

Supplementary Table S1. Interobserver reproducibility of the three-tier manual p53 immunohistochemistry interpretation

Two independent, blinded interpretations by the two pathologists who established the reference standard (n = 545 cores)

(A) Cross-classification of the two independent interpretations

Pathologist 1 (S.W.C.)	Null	Wild-type	Overexpression	Total
Pathologist 2 (K.K.)				
Null	57	4	0	61
Wild-type	21	319	12	352
Overexpression	0	8	124	132
Total	78	331	136	545

(B) Agreement statistics

Statistic	Value (95% CI)
Concordant cores, n (%)	500 (91.7)
Cohen's κ	0.845 (0.799–0.885)
Linear-weighted κ	0.861
Quadratic-weighted κ	0.885

CI, confidence interval; κ , Cohen's kappa.

^aThe two pathologists independently interpreted all cores using the three-tier system (null / wild-type / overexpression) while blinded to each other's assessment and to all clinicopathologic data. Discordant cores were subsequently reconciled in a joint review by three pathologists (S.W.C., K.E.K., and J.Y.A.), and the resulting consensus classification served as the reference standard for the extended-Youden and volume-under-the-ROC-surface analyses.

^bAgreement was assessed on all 545 cores for which both pathologists recorded a three-tier interpretation. The final analytic cohort (n = 528) is a subset of these cores, additionally requiring at least 100 detected tumor cells by both detection algorithms; because reproducibility of the visual interpretation does not depend on the digital cell count, all 545 paired readings were retained for this analysis.

^cAll 45 discordances occurred between adjacent categories (25 at the null/wild-type boundary; 20 at the wild-type/overexpression boundary). No core was classified as null by one pathologist and as overexpression by the other.

^d95% Confidence interval for κ estimated by bootstrap resampling (2,000 replicates).