

**Figure S4** Non-linear relationship between the total number of samples used for training and sensitivity at 85% specificity for colorectal cancer detection. The method selected by colorectal cancer analyses was trained again with k-fold except the number of training samples per fold were downsampled. The lower numbers are comparable to those available for balanced k-batch and were used to investigate decreased classifier performance due to smaller sample sizes in training.