## Supplementary Materials

### Brain tumour samples and histopathological grading

In Table S1, the measured samples with the corresponded details are listed.

Table S1 Overview of measured samples and the corresponding WHO grading and diagnostics.

|  |  |  |  |
| --- | --- | --- | --- |
|  | WHO grade | Type | Manuscript name |
| Sample set I | I | Fibrous meningioma | Sample A |
| II | Oligodendroglioma | Sample B |
| III | Anaplastic ependymoma | Sample C |
| IV | Glioblastoma | Sample D |
| Sample set II | I | Plexus papilloma | Sample E |
| II | Oligodendroglioma | Sample F |
| III | Anaplastic oligodendroglioma | Sample G |
| IV | Glioblastoma | Sample H  |

### Explanation of the confusion matrix terminology

A confusion matrix describes the performance of the classification model based on the Bayesian discriminant analysis. The main condition for building a confusion matrix is, that the true values, in our case the pathologist’s assumption, must be known. As an example, two confusion matrices are shown in Table S2 for the non-fixed and in Table S3 for the formalin-fixed model of the FTIR spectra of sample set I. The actual value is the neuropathologist’s assumption with information provided in the columns. “Predicted” means how a spectrum is classified by the model, listed in the rows. For example, in Table S2 for sample A 15 spectra are assumed as sample A and 15 are predicted as sample A. Similarly, for sample B, there are 15 spectra classified as sample B but only 13 spectra predicted as sample B. The other two spectra are predicted as sample D by the model. From the 15 assumed sample C spectra there are 14 predicted as sample C and one is predicted as sample D. The same is observable for the assumed sample D spectra: 14 spectra are predicted as sample D but one is classified as sample B by the model. In summary, the diagonal of Table S2 indicates the number of spectra where the brain tumour assignment of the pathologist coincides with the model prediction. Still, the overall accuracy is 93 %. This describes how many spectra are predicted by the model as true compared to the pathologist’s assumption. All spectra from the formalin-fixed cross-sections, which are assumed sample A, sample B and sample C are classified as the pathologist’s assumption in Table S3. Only for sample D there are two spectra which are predicted as sample B, and 13 spectra which are predicted as sample D.

The spatially resolved spectroscopic results do not match in each case the pathological assignment. A tumour is a heterogeneous specimen: in vicinity to the cancerous cells, there are healthy and/or less malignant cells. Therefore, a WHO-grade IV tumour cross-section can contain locations with lower tumour grades which are more sensitive to optical spectroscopy than to classical histopathological characterization. Additionally, a second pathway in tumour genesis is possible for malignant gliomas [87]. In the second case, a glioma arises directly to WHO-grade IV or develops from a low-grade tumour by genetic alterations [87].

Several model quality parameters like sensitivity (Equation 1), specificity (Equation 2), false positive rate (Equation 3) and precision (Equation 4) were calculated from confusion matrices for each sample as following:

$Sensitivity= \frac{True Positives}{True Positives+False Negatives}$ (1)

$Specificity= \frac{True Negatives}{True Negatives+False Positives}$(2)

$False positive rate= \frac{False Positives}{True Negatives+False Positives}$(3)

$Precision= \frac{True Positives}{True Positives+False Positives} $(4)

Due to the identical size of each class, the overall sensitivity, specificity, false positive rate and precision for each model are calculated by the arithmetic mean. For an example see Equation 5:

$Overall sensitivity= \frac{Sensitivity\_{I}+ Sensitivity\_{II}+ Sensitivity\_{III}+ Sensitivity\_{IV}}{4} $(5)

As an example, the model quality parameters are shown in Table S4 for the formalin-fixed model of the FTIR spectra.

Table S Confusion matrix of the FTIR spectra of the non-fixed brain tumour cross-sections of sample set I (accuracy 93 %).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
|  Sample A | 15 | 0 | 0 | 0 |
|  Sample B | 0 | 13 | 0 | 1 |
|  Sample C | 0 | 0 | 14 | 0 |
|  Sample D | 0 | 2 | 1 | 14 |

Table S Confusion matrix of the FTIR spectra of the formalin-fixed brain tumour cross-sections of sample set I (accuracy 97 %).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
|  Sample A | 15 | 0 | 0 | 0 |
|  Sample B | 0 | 15 | 0 | 2 |
|  Sample C | 0 | 0 | 15 | 0 |
|  Sample D | 0 | 0 | 0 | 13 |

Table S4 Model quality parameters for formalin-fixed brain tumours of sample set I (FTIR microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sensitivity[%] | Specificity[%] | False positive rate[%] | Precision[%] |
| Sample A | 100 | 100 | 0 | 100 |
| Sample B | 100 | 96 | 4 | 88 |
| Sample C | 100 | 100 | 0 | 100 |
| Sample D  | 87 | 100 | 0 | 100 |
| Average | 97 | 99 | 1 | 97 |

### Ultraviolet absorption microspectroscopy of non-fixed and formalin-fixed human brain tumours

**Derived mean spectra**

The spectral derivation as data pre-processing highlights small changes within the spectra and thus makes spectral variations easily visible. Based on emphasizing effects, corresponding PCA-models can be improved and Loadings plots are easier to interpret. Exemplarily, pre-processed mean spectra for each sample are shown in Figure S1. Due to the spectral derivation, a comparison of the derived spectrum with the main variance in the Loadings of each PC in the main manuscript (Figure 2) is possible. By this means the shape of the derivated spectra could be compared to the main variance in the Loadings of each PC in the main manuscript (Figure 2).



Figure S1 UV absorption mean spectra: UV absorption mean spectra of (a) non-fixed and (b) formalin-fixed sample set I, (c) non-fixed and (d) formalin-fixed sample set II of brain tumour cross-sections in the range from 230 nm to 380 nm. The spectra are vertically displaced, SNV transformed and derivated (1st Savitzky-Golay).

**2D-scores plot**

3D-scores plots, as shown in the main manuscript (Figure 2), are more demanding to understand than the 2D-scores plots. Nevertheless, in the case of the UV PCA models 3 PCs are required. Thus, PC1 versus PC2 and PC1 versus PC3 is shown in Figure S2.

****

Figure S2 (a) and (e) 2D scores plots PC1 versus PC2 to differentiate between non-fixed tissue samples. (c) and (g) 2D scores plots PC1 versus PC3 for the same PCA models. (b) and (f) 2D scores plots PC1 versus PC2 to differentiate between formalin-fixed tissue samples. (d) and (h) 2D scores plots PC1 versus PC3 for the same PCA models. The representative tumour entities for the tissue samples A to D with different malignancies are fibrous meningioma (sample A), oligodendroglioma (sample B), anaplastic ependymoma (sample C) and glioblastoma (sample D). The representative tumour entities for the tissue samples E to H with different malignancies are plexus papilloma (sample E), oligodendroglioma (sample F), anaplastic oligodendroglioma (sample G) and glioblastoma (sample H).

**Confusion matrices**

For the comparison of non-fixed and formalin-fixed models, each PCA is combined with a Bayesian discriminant analysis with Mahalanobis distance. In Table S5-S8 the confusion matrices from the discriminant analysis are shown.

Table S5 Confusion matrix of non-fixed brain tumours of sample set I with an accuracy of 82 % (UV microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
|  Sample A | 15 | 0 | 1 | 10 |
|  Sample B | 0 | 15 | 0 | 0 |
|  Sample C | 0 | 0 | 14 | 0 |
|  Sample D | 0 | 0 | 0 | 5 |

Table S6 Confusion matrix of formalin-fixed brain tumours of sample set I with an accuracy of 95 % (UV microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
|  Sample A | 15 | 1 | 0 | 0 |
|  Sample B | 0 | 12 | 0 | 0 |
|  Sample C | 0 | 0 | 15 | 0 |
|  Sample D | 0 | 2 | 0 | 15 |

Table S7 Confusion matrix of non-fixed brain tumours of sample set II with an accuracy of 85 % (UV microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
|  Sample E | 10 | 0 | 0 | 0 |
|  Sample F | 0 | 12 | 1 | 0 |
|  Sample G | 0 | 2 | 14 | 0 |
|  Sample H | 5 | 1 | 0 | 15 |

Table S8 Confusion matrix of formalin-fixed brain tumours of sample set II with an accuracy of 75 % (UV microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
| Sample E | 7 | 1 | 0 | 0 |
| Sample F | 8 | 12 | 2 | 0 |
| Sample G | 0 | 1 | 12 | 1 |
| Sample H  | 0 | 1 | 1 | 14 |

### Elastic light scattering microspectroscopy of non-fixed and formalin-fixed human brain tumours

**Derived mean spectra**

Especially for the ELS spectroscopy, it is necessary to examine the pre-processed mean spectra for the interpretation of the PCA-models (Figure S2). Only due to the derivation, the finely corrugated structures caused by interferences are visible and could be compared to the corresponding Loading plots (Figure 3).



Figure S2 ELS mean spectra: ELS mean spectra of (a) non-fixed and (b) formalin-fixed for sample set I, (c) non-fixed and (d) formalin-fixed for sample set II of brain tumour cross-sections in the range from 380 nm to 700 nm. The spectra are vertically displaced, area normalized and derived (1st Savitzky-Golay).

**Confusion matrices**

For the comparison of non-fixed and formalin-fixed models, each PCA is combined with a Bayesian discriminant analysis with Mahalanobis distance. In Table S9-S12 the confusion matrices from the discriminant analysis are shown.

Table S9 Confusion matrix of non-fixed brain tumours of sample set I with an accuracy of 88 % (ELS microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
|  Sample A | 13 | 0 | 3 | 0 |
|  Sample B | 0 | 14 | 0 | 1 |
|  Sample C | 2 | 0 | 12 | 0 |
|  Sample D | 0 | 1 | 0 | 14 |

Table S10 Confusion matrix of formalin-fixed brain tumours of sample set I with an accuracy of 98 % (ELS microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
|  Sample A | 15 | 0 | 0 | 0 |
|  Sample B | 0 | 15 | 0 | 1 |
|  Sample C | 0 | 0 | 15 | 0 |
|  Sample D | 0 | 0 | 0 | 14 |

Table S11 Confusion matrix of non-fixed brain tumours of sample set II with an accuracy of 72 % (ELS microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
| Sample E | 9 | 0 | 1 | 4 |
| Sample F | 0 | 11 | 2 | 0 |
| Sample G | 5 | 4 | 12 | 0 |
| Sample H  | 1 | 0 | 0 | 11 |

Table S12 Confusion matrix of formalin-fixed brain tumours of sample set II with an accuracy of 98 % (ELS microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
| Sample E | 15 | 0 | 1 | 0 |
| Sample F | 0 | 15 | 0 | 0 |
| Sample G | 0 | 0 | 14 | 0 |
| Sample H  | 0 | 0 | 0 | 15 |

### Fourier-transform infrared microspectroscopy of non-fixed and formalin-fixed human brain tumours

**Derived mean spectra**

Connected to an easier understanding and interpretation of the corresponding PCA-models and especially to the Loadingplots the pre-processed mean spectra for each sample are shown in Figure S3. By this means the shape of the derivated spectra could be compared to the main variance in the Loadings of each PC (Figure 4).



Figure S3 FTIR mean spectra: FTIR mean spectra of (a) non-fixed and (b) formalin-fixed for sample set I, (c) non-fixed and (d) formalin-fixed for sample set II brain tumour cross-sections in the range from 4000 cm-1 to 700 cm-1. The wavenumbers between 2410 cm-1 and 2240 cm-1 are not displayed. The spectra are vertically displaced, unit vector normalized and derivated (1st Savitzky-Golay).

**Confusion matrices**

For the comparison of non-fixed and formalin-fixed models, each PCA is combined with a Bayesian discriminant analysis with Mahalanobis distance. In Table S13-S16 the confusion matrices from the discriminant analysis are shown.

Table S13 Confusion matrix of non-fixed brain tumours of sample set I with an accuracy of 93 % (FTIR microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
| Sample A | 15 | 0 | 0 | 0 |
| Sample B | 0 | 13 | 0 | 1 |
| Sample C | 0 | 0 | 14 | 0 |
| Sample D  | 0 | 2 | 1 | 14 |

Table S14 Confusion matrix of formalin-fixed brain tumours of sample set I with an accuracy of 97 % (FTIR microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
| Sample A | 15 | 0 | 0 | 0 |
| Sample B | 0 | 15 | 0 | 2 |
| Sample C | 0 | 0 | 15 | 0 |
| Sample D  | 0 | 0 | 0 | 13 |

Table S15 Confusion matrix of non-fixed brain tumours of sample set II with an accuracy of 87 % (FTIR microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
| Sample E | 15 | 0 | 2 | 0 |
| Sample F | 0 | 14 | 4 | 0 |
| Sample G | 0 | 1 | 8 | 0 |
| Sample H  | 0 | 0 | 1 | 15 |

Table S16 Confusion matrix of formalin-fixed brain tumours of sample set II with an accuracy of 93 % (FTIR microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
| Sample E | 15 | 0 | 0 | 0 |
| Sample F | 0 | 14 | 0 | 1 |
| Sample G | 0 | 0 | 13 | 0 |
| Sample H  | 0 | 1 | 2 | 14 |

### Combined datasets from non-fixed and formalin-fixed human brain tumours

Additionally, we build multivariate models for each spectroscopic method with all data acquired of non-fixed and formalin-fixed cross-sections. The same pre-processing of the spectra is used as described before for the single multivariate models of non-fixed and formalin-fixed cross sections. First, a PCA is calculated followed by a Bayesian discriminant analysis based on the calculated PCs. For the calculation of the distances, the same algorithms, Mahalanobis (UV and ELS) and Euclidean distance (FTIR), are used like in the single models before.

**Ultraviolet absorption microspectroscopy**

For the calculation of the combined PCA, the same UV-spectra for each sample set with the same pre-processing (SNV and 1st Savitzky-Golay) are used as described for the UV-models before. The PCA-models are shown in Figure S4.



Figure S4PCA of non-fixed and formalin-fixed brain tumour cross-sections of UV absorption spectra: UV spectra from 230 nm to 380 nm. (a) Scores plot of sample set I with PC1 (77 % explained variance) versus PC2 (13 % explained variance) and (b) scores plot of sample set I with PC3 (5 % explained variance) versus PC4 (2 % explained variance). Tissue sample A (blue), B (yellow), C (red) and D (black) are separated from each other (PC1 to PC4). The non-fixed spectra are marked by an unfilled circle, the formalin-fixed ones by a filled square. (c) Corresponding loadings plots of PC1 (black), PC2 (red), PC3 (blue) and PC4 (yellow) for the combined model of sample set I. (d) Scores plot of sample set II with PC1 (66 % explained variance) versus PC2 (20 % explained variance) and (e) scores plot of sample set II with PC3 (4 % explained variance) versus PC4 (2 % explained variance). Tissue sample E (blue), F (yellow), G (red) and H (black) are separated from each other (PC1 to PC4). (f) Corresponding loadings plots of PC1 (black), PC2 (red), PC3 (blue) and PC4 (yellow) for the combined model of sample set II.

For the comparison of the models, each PCA is combined with a Bayesian discriminant analysis with Mahalanobis distance. In Table S17-S18 the confusion matrices from the discriminant analysis are show{Amharref, 2006 #132}n.

Table S17 Confusion matrix of non-fixed and formalin-fixed brain tumours of sample set I with an accuracy of 91 % (UV microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
| Sample A | 30 | 0 | 1 | 2 |
| Sample B | 0 | 29 | 0 | 2 |
| Sample C | 0 | 0 | 25 | 1 |
| Sample D  | 0 | 1 | 4 | 25 |

Table S18 Confusion matrix of non-fixed and formalin-fixed brain tumours of sample set II with an accuracy of 84 % (UV microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
| Sample E | 27 | 2 | 0 | 0 |
| Sample F | 3 | 20 | 2 | 0 |
| Sample G | 0 | 6 | 25 | 1 |
| Sample H  | 0 | 2 | 3 | 29 |

**Elastic light scattering microspectroscopy**

For the calculation of the combined PCA, the same ELS-spectra for each sample set with the same pre-processing (area normalization and 1st Savitzky-Golay) are used as described for the ELS-models before. The PCA-models are shown in Figure S5.



Figure S5 PCA of non-fixed and formalin-fixed brain tumour cross-sections of ELS spectra: ELS spectra from 380 nm to 700 nm. The ELS spectra of sample set I build a model with (a) scores plot PC1 (52 % explained variance) versus PC2 (10 % explained variance) and (b) scores plot PC1 versus PC3 (5 % explained variance). Tissue sample A (blue), B (yellow), C (red) and D (black) are demarcated from each other (PC1 to PC3). The non-fixed spectra are marked by an unfilled circle, the formalin-fixed ones by a filled square. (c) Corresponding loadings plots of PC1 (black), PC2 (red) and PC3 (blue) for the combined model of sample set I. The ELS spectra of sample set II build a model with (d) scores plot PC1 (79 % explained variance) versus PC2 (4 % explained variance) and (e) scores plot PC3 (2 % explained variance) versus PC4 (2 % explained variance). Tissue sample E (blue), F (yellow), G (red) and H (black) are demarcated from each other (PC1 to PC5). (f) Corresponding loadings plots of PC1 (black), PC2 (red), PC3 (blue), PC4 (yellow) and PC5 (light blue) for the combined model of sample set II.

For the comparison of the models, each PCA is combined with a Bayesian discriminant analysis with Mahalanobis distance. In Table S19-S20 the confusion matrices from the discriminant analysis are shown.

Table S19 Confusion matrix of non-fixed and formalin-fixed brain tumours of sample set I with an accuracy of 82 % (ELS microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
| Sample A | 27 | 0 | 5 | 0 |
| Sample B | 0 | 26 | 0 | 5 |
| Sample C | 2 | 0 | 24 | 4 |
| Sample D  | 1 | 4 | 1 | 21 |

Table S20 Confusion matrix of non-fixed and fixed brain tumours of sample set II with an accuracy of 79 % (ELS microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
| Sample E | 22 | 0 | 1 | 0 |
| Sample F | 0 | 21 | 4 | 0 |
| Sample G | 3 | 9 | 25 | 3 |
| Sample H  | 5 | 0 | 0 | 27 |

**Fourier-transform infrared microspectroscopy**

For the calculation of the combined PCA, the same FTIR-spectra for each sample set with the same pre-processing (unit vector normalization and 1st Savitzky-Golay) are used as described for the FTIR-models before. The PCA-models are shown in Figure S6.



Figure S6 PCA of non-fixed and formalin-fixed brain tumour cross-sections of FTIR spectra: FTIR spectra from 4000 cm-1 to 700 cm-1. (a) Scores plot of sample set I with PC1 (65 % explained variance) versus PC2 (11 % explained variance) and (b) scores plot of sample set I with PC3 (7 % explained variance) versus PC4 (3 % explained variance). Tissue sample A (blue), B (yellow), C (red) and D (black) are separated from each other (PC1 to PC4). The non-fixed sample spectra are marked by an unfilled circle, the formalin-fixed ones by a filled square. (c) Corresponding loadings plots of PC1 to PC4 for the combined model of sample set I. (d) Scores plot of sample set II with PC1 (79 % explained variance) versus PC2 (16 % explained variance) and (e) scores plot of sample set II with PC3 (2 % explained variance) versus PC4 (1 % explained variance). Tissue sample E (blue), F (yellow), G (red) and H (black) are separated from each other (PC1 to PC4). (f) Corresponding loadings plots of PC1 to PC4 for the combined model of sample set II. The wavenumbers between 2410 cm-1 and 2240 cm-1 are not displayed.

For the comparison of the models, each PCA is combined with a Bayesian discriminant analysis with Euclidian distance. In Table S21-S22 the confusion matrices from the discriminant analysis are shown.

Table S21 Confusion matrix of non-fixed and formalin-fixed brain tumours of sample set I with an accuracy of 92 % (FTIR microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample A | Sample B | Sample C | Sample D |
| Sample A | 30 | 0 | 0 | 0 |
| Sample B | 0 | 27 | 0 | 5 |
| Sample C | 0 | 0 | 30 | 2 |
| Sample D  | 0 | 3 | 0 | 23 |

Table S22 Confusion matrix of non-fixed and formalin-fixed brain tumours of sample set II with an accuracy of 87 % (FTIR microspectroscopy).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Actual Predicted | Sample E | Sample F | Sample G | Sample H |
| Sample E | 30 | 0 | 1 | 0 |
| Sample F | 0 | 30 | 1 | 0 |
| Sample G | 0 | 0 | 15 | 1 |
| Sample H  | 0 | 0 | 13 | 29 |