1070-1142</sub> : A <sub>1217-1271</sub>	A <sub>1070-1142</sub> : A <sub>1321-1354</sub>	A <sub>1070-1142</sub> : A <sub>1483-1587</sub>	A <sub>1070-1142</sub> : A <sub>1587-1766</sub>	A <sub>1070-1142</sub> : A <sub>2947-3000</sub>
A <sub>1182-1217</sub> : A <sub>1427-1483</sub>	A <sub>1182-1217</sub> : A <sub>2862-2893</sub>	A <sub>1182-1217</sub> : A <sub>3000-3118</sub>	A <sub>1182-1217</sub> : A <sub>3118-3674</sub>	
A <sub>1217-1271</sub> : A <sub>1321-1354</sub>	A <sub>1217-1271</sub> : A <sub>1483-1587</sub>	A <sub>1217-1271</sub> : A <sub>1587-1766</sub>	A <sub>1217-1271</sub> : A <sub>3000-3118</sub>	
A <sub>1321-1354</sub> : A <sub>1373-1394</sub>	A <sub>1321-1354</sub> : A <sub>1483-1587</sub>	A <sub>1321-1354</sub> : A <sub>2893-2947</sub>	A <sub>1321-1354</sub> : A <sub>3000-3118</sub>	A <sub>1321-1354</sub> : A <sub>3118-3674</sub>
A <sub>1373-1394</sub> : A <sub>1587-1766</sub>	A <sub>1373-1394</sub> : A <sub>2816-2862</sub>	A <sub>1373-1394</sub> : A <sub>2947-3000</sub>	A <sub>1373-1394</sub> : A <sub>3000-3118</sub>	
A <sub>1427-1483</sub> : A <sub>1483-1587</sub>	A <sub>1427-1483</sub> : A <sub>2816-2862</sub>	A <sub>1427-1483</sub> : A <sub>2862-2893</sub>		
A <sub>1483-1587</sub> : A <sub>1587-1766</sub>	A <sub>1483-1587</sub> : A <sub>3118-3674</sub>			
A <sub>1587-1766</sub> : A <sub>2816-2862</sub>	A <sub>1587-1766</sub> : A <sub>3000-3118</sub>			
A <sub>2816-2862</sub> : A <sub>2893-2947</sub>	A <sub>2816-2862</sub> : A <sub>2947-3000</sub>			
A <sub>2862-2893</sub> : A <sub>2893-2947</sub>				

**Table S3: Detailed description of the important features in the different classification models obtained using the FT-IR data. Abbreviations: H: Peak Height; A: Peak Area at the indicated the wavenumber. For instance, the H<sub>1545</sub>: H<sub>1653</sub> indicates the peak height ratio of 1545 cm<sup>-1</sup> and 1653 cm<sup>-1</sup> wavenumbers.**

Metric	Biochemical Significance
H <sub>1545</sub> : H <sub>1653</sub>	Amide II to Amide I ratio indicates protein profile of the sample both in composition and secondary structure
A <sub>1387</sub> : H <sub>1653</sub>	Relative concentration of aliphatic groups compared to proteins and also composition of different proteins
A <sub>2873</sub> : H <sub>2916</sub>	Aliphatic chain length and the extent of branching. This is important to distinguish between the different types of stroma and capturing differences in the composition of the amino groups
A <sub>1080</sub> : A <sub>1340</sub>	DNA/RNA to CH <sub>2</sub> ratio for nuclear to cytoplasmic ratio, which is known to change during disease progression
A <sub>1032</sub> : H <sub>1080</sub>	DNA, RNA and glycosylation profiles for epithelial segregation and differentiating between reactive and other stroma.
A <sub>1080</sub> : A <sub>1238</sub>	Relative concentration of nucleic acids to collagen and glycogen glycosylation
H <sub>2873</sub> : H <sub>2956</sub> & A <sub>2916</sub> : H <sub>2848</sub>	Relative CH <sub>2</sub> to CH <sub>3</sub> concentration important for stroma model to separate between the different types of stroma
A <sub>1032</sub> : A <sub>3080</sub>	Relative concentration of DNA/RNA to proteins

**Table S4: 7 metric definitions used in all the QCL classification models. Abbreviations: H: Peak Height; A: Peak Area at the indicated the wavenumber. For instance, the H<sub>1546</sub>: H<sub>1658</sub> indicates the peak height ratio of 1546 cm<sup>-1</sup> and 1658 cm<sup>-1</sup> wavenumbers and A<sub>1483,1587,1658,1765</sub> represents the spectral area of the 1658 cm<sup>-1</sup> peak (underlined) calculated from these bounds.**

#### Metric Definitions for QCL Data

H <sub>1082</sub> : H <sub>1238</sub>	H <sub>1386</sub> : H <sub>1238</sub>	H <sub>1658</sub> : H <sub>1546</sub>
A <sub>984,<u>1082</u>,1143</sub> : A <sub>1143,<u>1238</u>,1278</sub>	A <sub>1143,<u>1238</u>,1278</sub> : A <sub>1353,<u>1386</u>,1483</sub>	A <sub>1483,<u>1546</u>,1587,1765</sub> : A <sub>1483,1587,<u>1658</u>,1765</sub> A <sub>1353,<u>1386</u>,1483</sub> : A <sub>1483,1587,<u>1658</u>,1765</sub>

**Table S5: Biochemical significance of absorbance peaks (Functional group and vibrational modes are listed below as described by Rehman IU, Movasaghi Z, Rehman S (2013) Vibrational spectroscopy for tissue analysis. CRC Press, Taylor and Francis Group, FL)**

Peak (cm <sup>-1</sup> )	Commonly accepted Biological species/ Functional Group Assignment	IR Band Assignments
1082	DNA/RNA	PO <sub>2</sub> symmetric stretch
1238	DNA/RNA	PO <sub>2</sub> asymmetric stretch
1386	Aliphatic groups	CH <sub>3</sub> symmetric bend
1546	Amide backbone	N-H bend and C-H stretch
1658	Amide backbone	C=O stretch