

Supplementary Table S8. Confusion matrices between the three-tier classification obtained with the manual H-score cutoffs and those obtained with each of the four data-derived cutoff strategies

Final analytic cohort (n = 528). Rows show the classification obtained with the manual H-score cutoffs (8.0/106.0), used as the common comparator; columns show the classification obtained with each data-derived method.

(A) Tertiles, 27.6/78.6

Manual H-score cutoffs	Low	Wild-type	Overexpression	Total
Low	70	0	0	70
Wild-type	106	176	42	324
Overexpression	0	0	134	134
Total	176	176	176	528

(B) Log-rank optimization, 21.3/95.7

Manual H-score cutoffs	Low	Wild-type	Overexpression	Total
Low	70	0	0	70
Wild-type	66	249	9	324
Overexpression	0	0	134	134
Total	136	249	143	528

(C) Extended Youden index, 14.5/100.8

Manual H-score cutoffs	Low	Wild-type	Overexpression	Total
Low	70	0	0	70
Wild-type	29	288	7	324
Overexpression	0	0	134	134
Total	99	288	141	528

(D) ROC surface (VUS), 12.4/103.8

Manual H-score cutoffs	Low	Wild-type	Overexpression	Total
Low	70	0	0	70
Wild-type	18	302	4	324
Overexpression	0	0	134	134
Total	88	302	138	528

VUS, volume under the ROC surface.

^aFor every cutoff strategy, all discordances occurred between adjacent categories; no core classified as low by the manual cutoffs was assigned to the overexpression tier by a data-derived method, or vice versa.

^bBecause the manual lower cutoff (8.0) was the lowest of the five, every core assigned to the low tier by the manual cutoffs was also assigned to the low tier by each data-derived method; and because the manual upper cutoff (106.0) was the highest, every core assigned to the overexpression tier by the manual cutoffs remained in that tier. All discordance therefore arises from wild-type cores being reassigned to the low or overexpression tiers.

^cThe tertile-based cutoffs reassigned 106 of 324 manually defined wild-type cores to the low tier and 42 to the overexpression tier, reflecting the fact that this strategy partitions the H-score distribution without reference to staining phenotype. The ROC-surface cutoffs reassigned only 22 wild-type cores in total.

^dThe extended-Youden and ROC-surface cutoffs were derived against the manual three-tier immunohistochemical interpretation of the same images as the reference standard; their close agreement with the manual H-score cutoffs therefore reflects internal consistency rather than independent validation. The tertile and log-rank cutoffs were derived without reference to the manual interpretation.

^eSummary agreement statistics (cutoff deviations, concordance, Cohen's kappa, weighted kappa, sensitivity, and specificity) are given in Table 3.