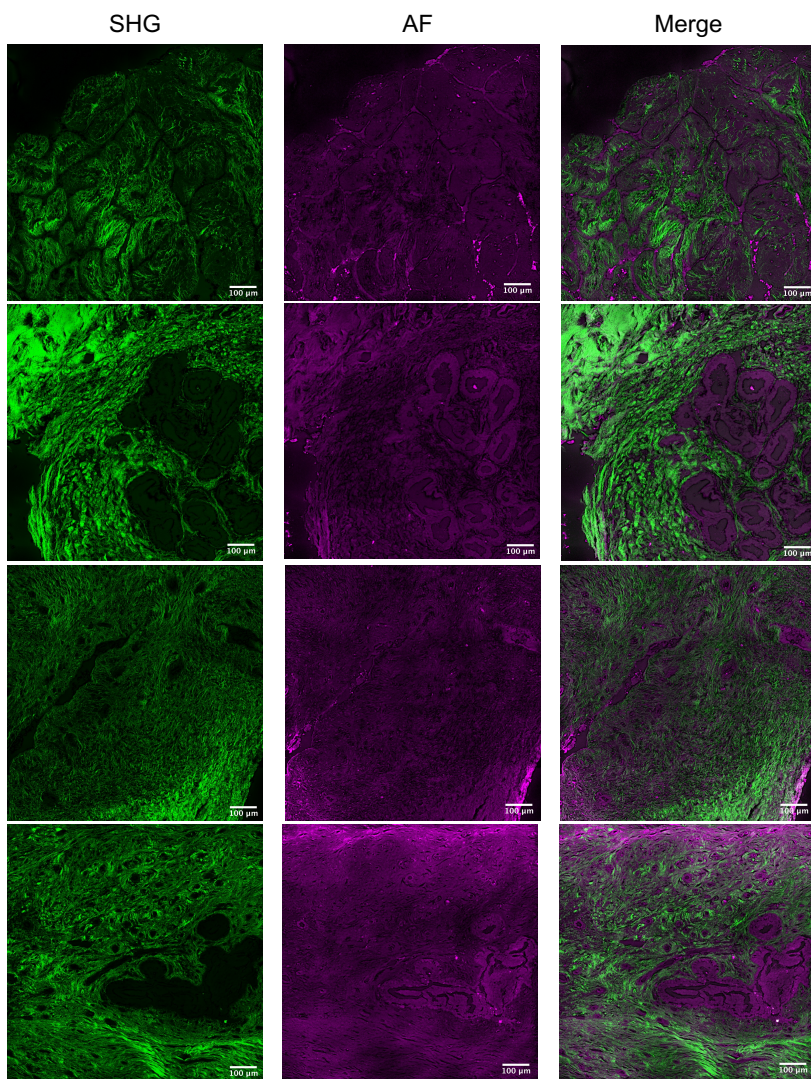
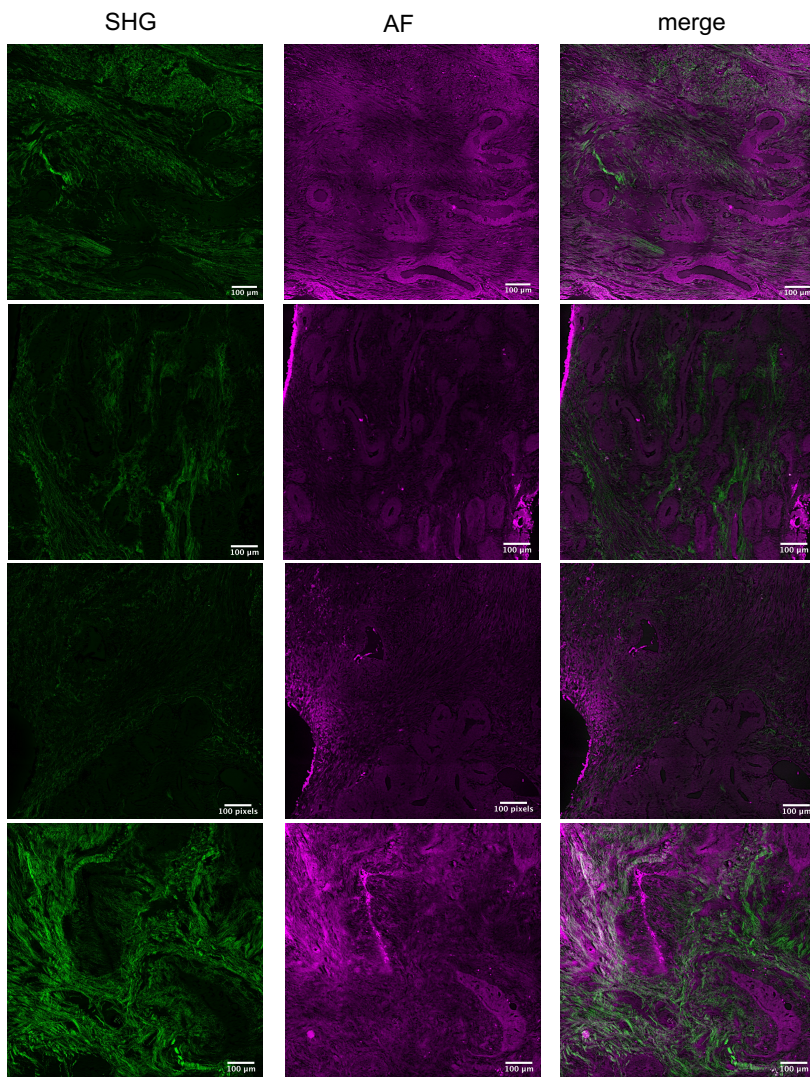


Supplementary Figure S1. Whole view images of needle biopsied tissue sections with multi-photon microscopy and HE stained samples. Several fields indicated as white frames were $1\text{ mm} \times 1\text{ mm}$ FoV images, and were chosen without overlap and almost the whole region was filled when combining those together. Scale bar, 1 mm.



Supplementary Figure S2. Multi-photon microscopy images of tissue sections for FA lesions. Each image was from different patient samples. Scale bar, 100 µm.



Supplementary Figure S3. Multi-photon microscopy images of tissue sections for PT lesions. Each image was from different patients samples. Scale bar, 100 μm .

A

Original

Difference

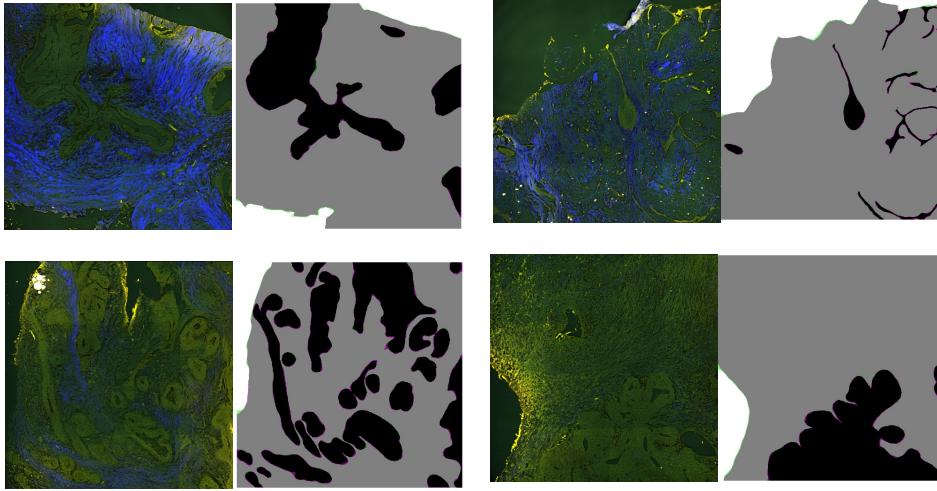
Original

Difference

FA

PT

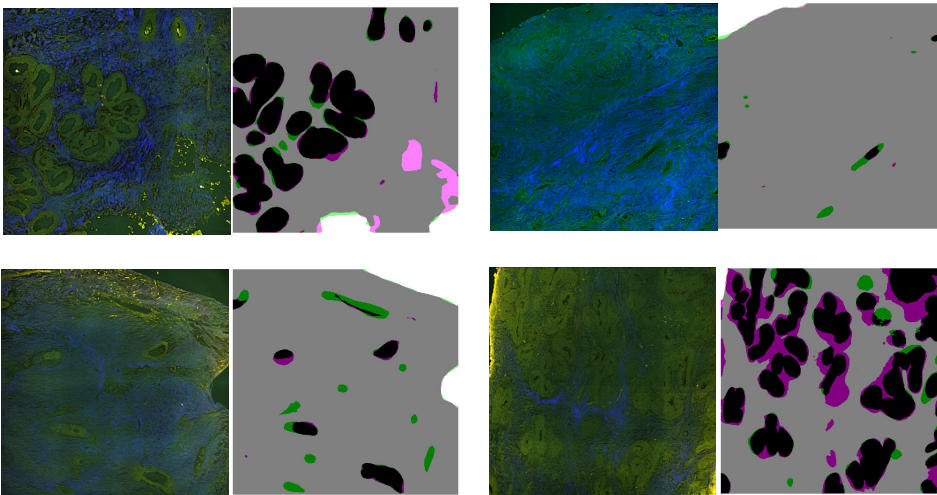
Training images

**B**

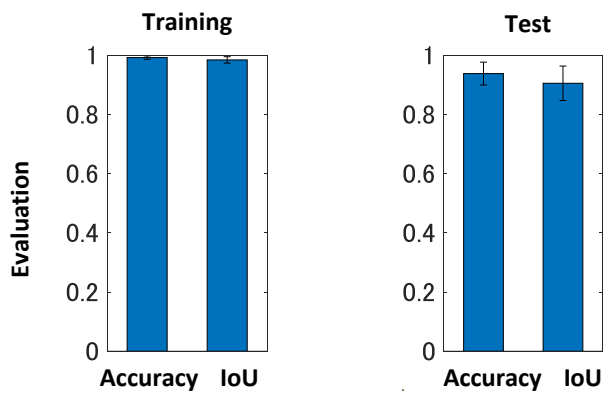
FA

PT

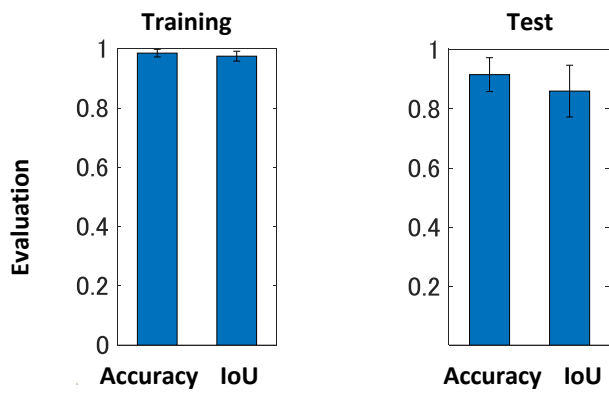
Test images



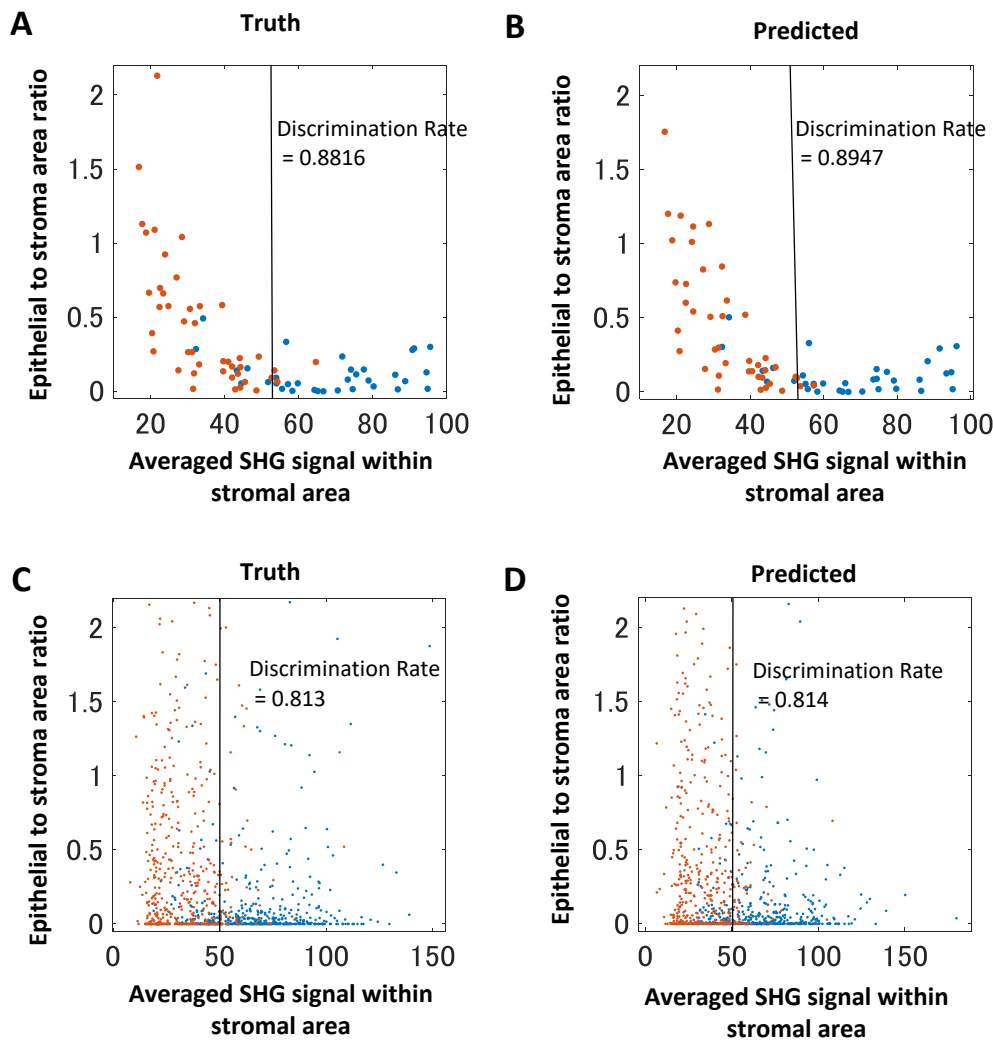
Supplementary Figure S4. Results of image segmentation by a deep learning based framework, SegNet. (A) Results of training image sets. Original multi-photon microscopy images and difference images between ground-truth and predicted images were shown for both FA and PT images. Differences in images indicated FN areas as magenta and FP areas as green. (B) Results of test image sets.



Supplementary Figure S5. Numerical evaluation of the segmentation results for the case where another 50% randomly selected images were used for training. The total accuracy between the ground-truth and predicted images, and the weighted IoU are shown for training and test data sets. These scores are evaluated for each image in training and test cases, and statistics such as mean and standard deviation are calculated. The bar denotes average; the error bar denotes standard deviation over the data calculated from image sets.



Supplementary Figure S6. Numerical evaluation of the segmentation results for the case where 20% randomly selected images were used for training. The total accuracy between the ground-truth and predicted images, and the weighted IoU are shown for training and test data sets. These scores are evaluated for each image in training and test cases, and statistics such as mean and standard deviation are calculated. The bar denotes average; the error bar denotes standard deviation over the data calculated from image sets.



Supplementary Figure S7. Scatter plots and discrimination analysis of the two quantification scores. (A,B) Scatter plot of the ground-truth (A) and the predicted (B) data for original (512×512 px) image set. Blue and red circles denote FA and PT data, respectively. The lines represent the fitted lines for discrimination between FA and PT scores. (C,D) Scatter plot of the ground-truth (A) and the predicted (B) data for small size (128×128 px) image set. The small size image sets are generated by dividing the original (512×512 px) image into 4×4 blocks. From these images, 500 images are randomly selected (excluding the stromal area = 0 block) for both FA and PT. Score calculation is performed for each image. Blue and red circles denote FA and PT data, respectively. The lines represent the fitted lines for discrimination between FA and PT scores.