This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material

Table S1. Sample size

	TP0	TP1	TP2
Control	23	23	12
Treatment A	18	17	18
Treatment B	22	21	10

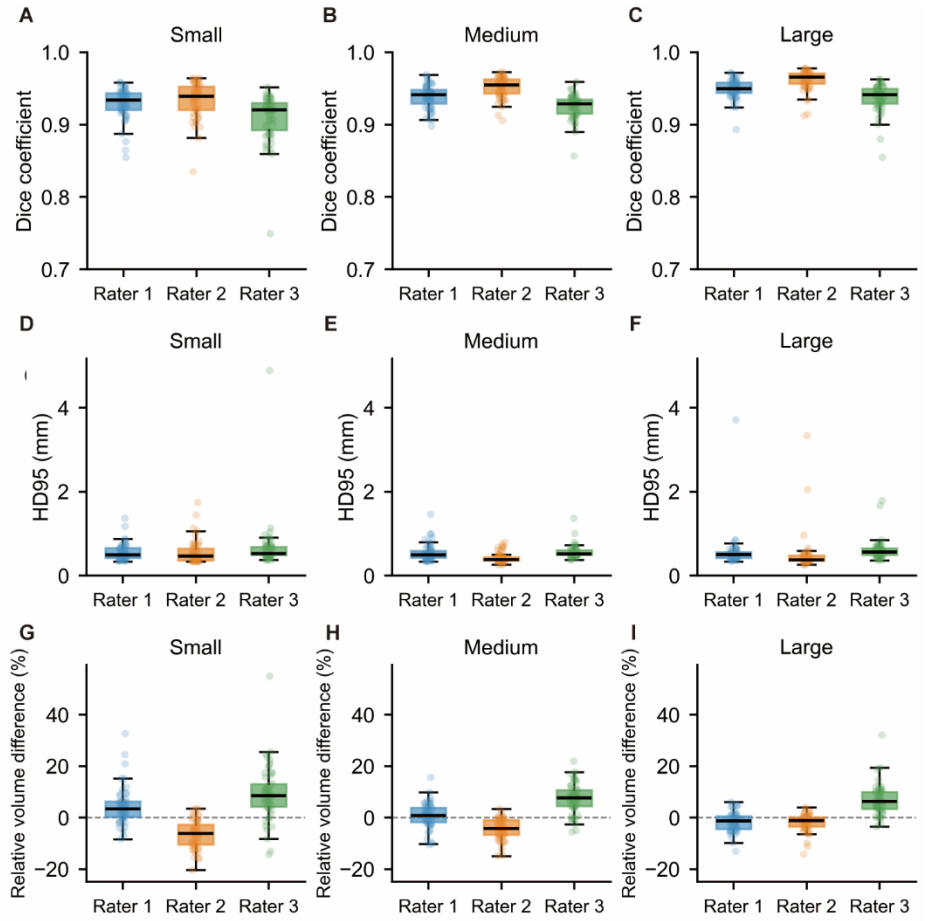

Fig S1. Size-stratified rater-to-consensus agreement and volume differences. Rater-to-STAPLE agreement was evaluated using Dice coefficient (A–C), HD95 (D–F), and relative volume difference (G–I) for small (A, D, G), medium (B, E, H), and large (C, F, I) tumors. Tumors were stratified according to tertiles of the STAPLE consensus volume distribution. Relative volume difference was calculated relative to the STAPLE

consensus volume, with positive and negative values indicating larger and smaller rater-derived volumes than the STAPLE consensus, respectively.

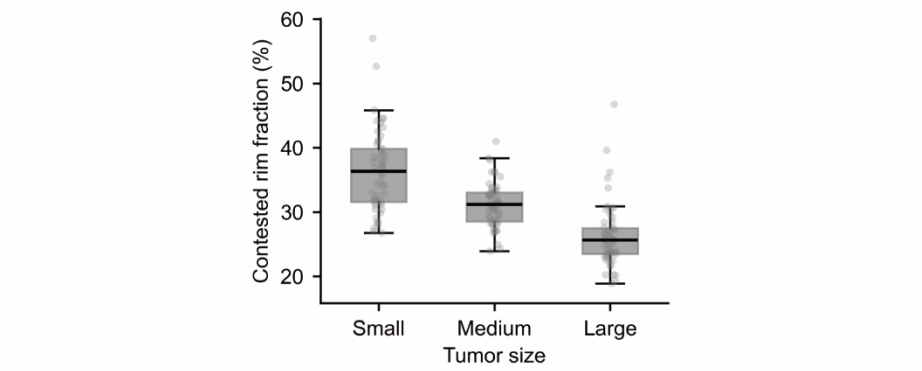

Fig S2. Tumor size dependence of the contested rim fraction. The contested rim was defined as the region included in the union of the three rater masks but not in their unanimous intersection and was expressed as a percentage of the union volume. Tumors were stratified into small, medium, and large groups according to tertiles of the STAPLE consensus volume distribution.

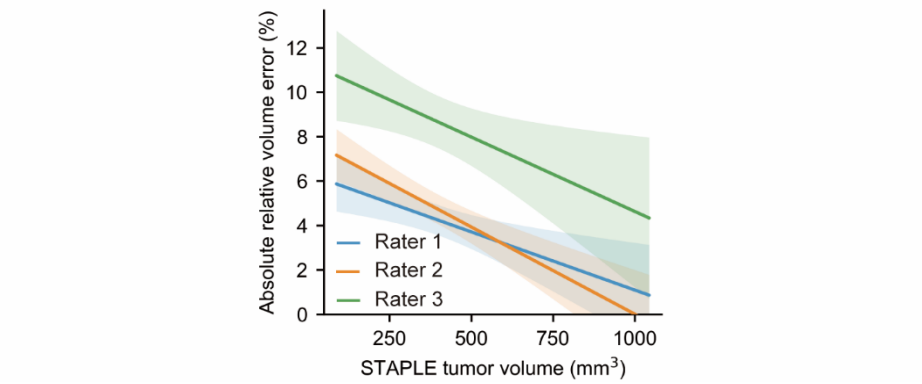

Fig S3. Association between tumor size and absolute relative volume error. Absolute relative volume error was evaluated as a function of STAPLE consensus volume separately for each rater using linear mixed-effects models, with mouse ID included as a random intercept to account for repeated longitudinal measurements. Lines indicate model estimates and shaded areas represent 95% confidence intervals. Model slopes (β) represent the change in absolute relative volume error, in percentage points, per 100 mm^3 increase in STAPLE consensus volume. Absolute relative volume error decreased significantly with increasing tumor volume for all three raters (Rater 1: $\beta = -0.524$, $P < 0.002$; Rater 2: $\beta = -0.785$, $P < 0.0001$; Rater 3: $\beta = -0.67$, $P = 0.013$).