Supplement material

Automatic segmentation for image dataset preparation

- 21 WSIs with CRC
- 5 WSIs with collitis ulcerosa
- 1 WSI with HGIEN
- 1 WSI with LGIEN
- 1 WSI with healthy tissue

Manual WSI annotation and tiling

Training, validation and testing 13 class[1] semantic segmentation U-Net model

A

Data preparation for classification

212,849 tiles from Graz collective

Excluding tiles without tissue using WSI-tools

Determination of cancerous tissue for each tile with U-Net

Separation of tiles into groups according to their amount of tumor area

Tumor boarder area with 50 - 75% of the tile covered by CRC

Tumor central area with at least 95% of the tile covered by CRC

Colour normalisation using GAN[2]

B

Classification

Border tumor area tiles

All tiles

Central tumor area tiles

Training one ResNet and MIL model for each input dataset

Nodal status

C

Figure S1: Overview of the complete workflow

A) 29 whole slide images (WSI) of colorectal tissue samples were included. 21 with colorectal carcinoma (CRC), seven other pathologies and one healthy tissue were manually annotated for 13 classes[1]. After further data preprocessing we used the WSI tiles to train, validate and test the segmentation U-Net model.

B) With the U-Net we segmented the tiled WSIs from the Graz collective to further determine the amount of tumor tissue in each tile. We split the collective into two groups for tumor border tiles and central tumor tiles.

C) Residual neural network (ResNet) and MIL models were trained to predict outcome variables for each group.

(HGIEN: high-grade intraepithelial neoplasia, LGIEN: low-grade intraepithelial neoplasia, GAN: gernerative adversarial network)

[1]: for class labels and distribution see table S2

[2]: code for the GAN provided and implemented by Runz et al.
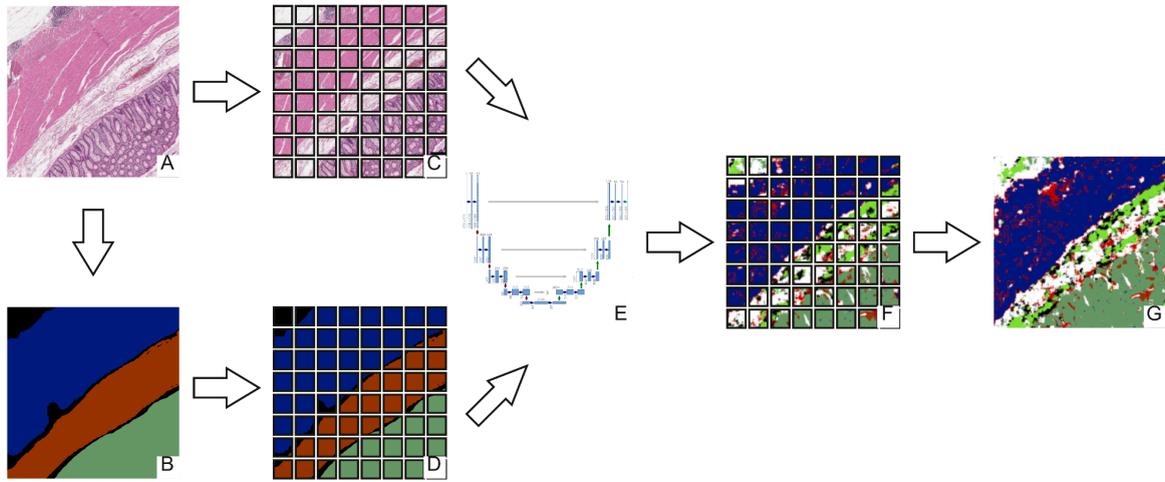
Figure S2: Workflow for the segmentation of a HE-stained histological images
The original HE-image A) was annotated by hand in accordance with the underlying classes of tissue. In the annotated image B) every class was represented by its own color. Color coding is generic in this workflow and not in accordance with table S1. The HE-image and the annotated image were tiled into multiple equal-sized blocks C, D). Subsequently, the pairs annotation and tile were used to train a segmentation model (U-Net) E). After training the network was able to produce annotated tiles on the basis of input HE-tiles F) with an accuracy of 72%. These output-tiles could be reassembled to form a completely annotated version of the original HE-image G).

Table S1: Palette for segmentation labels

13 classes in total were annotated each with a unique color coding. The representation of a label area as a fraction of the whole dataset is given in the last column

| color	label	fraction [%]
	background	18.5
	tumor	33.1
	mucosa	6.6
	submucosa	2.8
	muscularis	12.8
	fatty tissue	12.6
	stromal reaction	7.1
	artefact	0.3
	inflammation	1.1
	necrosis	0.5
	else	1.4
	mucus	0.8
	non tumor pathology	2.5
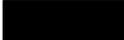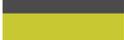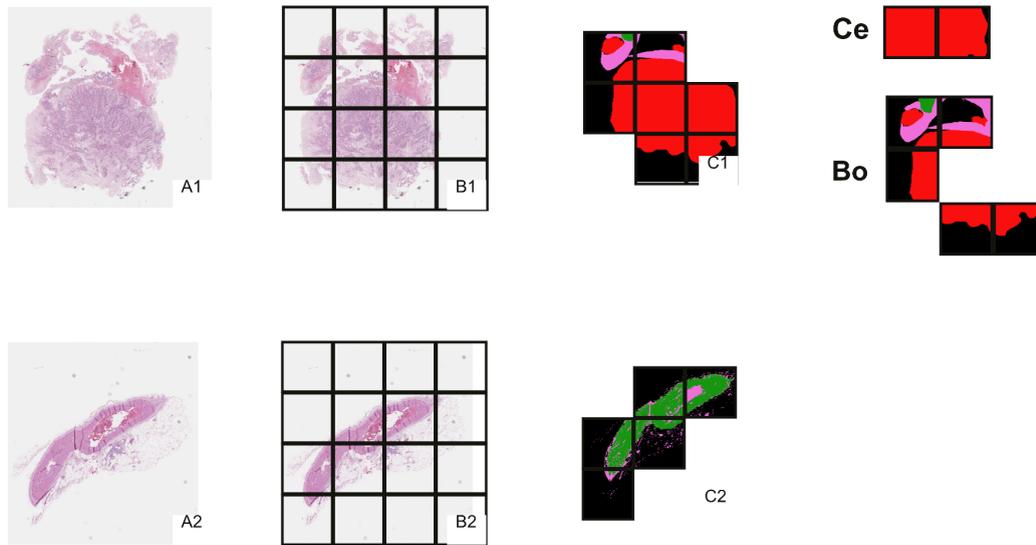
Figure S3: Workflow for the classification of HE-images
The WSIs A1, A2 from the Graz collective were tiled B1, B2). Tiles containing no tissue at all were discarded. Based on the segmentation result of the previously trained U-Net the pixels classified as tumor were counted for each tile C1, C2). Tiles with ≥ 95\% tumor were assigned to the input group tumor central Ce) and tiles with with 50\% - 75\% were assigned to the input group tumor border Bo).
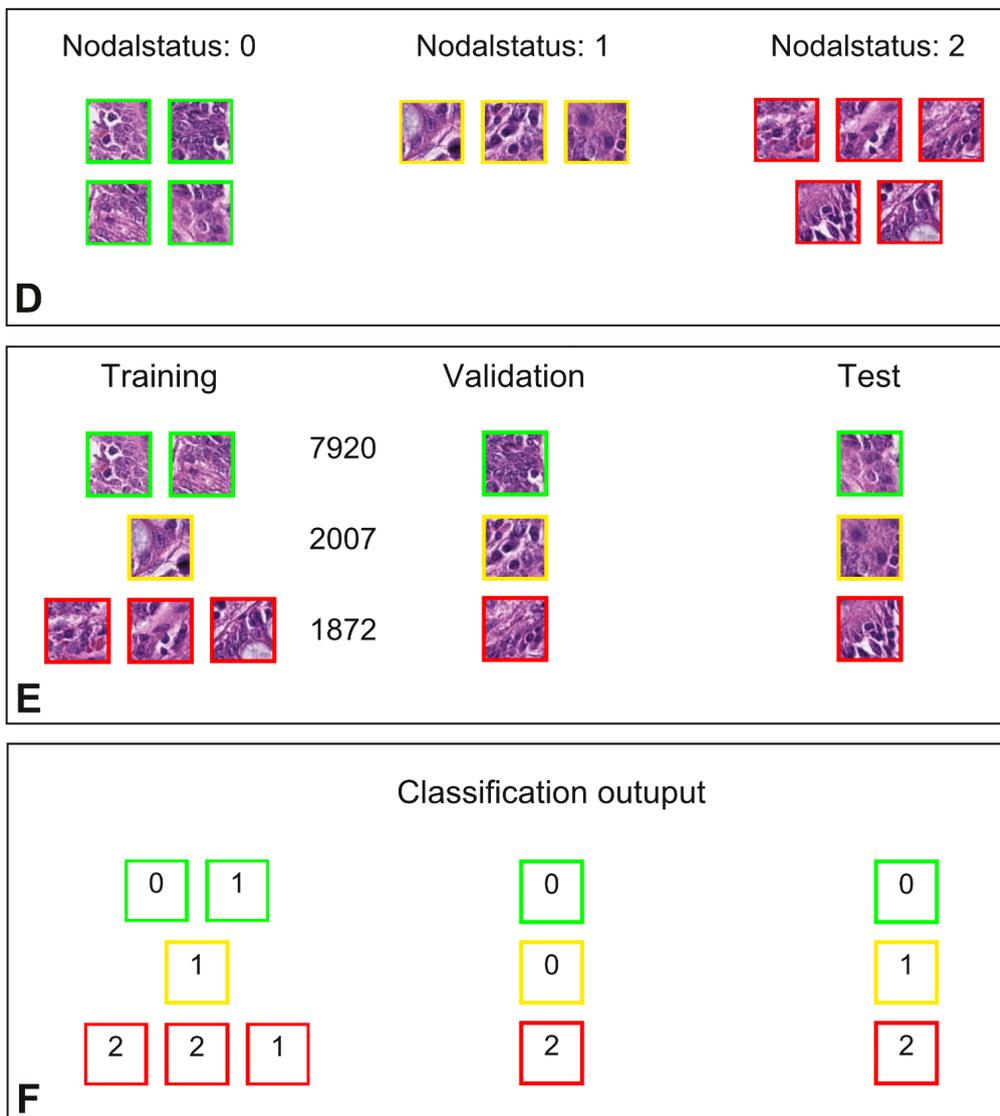
Figure S4github: Data set preparation for the classification of HE-images
After selection of suitable tiles the tiles underwent a special sorting for each classification task. In this example the procedure for a classification for nodal status based on central tumor areas (Ce2No) is shown. D) At first the tiles were put into groups with the same nodal status. E) Afterwards the tiles of every status were randomly and without any regards to their actual case spread into groups for training, validation and testing of the model. The numbers state the amount of tiles in each phase. With nodal status 0 there were 7920 tiles in the training phase, 1584 in the validation phase and 1056 for testing. F) For each tile the residual neural network (ResNet) predicts the nodal status. Validation was performed during training after each training epoch and testing was performed after training and validation were completed.