iScience, Volume *27*

Supplemental information

Facilitating clinically relevant skin tumor diagnostics with spectroscopy-driven machine learning

Emil Andersson, Jenny Hult, Carl Troein, Magne Stridh, Benjamin Sjögren, Agnes Pekar-Lukacs, Julio Hernandez-Palacios, Patrik Edén, Bertil Persson, Victor Olariu, Malin Malmsjö, and Aboma Merdasa

Table S1: Patient data, related to STAR Methods.

Lesion (patient) no.	Sex	Age	Fitzpatrick skin type (1-VI)	Diagnosis
1 (1)	Male	60's	I	Malignant melanoma, in situ
2 (2)	Male	50's	II	Dysplastic nevus
3 (3)	Male	60's	II	Fibrohistiocytoma
4 (4)	Male	60's	II	Basal cell carcinoma
5 (5)	Male	50's	N/I	Papilloma
6 (6)	Female	40's	II	Dysplastic nevus
7 (7)	Male	80's	II	Squamous cell carcinoma
8 (8)	Male	20's	IV	Blue nevus
9 (9)	Female	40's	II	Nevus
10 (10)	Female	70's	II	Basal cell carcinoma
11 (11)	Female	60's	II	Malignant melanoma, in situ
12 (12)	Female	70's	II	Basal cell carcinoma
13 (13)	Female	60's	II	Squamous cell carcinoma
14 (14)	Male	60's	II	Nevus
15 (15)	Male	60's	II	Nevus and blue nevus
16 (16)	Female	60's	II	Basal cell carcinoma
17 (17)	Male	40's	II	Malignant melanoma, invasive
18 (17)	Male	40's	II	Nevus
19 (18)	Male	80's	II	Malignant melanoma, invasive

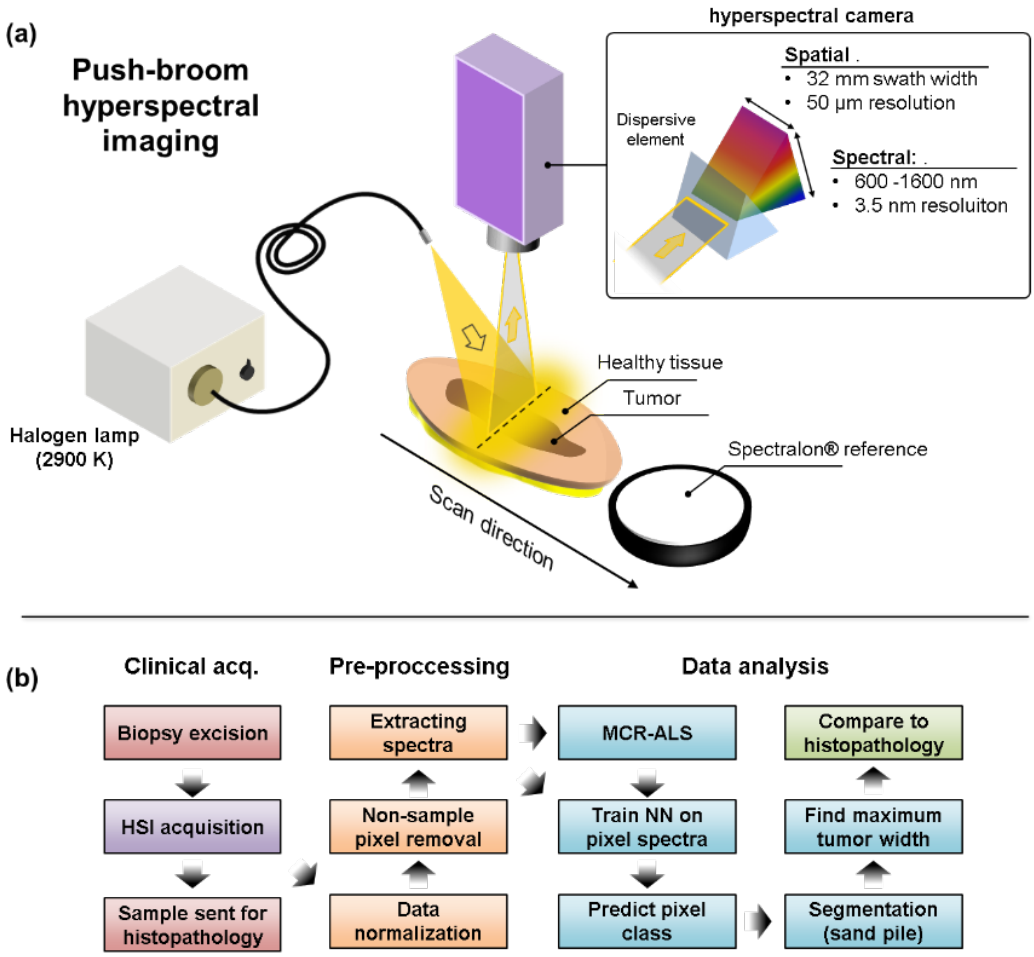

Figure S1: Method overview, related to STAR Methods.
(a) Schematic demonstrating the spatial scanning hyperspectral measurement geometry in which a white incandescent lamp is used to illuminate the sample surface. A line is thereafter imaged, via a slit, onto an area detector. Prior to reaching the detector, the light is dispersed by a prism at each point along the line such that one dimension of the detector area captures the spectral information, while the other dimension captures the spatial information. By scanning the entire system across a surface, a hyperspectral image is obtained. A white reference placed at the same height as the sample is also scanned in every measurement. (b) Flow chart demonstrating the procedure going from acquisition of data from patient using HSI, through pre-processing of the data and finally to the data analysis.

Table S2: Data properties, related to Figure 3 and STAR Methods.
The number of pixels each sample consists of, the number of pixels in the selected tumor and healthy tissue areas, and the samples' histopathological widths.

Lesion no.	# sample pixels	# tumorous pixels	# healthy pixels	Histopathological width (mm)
1	93363	709	754	6.89
2	210700	1961	2514	15.7
3	152591	1009	754	6.78
4	30534	197	226	7.83
6	27404	197	226	2.39
7	136424	1257	1226	13
8	174442	1009	1226	9.45
9	49739	1257	298	4.89
10	49468	1961	2018	3.45
11	149952	1257	1594	8.58
12	143629	1653	1594	6.76
13	46080	529	394	4.97
14	62715	709	506	2.86
15	18392	317	298	1.6
16	55359	709	754	6.61
17	112117	1653	1594	8.31
18	52377	613	754	4.84
19	145197	1961	2018	8.54

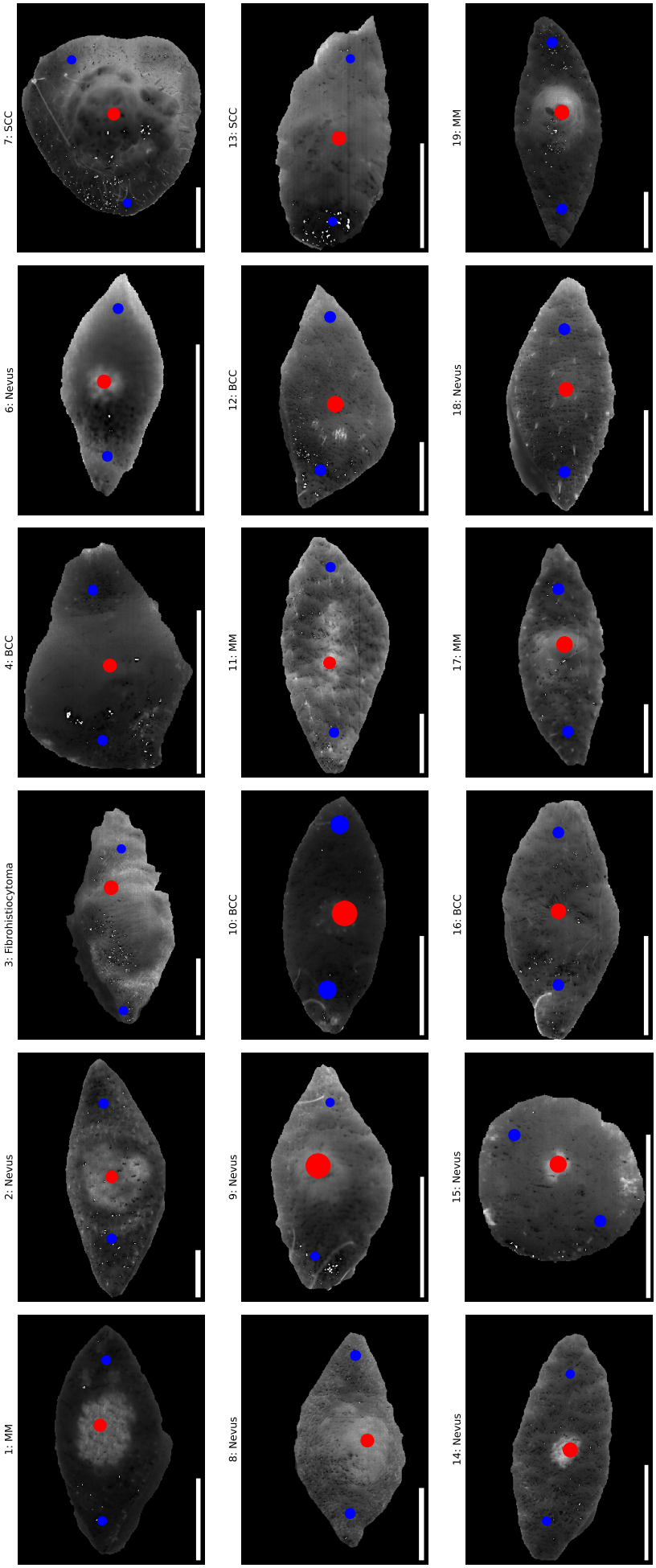

Figure S2: Training regions, related to Figure 3.
The training regions used for each sample with blue representing healthy tissue and red tumor. The tumor samples are shown at 640 nm.

Table S3: Model hyperparametes, related to STAR Methods.

(a) Grid search parameters.

Hyperparameter	Search values
Learning rate	$[0.025, 0.07, 0.01, 0.15]$
Batch size	$[0.2, 0.5, 1.0]$

(b) Network architectures. The square brackets denotes an architecture where each item is a hidden layer. All convolutional filters and maxpoolings are 1-dimensional. XC denotes number of channels and YS denotes stride Y.

Model	Architecture
MLP, MLP: MCR-ALS	[2], [8], [20], [100], [2,2], [8,8], [100,100], [2,2,2], [8, 8, 8]
CNN	[Conv(10, 10C, 10S), MaxPool(2, 2S), Conv(5, 20C, 1S), Dense(10)], [Conv(10, 10C, 10S), MaxPool(2, 2S), Conv(5, 20C, 1S), MaxPool(2, 2S), Conv(5, 20C, 1S), Dense(10)], [Conv(10, 10C, 5S), MaxPool(2, 2S), Conv(5, 20C, 5S), MaxPool(2, 2S), Conv(3, 20C, 2S), Dense(10)], [Conv(20, 10C, 10S), MaxPool(2, 2S), Conv(10, 20C, 5S), Dense(10)], [Conv(5, 10C, 1S), Conv(10, 10C, 1S), MaxPool(2, 2S), Conv(10, 20C, 1S), Dense(10)], [Conv(20, 10C, 5S) MaxPool(2, 2S), Conv(10, 10C, 2S), MaxPool(2, 2S), Conv(5, 10C, 1S), Dense(10)]

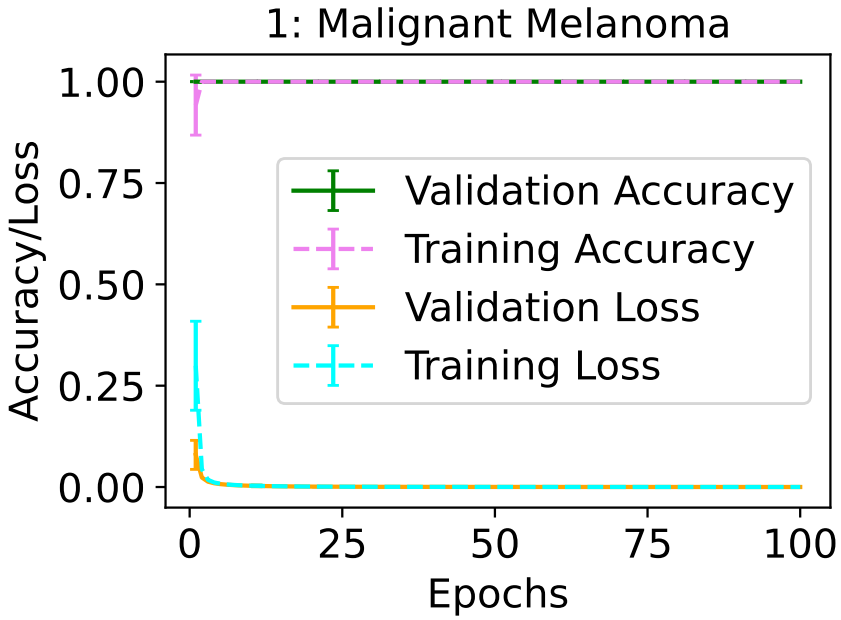

Figure S3: Training and validation performance history, related to STAR Methods.
Training and validation accuracy and loss history for a representative sample during training. The errorbars indicate the standard deviation.

Table S4: Parameters for the active contour algorithm, related to STAR Methods

Time step	β	γ	ω
$0 < t < 30$	0	10	0
$30 \leq t < 500$	1	1	0.5

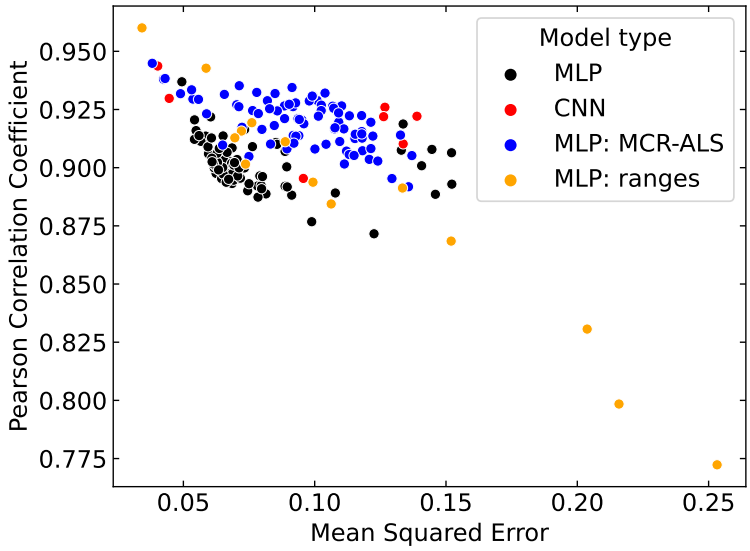

Figure S4: Model validation, related to STAR Methods and Figures 5 & 6.
Each dot represent a model from the grid search (Table S3) in the realative MSE - Pearson Correlation Coefficient space. The color of the dots represent the different kinds of models. The orange orange "MLP: ranges" refer to the MLP trained of various sets of ranges. The top-left corner represents the lowest error and highest correlation.

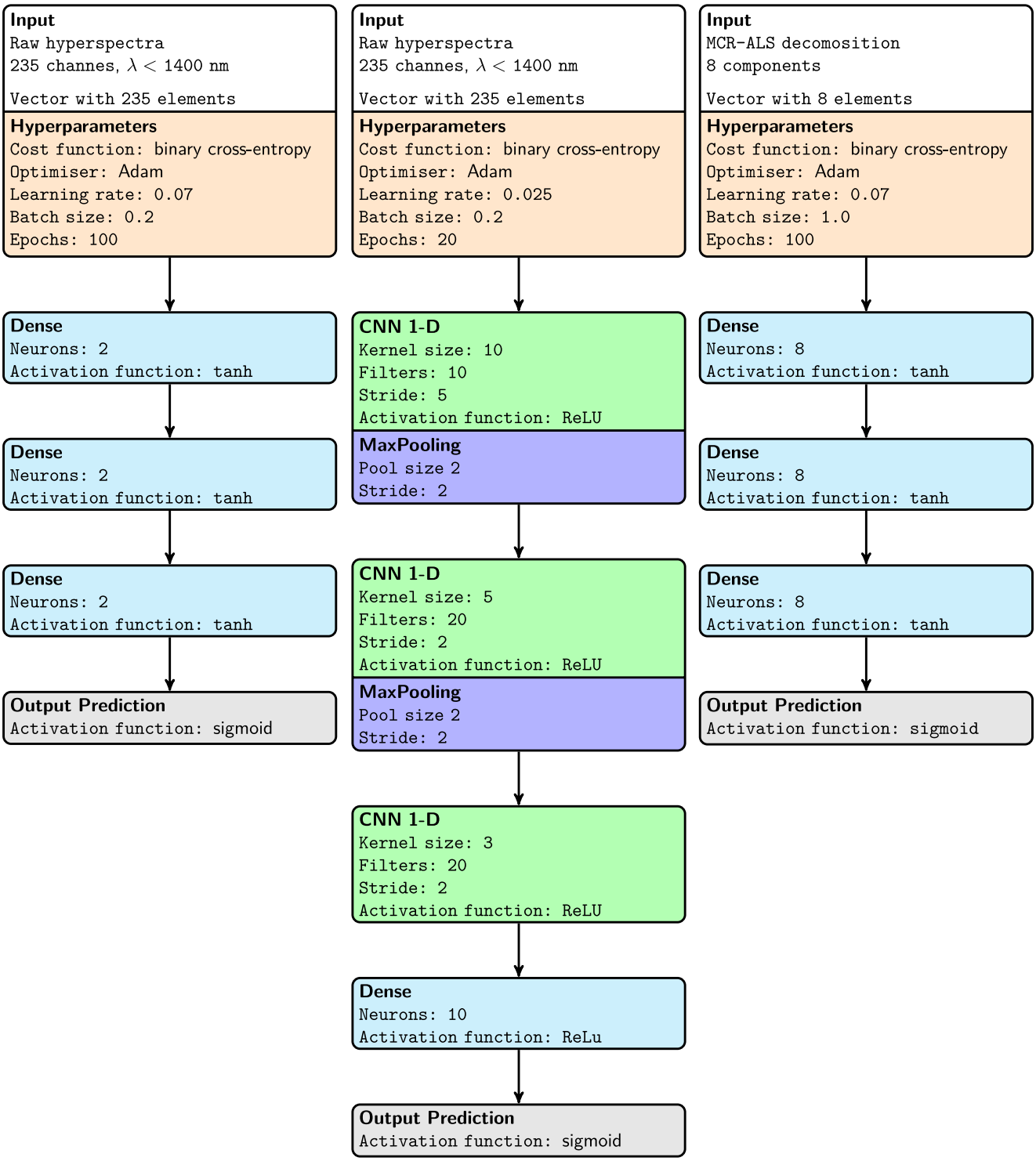

Figure S5: Specifications of final models, related to STAR Methods.
Model specifications for (left) the MLP, (middle) the CNN, and (right) the MLP on the MCR-ALS decomposition.

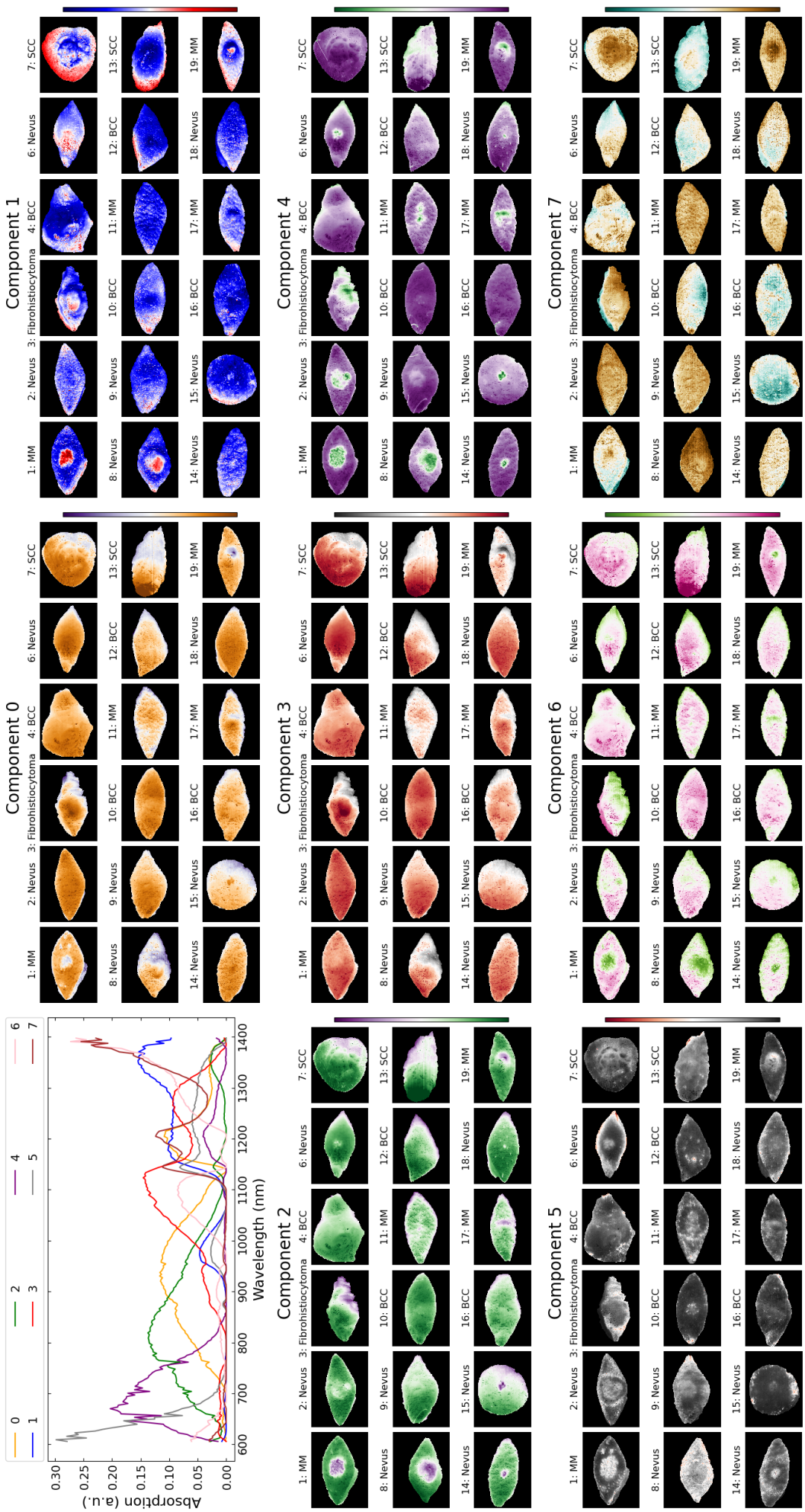

Figure S6: MCR-ALS decomposition for all tumors, related to Figure 2.
The top-left panel shows the 8 spectral components, while the remaining 8 panels show the contribution of each component for all the tumor samples.

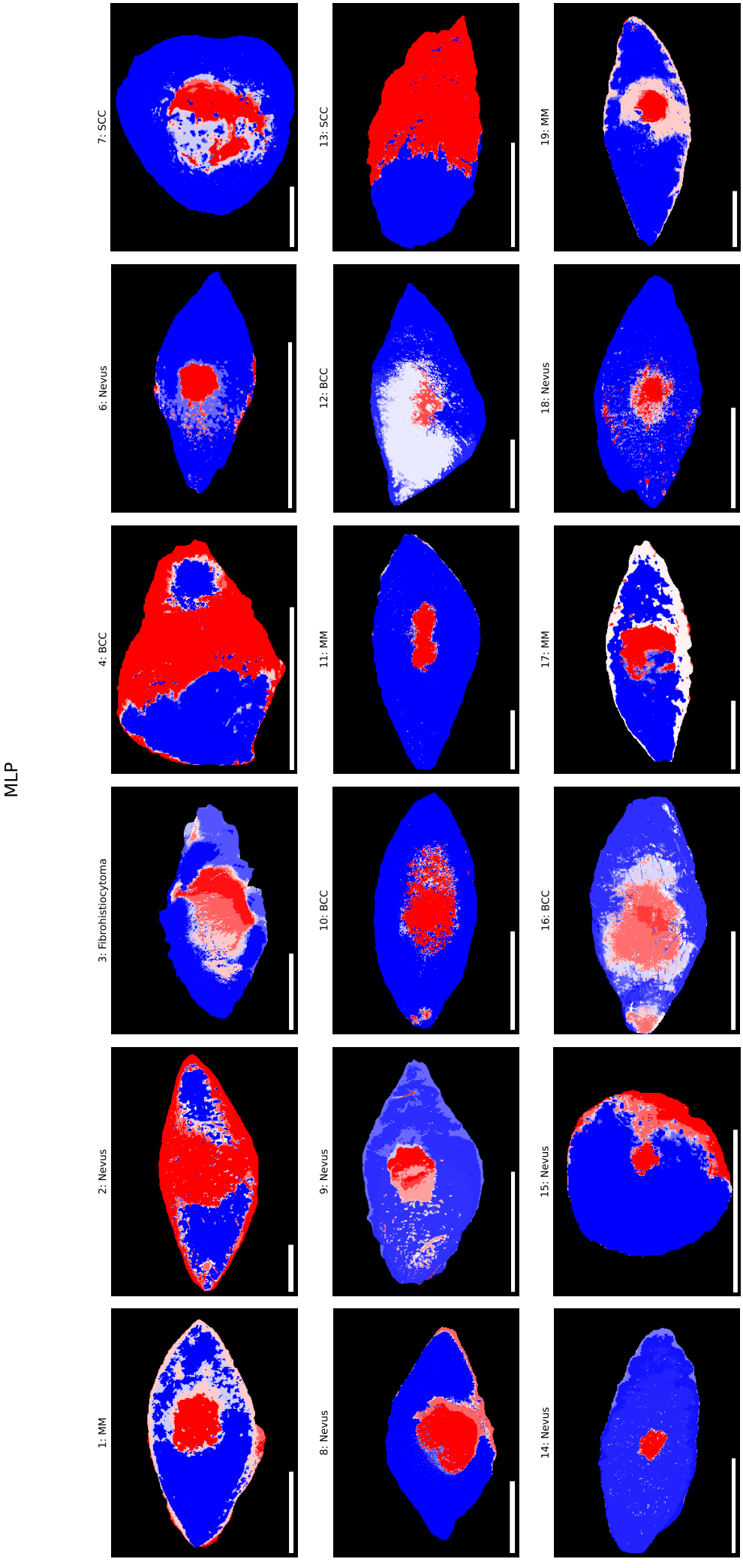

Figure S7: Prediction map for the MLP, related to Figure 3.

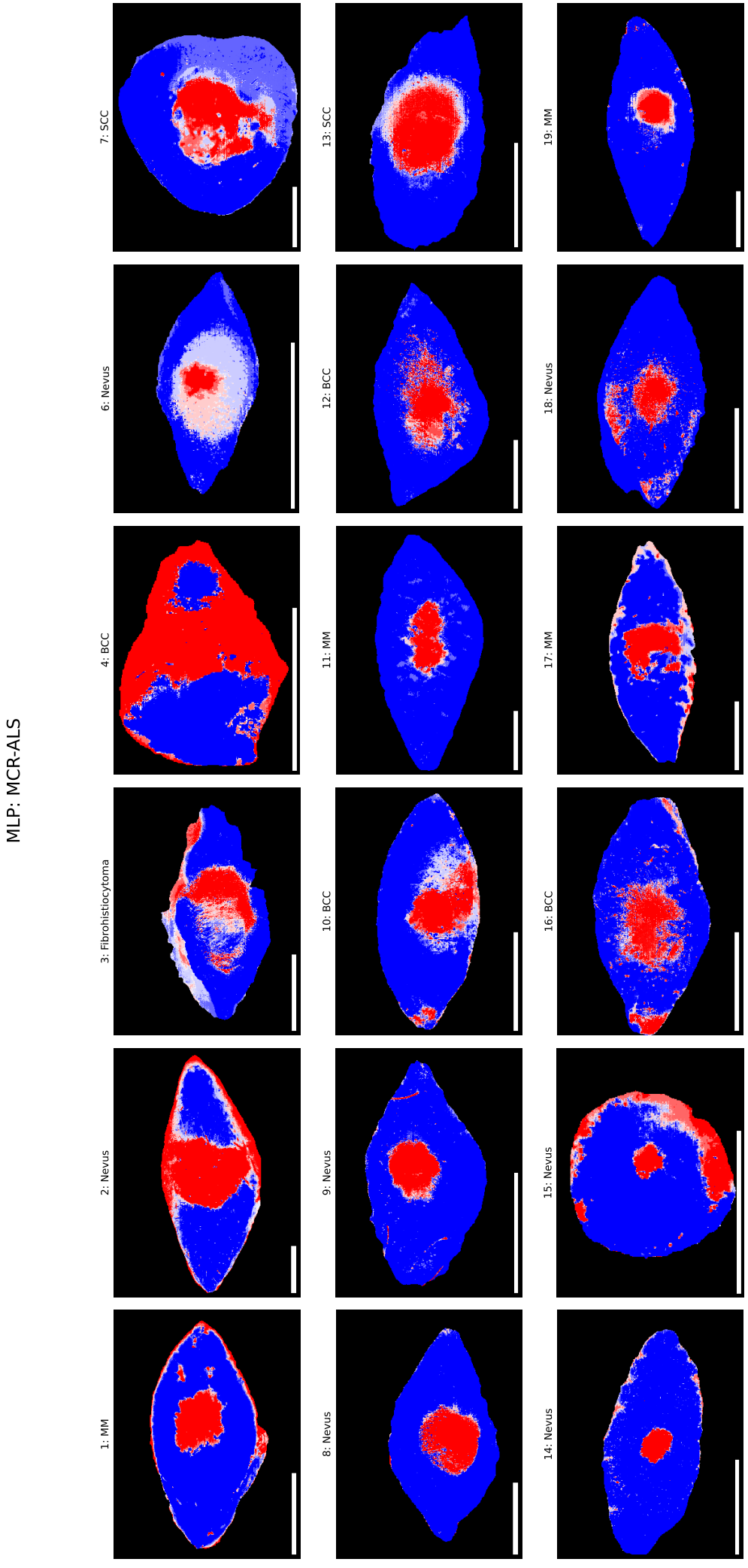

Figure S8: Prediction map for the MLP: MCR-ALS, related to Figure 3.

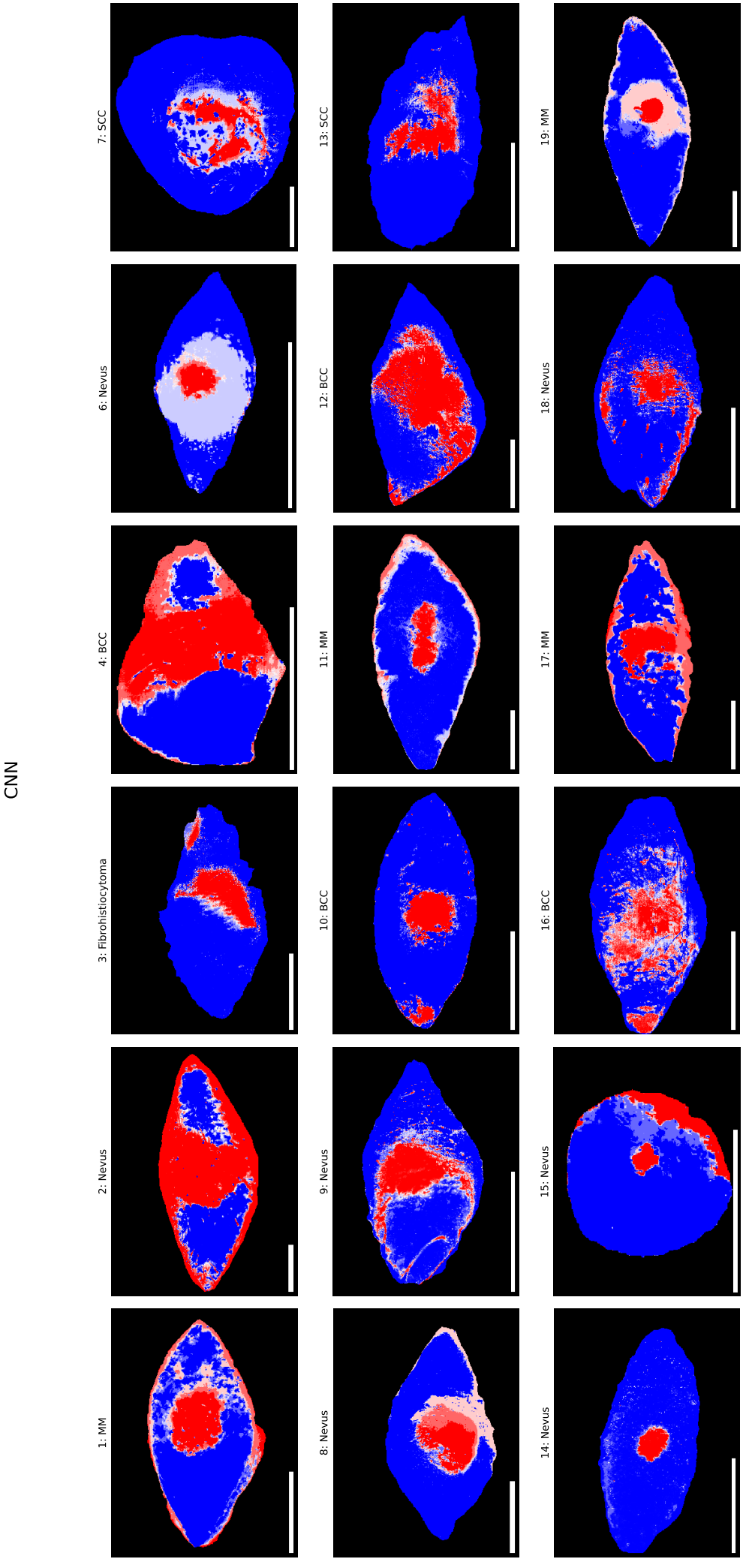

Figure S9: Prediction map for the CNN, related to Figure 3.

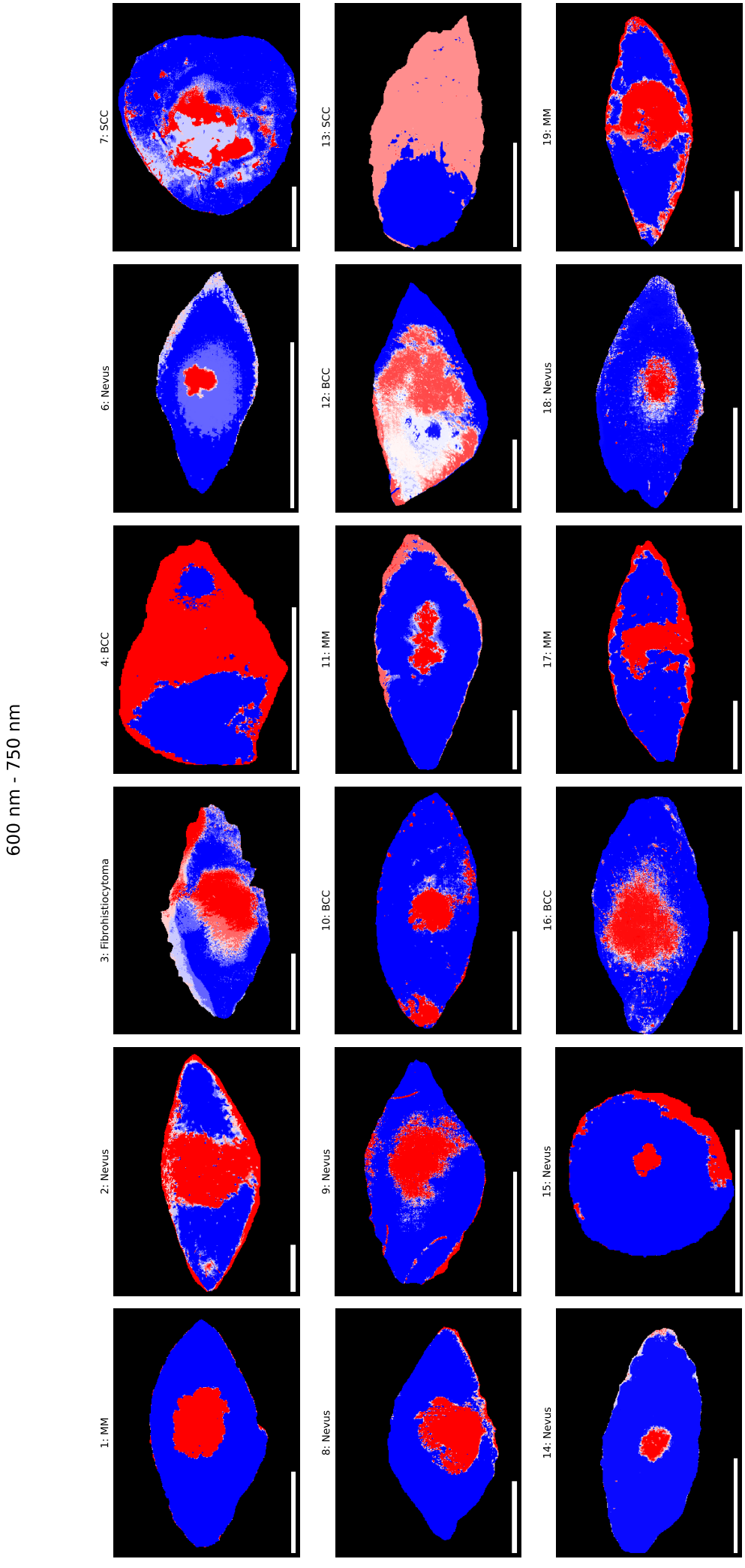

Figure S10: Prediction map for the MLP in the spectral range 600 nm - 750 nm, related to Figure 6.

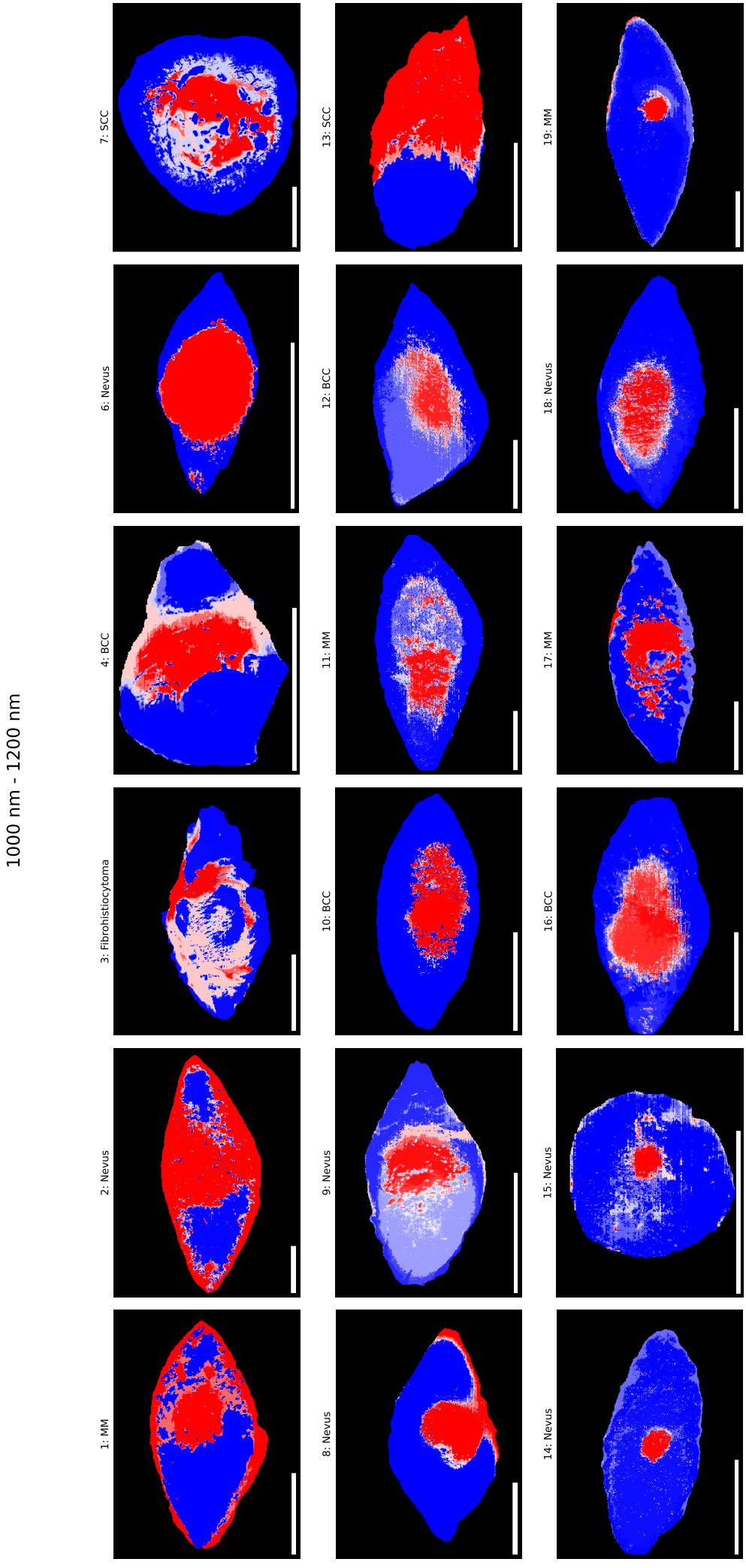

Figure S11: Prediction map for the MLP in the spectral range 1000 nm - 1200 nm, related to Figure 6.

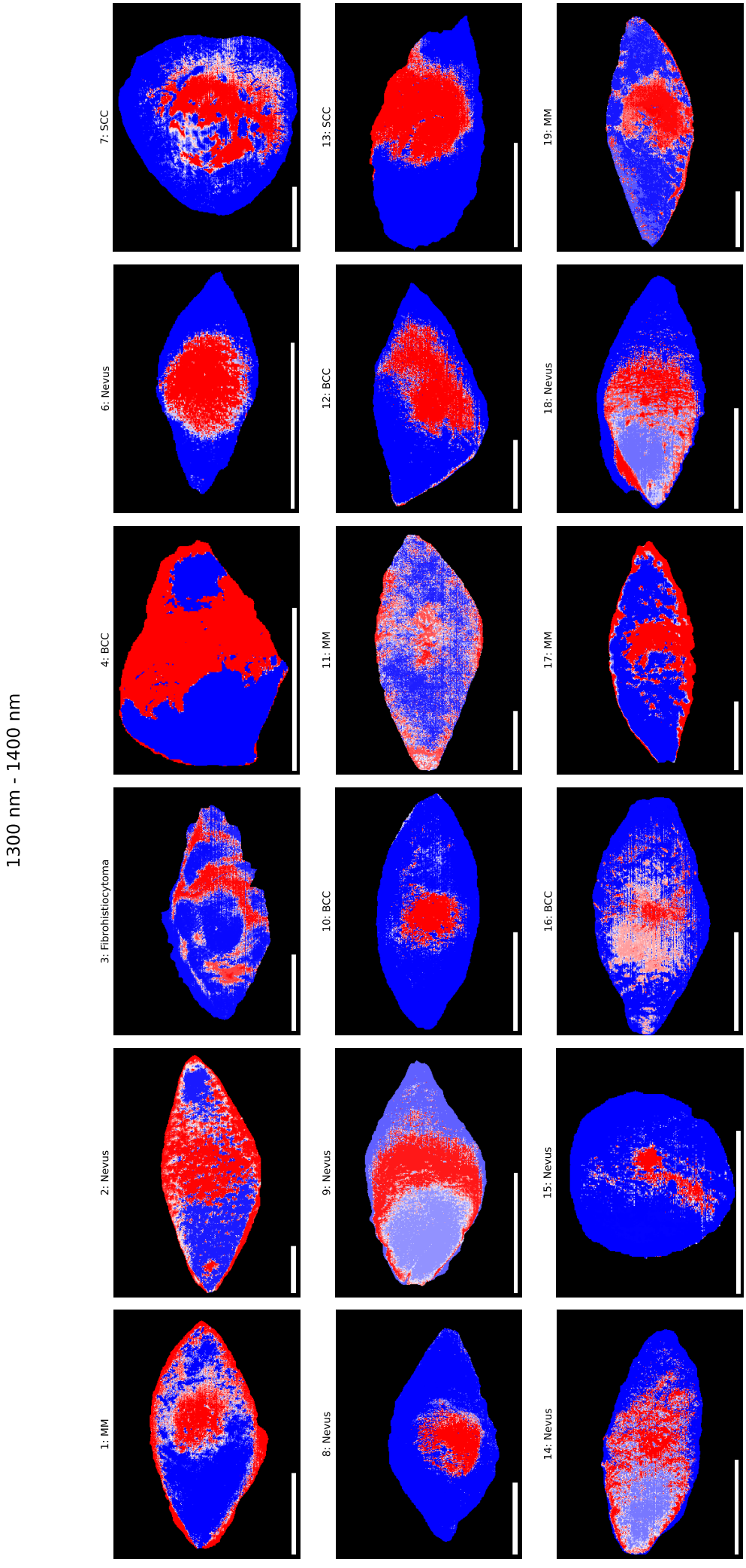

Figure S12: Prediction map for the MLP in the spectral range 1300 nm - 1400 nm, related to Figure 6.

Table S5: Model widths, related to Figures 5 & 6.
Measured histology width and predicted tumor widths in mm for each of the 6 presented models for every tumor.

Lesion no.	Histology	MLP	CNN	MLP: MCR-ALS	MLP: 600-750 nm	MLP: 1000-1200 nm	MLP: 1300-1400 nm
1	6.89	6.90	7.20	5.95	7.10	7.75	7.10
2	15.7	15.8	15.8	15.7	15.9	13.0	13.1
3	6.78	6.90	6.45	6.95	6.25	5.80	6.65
4	7.83	5.65	3.35	3.30	-	4.60	3.45
6	2.39	2.36	2.67	2.77	1.96	5.53	4.33
7	13	16.0	13.4	16.5	14.2	19.8	19.9
8	9.45	9.25	9.25	8.25	9.15	8.65	8.65
9	4.89	3.77	6.13	4.17	5.12	6.08	5.17
10	3.45	5.26	5.01	5.01	4.56	5.61	5.11
11	8.58	4.40	4.35	6.00	5.70	9.45	2.55
12	6.76	4.55	11.6	7.61	12.1	7.40	9.71
13	4.97	7.70	6.49	6.69	8.15	7.85	4.68
14	2.86	2.60	2.91	2.96	2.91	2.86	6.28
15	1.6	1.75	1.70	1.85	1.75	2.15	2.30
16	6.61	6.25	7.10	5.60	5.75	6.10	6.60
17	8.31	8.35	8.20	8.35	8.30	8.35	7.40
18	4.84	4.05	4.60	4.35	3.10	5.05	6.85
19	8.54	9.90	9.80	7.15	10.7	4.60	10.3

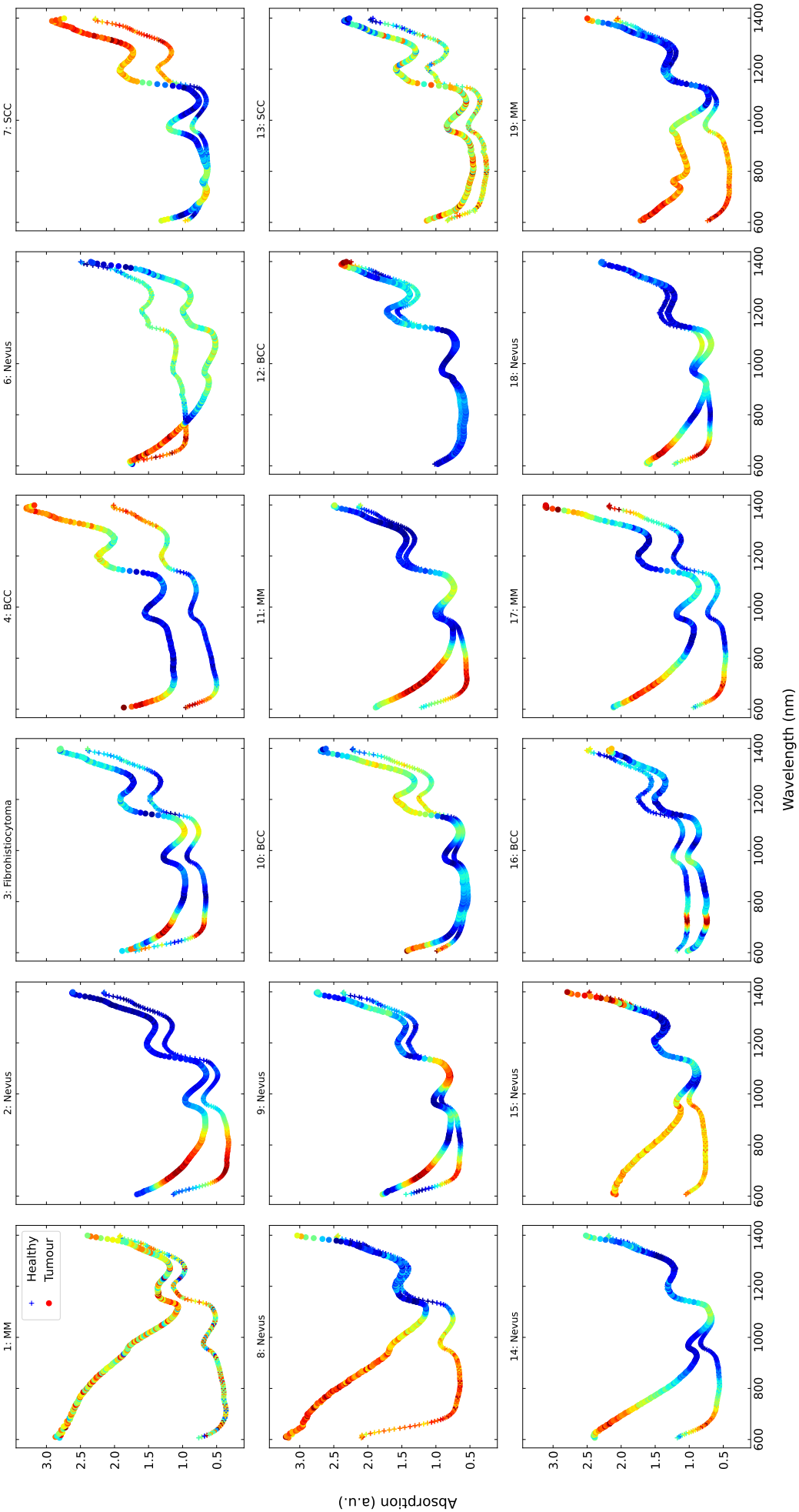

Figure S13: Saliency for all the samples, related to Figure 6.