Online Resource 3: Supplementary tables

Article: When Does Supervision Beat Reconstruction? Magnification Shift and the Robustness of AI Quality Control in Digital Pathology

Journal: Journal of Medical Systems

Authors: Konstantinos Moutselos, Ilias Maglogiannis

Corresponding author: Konstantinos Moutselos, Department of Digital Systems, University of Piraeus, Piraeus, Greece; kmouts@unipi.gr

Reference numbers in square brackets refer to the reference list of the article. Sections, tables and figures without the prefix S are those of the article.

Table S1 Datasets, roles and magnification routes

Dataset	Role	Content used	Source format and pixel size	Read route to the working resolution (≈1.0 µm/px)	Reference used
AIRAQc (TCGA diagnostic slides) [14]	Training; cross-validation; development benchmark	50 released slides from 50 patients. 8 excluded (no usable patches or too little tissue) and 2 held out (training monitor only). 16 training slides in 4 folds (4 clean); 24 benchmark slides (11 clean, 13 with artifacts)	Whole-slide pyramids (GDC)	Pyramid level nearest to 1.0 µm/px. Four slides whose nearest level was ≈2.0 µm/px (two training, two benchmark) were re-read from a ≈0.5 µm/px level downsampled ×2 (box). All results use the 10× read	Released pixel annotations (four types; tissue only); secondary reference with teal ink as pen marking
SlideGuard [5]	Blind external evaluation	53 WSIs: TCGA 25; DHMC lung 10 (6 of high quality); lymphatic tissue (MSCNRIO) 6; mouse skin 12	Whole-slide pyramids	Resolution guard (0.9–1.1 µm/px; no upsampling). Read at 0.907–1.011 µm/px: TCGA 20 direct and 5 ×2 box; DHMC 10 direct (1.005–1.008); MSCNRIO 6 direct (0.971); mouse skin 12 box-filter, f ≈ 1.285 (1.000)	Hidden; scored by the organizers' server (per-slide artifact Dice, precision, recall, artifact-free Dice)
SlideInspect [16]	External, shifted magnification	289 H&E cases (135 with out-of-focus annotation); 55 clean controls	Images at 5×	Upsampled ×2 (bilinear)	Pixel masks: out-of-focus, fold, mounting
Fırat [18, 19]	External, matching magnification; resampling experiment (Δ-F)	2,127 images from 302 slides, five organs: 899 clean (150 slides), 1,228 with folds (152 slides); 2 fold images with empty masks excluded	3840 × 2160 px RGB, 10× objective; 0.49 µm/px (dataset value; reproduced from the scale bar, 204 px per 100 µm)	Downsampled ×2 (box) to 1920 × 1080 px at 0.98 µm/px. Scale-bar rectangle masked for every detector	Binary fold masks

Dataset	Role	Content used	Source format and pixel size	Read route to the working resolution (≈1.0 µm/px)	Reference used
		from fold endpoints			

Notes to S1:

- No external slide overlaps the development slides at slide or patient level. The SlideGuard TCGA cohort is reported separately, because the generative backbone and the second encoder were pretrained on corpora that include TCGA.
- No external file was rejected or flagged by the resolution guard.
- The SlideGuard cohorts were assigned from file names, group sizes and scanner metadata (Online Resource 1, S1.2).
- Abbreviations: DHMC, Dartmouth-Hitchcock Medical Center; MSCNRIO, Maria Skłodowska-Curie National Research Institute of Oncology.

Table S2 Sensitivity of the conclusions to the scoring region (true 10× read). (a) Each comparison × metric cell (88 in all) was re-scored in alternative tissue regions, with every decision setting fixed and the rule re-applied within each region. The regions in each cell are, in order, cross-validation SAT, ENT and LAZ (clean FP: SAT and ENT) and development benchmark SAT, LAZ and LAZT. ▲ marks a 95% interval above 0 (above 1 for the development clean-FP ratio), ▼ one below, and ○ one that contains it. Region-dependent cells (bold), where the sign or the position relative to the null differs between regions, number 17; the other 71 cells are region-robust. (b) ΔF1 under the ink-augmented reference in SAT and LAZ, for four contrasts.

a Region sensitivity (bold: region-dependent cell)

Contrast	Macro sensitivity / ΔF1	Out-of-focus	Fold	Pen marking	Air bubble	Clean FP (Δ / ratio)
k-fold (16 slides)						
H2 − A	▲▲▲	○○○	▲▲▲	▲▲▲	▲▲▲	○○
H2s7 − As7	▲▲▲	○○○	▲▲▲	▲▲▲	○○○	○▼
H2 − K2	▲○○	○○○	▲▲▲	▲▲▲	○○○	○▼
H2s7 − K2s7	▲○○	○▼▼	▲▲▲	▲▲▲	○○○	○▼
probe − H2	▲▲▲	—	—	—	—	▼▼
probe − H2s7	▲▲▲	—	—	—	—	▼▼
dev24 (24 slides)						
H2all − K2all	▲▲▲	○▼▼	▲▲▲	▲▲▲	○○○	▼▼▼
H2all_s7 − K2all_s7	▲▲▲	○▼▼	▲○○	▲▲▲	○○○	▼▼▼
H2all − freshcond	▲▲▲	○▼▼	▲▲▲	▲▲▲	○○○	▼▼▼
H2all_s7 − freshcond	▲▲▲	○▼▼	○○○	▲▲▲	○○○	▼▼▼
probe − H2all	○○○	○○○	○○○	○○○	○▲▲	▼▼▼
probe − H2all_s7	○○○	○○○	○○○	○○▲	▲▲▲	▼▼▼
probe − freshcond	○▲▲	▼▼▼	○○○	▲▲▲	▲▲▲	▼▼▼
dev21 (21 slides with GrandQC output)						
H2all − GrandQC	○○○	○○○	○○○	▼▼▼	▼▼▼	○○○
H2all_s7 − GrandQC	○○○	○○○	○○○	▼▼▼	▼▼▼	○○○
probe − GrandQC	▲▲▲	○○○	○○○	▼○○	▼▼▼	▼○○

b Ink-augmented reference

Contrast	ΔF1, SAT [95% CI]	ΔF1, LAZ [95% CI]
H2all − K2all	+0.0387 [−0.0003, +0.0782]	+0.0642 [+0.0230, +0.1274]
H2all_s7 − K2all_s7	+0.0447 [+0.0121, +0.0832]	+0.0659 [+0.0291, +0.1267]
probe − H2all	−0.0694 [−0.1148, +0.0046]	−0.0222 [−0.0520, +0.0185]
probe − freshcond	−0.0138 [−0.0880, +0.0810]	+0.1564 [+0.0572, +0.3029]

Table S3 Fırat resampling (×0.5 → ×2 per patch, Δ-F) by organ, in (a) SAT-F, the primary region, and (b) COV-F (Table 3). Clean false positives and fold sensitivity, original → resampled, are pooled within each organ. Counts are slides; the two fold images with empty masks are excluded. The probe share is the organ's part of the probe's pooled increase in flagged clean grid cells. Points only; pooled intervals are in Table 3. (c) Per clean slide: the number of slides on which clean false positives rose, fell or did not change after resampling, and the number on which the probe's change exceeded, fell below or equalled H2all's.

a SAT-F (primary); original → resampled

Organ	Clean slides	Fold slides	K2all clean FP	K2all fold sens.	H2all clean FP	H2all fold sens.	Probe clean FP	Probe fold sens.	Probe share of clean-FP increase
Brain	25	26	0.0270 → 0.0140	0.3703 → 0.5333	0.0341 → 0.0046	0.4401 → 0.5757	0.0011 → 0.0100	0.5019 → 0.6401	0.021
Kidney	32	30	0.0110 → 0.0024	0.4352 → 0.3057	0.0067 → 0.0053	0.4983 → 0.4427	0.0275 → 0.1793	0.6868 → 0.7895	0.537
Liver	31	44	0.0164 → 0.0043	0.3608 → 0.3442	0.0086 → 0.0074	0.4511 → 0.4741	0.0054 → 0.0087	0.4837 → 0.5813	0.010
Small intestine	35	27	0.2009 → 0.0731	0.6643 → 0.6115	0.1398 → 0.0822	0.6860 → 0.6989	0.2006 → 0.4258	0.6718 → 0.8079	0.410
Testis	27	25	0.0399 → 0.0152	0.2688 → 0.3079	0.0363 → 0.0077	0.3100 → 0.3761	0.0020 → 0.0115	0.3986 → 0.5031	0.022
Pooled	150	152	0.0467 → 0.0170	0.3874 → 0.3707	0.0359 → 0.0168	0.4455 → 0.4705	0.0372 → 0.1135	0.5331 → 0.6422	—

b COV-F; original → resampled

Organ	Clean slides	Fold slides	K2all clean FP	K2all fold sens.	H2all clean FP	H2all fold sens.	Probe clean FP	Probe fold sens.	Probe share of clean-FP increase
Brain	25	26	0.0180 → 0.0121	0.3945 → 0.5658	0.0239 → 0.0048	0.4754 → 0.6233	0.0044 → 0.0221	0.6736 → 0.8403	0.034
Kidney	32	30	0.0683 → 0.0057	0.5034 → 0.3176	0.0420 → 0.0084	0.5665 → 0.5199	0.0605 → 0.4057	0.8396 → 0.9261	0.554
Liver	31	44	0.0203 → 0.0039	0.3737 → 0.3384	0.0141 → 0.0083	0.4811 → 0.5141	0.0187 → 0.0286	0.6070 → 0.7075	0.015
Small intestine	35	27	0.1983 → 0.0540	0.6976 → 0.6321	0.1254 → 0.0638	0.7274 → 0.7657	0.3071 → 0.5833	0.8172 → 0.9338	0.346
Testis	27	25	0.1710 → 0.0766	0.3486 → 0.3028	0.1120 → 0.0348	0.3752 → 0.4352	0.0086 → 0.0407	0.6291 → 0.8028	0.052
Pooled	150	152	0.0891 → 0.0293	0.4300 → 0.3678	0.0602 → 0.0218	0.4864 → 0.5200	0.0675 → 0.1946	0.6981 → 0.8295	—

c Per-slide direction of change on the clean slides

Region	Arm	Clean slides	Rose / exceeds	Fell / below	Unchanged / equal
SAT-F	K2all	150	15	127	8
SAT-F	H2all	150	22	119	9
SAT-F	probe (τ 0.5)	150	119	7	24
SAT-F	probe change vs H2all change	150	137	9	4
COV-F	K2all	150	10	137	3
COV-F	H2all	150	16	133	1
COV-F	probe (τ 0.5)	150	127	8	15
COV-F	probe change vs H2all change	150	138	12	0

Table S4 FP-matched fold sensitivity on the probe's τ curve (τ 0.5–0.99), Fırat, original and after ×0.5 → ×2 resampling (Δ-F). † Intervals use only the bootstrap replicates in which the arm's clean false positives fall within the range of the probe's τ curve (SAT-F: 8,332 and 7,809 of 20,000 replicates fell above it for H2all and H2all_s7). § K2all's clean false positives exceed the curve maximum, so no point is defined. The R0-curve counterparts are in Table 3.

Contrast (fold sensitivity at the arm's clean FP)	SAT-F [95% CI]	COV-F [95% CI]
Original magnification		

Contrast (fold sensitivity at the arm's clean FP)	SAT-F [95% CI]	COV-F [95% CI]
H2all − probe (τ curve)†	−0.079 [−0.125, +0.018]	−0.185 [−0.252, −0.088]
H2all_s7 − probe (τ curve)†	−0.099 [−0.149, −0.0004]	−0.203 [−0.265, −0.098]
K2all − probe (τ curve)	undefined§	—
K2all_s7 − probe (τ curve)	undefined§	—
After resampling (Δ-F)		
H2all − probe (τ curve)	+0.1824 [+0.0981, +0.2529]	+0.1458 [+0.0284, +0.2405]
K2all − probe (τ curve)	+0.0809 [−0.0069, +0.1535]	−0.0579 [−0.1928, +0.0609]

Table S5 Read-level sensitivity: every in-distribution analysis at the earlier ~5× read and at the true 10× read. Four AIRAQc slides had been read at about 2.0 µm per pixel: TCGA-HU-8610 and TCGA-W5-AA30 (cross-validation) and TCGA-HU-A4GC and TCGA-HU-A4GF (development benchmark, both clean); the manuscript reports the 10× read. Sections: (A) cross-validation levels and contrasts; (B) development benchmark; (C) region robustness; (D) controlled resampling of the development slides for the four probes; (E) the frozen combination rules. Each row gives the ~5× and 10× values; no conclusion changes between the reads. Probe rule 2 is τ selected by leave-one-slide-out cross-validation (LOOCV). The upper clip level is the 75th percentile of the arm's score-map values pooled over the held-out slides of the fold, within coverage and tissue (the upper level of R0 65/75); a grid cell is at or above it when its smoothed score reaches it. Point not bracketed: on the full data the arm's clean false positives lie outside the range spanned by the probe's R0 curve, so the point difference is not defined; the interval is taken over the bootstrap replicates in which it lies within that range. Signs and positions in section C are relative to 0 for differences and to 1 for the clean-FP ratio.

Quantity	~5× read	True 10× read
A Cross-validation (16 slides)		
k-fold		
K2 SAT sensitivity OOF	0.6697	0.6924
K2 SAT sensitivity fold	0.1485	0.1385
K2 SAT sensitivity penmark	0.5841	0.5871
K2 SAT sensitivity bubble	0.1020	0.1020
K2 SAT sensitivity macro	0.3761	0.3800
K2s7 SAT sensitivity OOF	0.6740	0.6969
K2s7 SAT sensitivity fold	0.1387	0.1289
K2s7 SAT sensitivity penmark	0.5904	0.5936
K2s7 SAT sensitivity bubble	0.1082	0.1082
K2s7 SAT sensitivity macro	0.3778	0.3819
A SAT sensitivity OOF	0.6476	0.6660
A SAT sensitivity fold	0.1323	0.1232
A SAT sensitivity penmark	0.4868	0.4904
A SAT sensitivity bubble	0.0748	0.0748
A SAT sensitivity macro	0.3354	0.3386
As7 SAT sensitivity OOF	0.6307	0.6495
As7 SAT sensitivity fold	0.1250	0.1156
As7 SAT sensitivity penmark	0.4997	0.5030
As7 SAT sensitivity bubble	0.0748	0.0748
As7 SAT sensitivity macro	0.3326	0.3357
H2 SAT sensitivity OOF	0.6489	0.6809
H2 SAT sensitivity fold	0.2736	0.2520
H2 SAT sensitivity penmark	0.7420	0.7467
H2 SAT sensitivity bubble	0.0910	0.0910
H2 SAT sensitivity macro	0.4389	0.4426
H2s7 SAT sensitivity OOF	0.6658	0.6944
H2s7 SAT sensitivity fold	0.2603	0.2392
H2s7 SAT sensitivity penmark	0.7641	0.7685
H2s7 SAT sensitivity bubble	0.0873	0.0873
H2s7 SAT sensitivity macro	0.4444	0.4474
probe SAT sensitivity OOF	0.8104	0.8104
probe SAT sensitivity fold	0.5818	0.5888

Quantity	~5× read	True 10× read
probe SAT sensitivity penmark	0.8800	0.8800
probe SAT sensitivity bubble	0.8687	0.8687
probe SAT sensitivity macro	0.7852	0.7870
K2 clean FP (ENT)	0.0157	0.0218
K2s7 clean FP (ENT)	0.0117	0.0191
A clean FP (ENT)	0.0072	0.0159
As7 clean FP (ENT)	0.0072	0.0160
H2 clean FP (ENT)	0.0041	0.0086
H2s7 clean FP (ENT)	0.0032	0.0081
probe clean FP (ENT)	0.000574	0.000000
K2 clean-FP cells (ENT)	49,474	74,981
K2s7 clean-FP cells (ENT)	36,752	65,850
A clean-FP cells (ENT)	22,614	54,633
As7 clean-FP cells (ENT)	22,656	55,231
H2 clean-FP cells (ENT)	12,884	29,566
H2s7 clean-FP cells (ENT)	10,116	27,962
probe clean-FP cells (ENT)	1,810	0
k-fold contrasts		
H2 − A macro-mean share of annotated cells at or above the upper clip level (SAT)	+0.0653 [+0.0286, +0.1524]	+0.0667 [+0.0314, +0.1430]
H2 − A macro-mean sensitivity (SAT)	+0.1035 [+0.0495, +0.1810]	+0.1041 [+0.0490, +0.1732]
H2 − A clean FP (ENT)	−0.0031 [−0.0481, +0.0021]	−0.0073 [−0.0503, +0.0020]
H2s7 − As7 macro-mean share of annotated cells at or above the upper clip level (SAT)	+0.0822 [+0.0387, +0.1688]	+0.0832 [+0.0402, +0.1588]
H2s7 − As7 macro-mean sensitivity (SAT)	+0.1118 [+0.0474, +0.1981]	+0.1116 [+0.0459, +0.1899]
H2s7 − As7 clean FP (ENT)	−0.0040 [−0.0531, −0.0001]	−0.0079 [−0.0511, −0.0002]
A − K2 macro-mean share of annotated cells at or above the upper clip level (SAT)	−0.0243 [−0.0845, +0.0042]	−0.0262 [−0.0752, +0.0024]
A − K2 macro-mean sensitivity (SAT)	−0.0407 [−0.1112, +0.0087]	−0.0414 [−0.0985, +0.0084]
A − K2 clean FP (ENT)	−0.0085 [−0.0570, +0.0351]	−0.0059 [−0.0570, +0.0354]
As7 − K2s7 macro-mean share of annotated cells at or above the upper clip level (SAT)	−0.0317 [−0.0900, +0.0016]	−0.0331 [−0.0813, +0.0009]
As7 − K2s7 macro-mean sensitivity (SAT)	−0.0453 [−0.1164, +0.0103]	−0.0462 [−0.1062, +0.0094]
As7 − K2s7 clean FP (ENT)	−0.0045 [−0.0413, +0.0419]	−0.0031 [−0.0413, +0.0284]
H2 − K2 macro-mean share of annotated cells at or above the upper clip level (SAT)	+0.0410 [+0.0176, +0.0901]	+0.0405 [+0.0187, +0.0877]
H2 − K2 macro-mean sensitivity (SAT)	+0.0628 [+0.0390, +0.1040]	+0.0626 [+0.0350, +0.1049]
H2 − K2 clean FP (ENT)	−0.0116 [−0.0609, −0.0031]	−0.0132 [−0.0562, −0.0063]
H2s7 − K2s7 macro-mean share of annotated cells at or above the upper clip level (SAT)	+0.0505 [+0.0229, +0.0997]	+0.0501 [+0.0233, +0.0979]
H2s7 − K2s7 macro-mean sensitivity (SAT)	+0.0665 [+0.0362, +0.1150]	+0.0654 [+0.0339, +0.1119]
H2s7 − K2s7 clean FP (ENT)	−0.0084 [−0.0480, −0.0023]	−0.0110 [−0.0450, −0.0038]
H2 − probe: macro sensitivity at H2's clean FP (ENT), probe R0 curve	point not bracketed; [−0.3542, +0.0215]	−0.2356 [−0.4050, −0.0165]
H2 − probe: macro sensitivity at H2's clean FP (ENT), extended probe R0 curve	+0.0847 [−0.3022, +0.4102]	−0.2356 [−0.4045, +0.3780]
H2s7 − probe: macro sensitivity at H2s7's clean FP (ENT), probe R0 curve	point not bracketed; [−0.3290, +0.1282]	−0.2265 [−0.4179, −0.0058]
H2s7 − probe: macro sensitivity at H2s7's clean FP (ENT), extended probe R0 curve	+0.1450 [−0.2560, +0.4397]	−0.2265 [−0.4178, +0.3932]
H2 − probe: clean FP (ENT), both at R0 65/75	−0.0018 [−0.0631, +0.0238]	+0.0075 [+0.0001, +0.0519]
H2s7 − probe: clean FP (ENT), both at R0 65/75	−0.0027 [−0.0666, +0.0130]	+0.0070 [+0.0000, +0.0489]
K2 − probe: clean FP (ENT), both at R0 65/75	+0.0098 [−0.0309, +0.0737]	+0.0207 [+0.0085, +0.0737]
H2 − probe (τ 0.5): precision, 12 artifact slides	−0.2534 [−0.3683, −0.1077]	−0.2508 [−0.3683, −0.1045]
H2 − probe (τ 0.5): Dice, 12 artifact slides	−0.4188 [−0.6161, −0.1927]	−0.4133 [−0.6149, −0.1830]
H2s7 − probe (τ 0.5): precision, 12 artifact slides	−0.2487 [−0.3684, −0.1025]	−0.2468 [−0.3689, −0.0998]
H2s7 − probe (τ 0.5): Dice, 12 artifact slides	−0.4125 [−0.6161, −0.1792]	−0.4076 [−0.6145, −0.1711]
K2 − probe (τ 0.5): precision, 12 artifact slides	−0.2653 [−0.4181, −0.0856]	−0.2644 [−0.4202, −0.0844]
K2 − probe (τ 0.5): Dice, 12 artifact slides	−0.4526 [−0.6212, −0.2295]	−0.4488 [−0.6212, −0.2217]
H2 − probe (R0 65/75): precision, 12 artifact slides	−0.1522 [−0.3067, +0.0966]	−0.1413 [−0.3022, +0.1095]
H2 − probe (R0 65/75): Dice, 12 artifact slides	−0.1992 [−0.4041, +0.1343]	−0.1939 [−0.4011, +0.1392]
probe: cell-level AUROC (SAT), cross-validation fold 1	0.9672	0.9675
probe: cell-level AUROC (SAT), cross-validation fold 2	0.9472	0.9472
probe: cell-level AUROC (SAT), cross-validation fold 3	0.9264	0.9264
probe: cell-level AUROC (SAT), cross-validation fold 4	0.8895	0.8950
B Development benchmark (24 slides)		
dev24		
freshcond rel SAT sensitivity OOF	0.9191	0.9173

Quantity	~5× read	True 10× read
freshcond rel SAT sensitivity fold	0.4781	0.4782
freshcond rel SAT sensitivity penmark	0.4857	0.4886
freshcond rel SAT sensitivity bubble	0.5239	0.5246
freshcond aug SAT sensitivity penmark	0.6218	0.6239
K2all rel SAT sensitivity OOF	0.9015	0.9000
K2all rel SAT sensitivity fold	0.4853	0.4766
K2all rel SAT sensitivity penmark	0.5558	0.5588
K2all rel SAT sensitivity bubble	0.5939	0.5939
K2all aug SAT sensitivity penmark	0.6733	0.6756
K2all aug SAT sensitivity bubble	0.5687	0.5231
H2all rel SAT sensitivity OOF	0.9012	0.9013
H2all rel SAT sensitivity fold	0.5798	0.5827
H2all rel SAT sensitivity penmark	0.7593	0.7625
H2all rel SAT sensitivity bubble	0.6067	0.6060
H2all aug SAT sensitivity penmark	0.8223	0.8246
probe rule 1 rel SAT sensitivity OOF	0.7393	0.7393
probe rule 1 rel SAT sensitivity fold	0.6407	0.6407
probe rule 1 rel SAT sensitivity penmark	0.8954	0.8954
probe rule 1 rel SAT sensitivity bubble	0.8748	0.8748
probe rule 1 aug SAT sensitivity penmark	0.6757	0.6757
K2all_s7 rel SAT sensitivity OOF	0.9062	0.9024
K2all_s7 rel SAT sensitivity fold	0.4922	0.4736
K2all_s7 rel SAT sensitivity penmark	0.5541	0.5571
K2all_s7 rel SAT sensitivity bubble	0.5952	0.5936
K2all_s7 aug SAT sensitivity penmark	0.6721	0.6743
K2all_s7 aug SAT sensitivity bubble	0.5910	0.5689
H2all_s7 rel SAT sensitivity OOF	0.9058	0.9037
H2all_s7 rel SAT sensitivity fold	0.5160	0.5175
H2all_s7 rel SAT sensitivity penmark	0.7590	0.7618
H2all_s7 rel SAT sensitivity bubble	0.6078	0.6051
H2all_s7 aug SAT sensitivity penmark	0.8221	0.8241
freshcond unannotated teal flagged (SAT)	0.9943	0.9943
K2all unannotated teal flagged (SAT)	0.9944	0.9944
H2all unannotated teal flagged (SAT)	0.9944	0.9944
K2all_s7 unannotated teal flagged (SAT)	0.9944	0.9944
H2all_s7 unannotated teal flagged (SAT)	0.9944	0.9944
probe unannotated teal flagged (SAT)	0.0756	0.0756
probe T0 unannotated teal flagged (SAT)	0.0756	0.0756
probe T1 unannotated teal flagged (SAT)	0.1371	0.1371
probe T2 unannotated teal flagged (SAT)	0.0363	0.0363
probe T3 unannotated teal flagged (SAT)	0.9840	0.9840
dev24 (Table 1)		
freshcond; F1 rel (24)	0.6716	0.6741
freshcond; F1 aug (24)	0.7767	0.7815
freshcond; Clean FP cells rel (11 clean)	531,943	528,429
freshcond; Clean FP cells aug	330,568	259,807
K2all; F1 rel (24)	0.6896	0.6946
K2all; ΔF1 vs freshcond rel [95% CI]	+0.0179 [−0.0054, +0.0456]	+0.0205 [−0.0031, +0.0485]
K2all; F1 aug (24)	0.7951	0.7983
K2all; ΔF1 vs freshcond aug [95% CI]	+0.0184 [−0.0024, +0.0401]	+0.0169 [+0.0005, +0.0364]
K2all; Clean FP cells rel (11 clean)	440,254	351,079
K2all; Clean-FP ratio vs freshcond rel [95% CI]	0.828 [0.493, 1.024]	0.664 [0.269, 0.947]
K2all; Clean FP cells aug	177,775	185,547
K2all_s7; F1 rel (24)	0.6887	0.6895
K2all_s7; ΔF1 vs freshcond rel [95% CI]	+0.0171 [−0.0072, +0.0458]	+0.0154 [−0.0104, +0.0441]
K2all_s7; F1 aug (24)	0.7910	0.7956
K2all_s7; ΔF1 vs freshcond aug [95% CI]	+0.0143 [−0.0180, +0.0406]	+0.0141 [−0.0092, +0.0360]
K2all_s7; Clean FP cells rel (11 clean)	395,412	394,239
K2all_s7; Clean-FP ratio vs freshcond rel [95% CI]	0.743 [0.541, 0.842]	0.746 [0.519, 0.842]
K2all_s7; Clean FP cells aug	152,050	159,157
H2all; F1 rel (24)	0.7492	0.7536
H2all; ΔF1 vs freshcond rel [95% CI]	+0.0775 [+0.0158, +0.1597]	+0.0795 [+0.0221, +0.1599]
H2all; F1 aug (24)	0.8370	0.8371
H2all; ΔF1 vs freshcond aug [95% CI]	+0.0604 [+0.0072, +0.1171]	+0.0556 [+0.0016, +0.1122]
H2all; Clean FP cells rel (11 clean)	118,789	125,176
H2all; Clean-FP ratio vs freshcond rel [95% CI]	0.223 [0.122, 0.268]	0.237 [0.137, 0.276]
H2all; Clean FP cells aug	118,789	125,176
H2all_s7; F1 rel (24)	0.7488	0.7528
H2all_s7; ΔF1 vs freshcond rel [95% CI]	+0.0772 [+0.0215, +0.1579]	+0.0787 [+0.0256, +0.1583]
H2all_s7; F1 aug (24)	0.8403	0.8404
H2all_s7; ΔF1 vs freshcond aug [95% CI]	+0.0636 [+0.0152, +0.1198]	+0.0589 [+0.0096, +0.1141]
H2all_s7; Clean FP cells rel (11 clean)	178,821	156,281

Quantity	~5× read	True 10× read
H2all_s7; Clean-FP ratio vs freshcond rel [95% CI]	0.336 [0.237, 0.453]	0.296 [0.141, 0.445]
H2all_s7; Clean FP cells aug	89,298	94,601
probe rule 1 (τ 0.5); F1 rel (24)	0.8156	0.8157
probe rule 1 (τ 0.5); ΔF1 vs freshcond rel [95% CI]	+0.1440 [−0.0305, +0.3638]	+0.1416 [−0.0329, +0.3611]
probe rule 1 (τ 0.5); F1 aug (24)	0.7676	0.7677
probe rule 1 (τ 0.5); ΔF1 vs freshcond aug [95% CI]	−0.0090 [−0.0838, +0.0877]	−0.0138 [−0.0880, +0.0810]
probe rule 1 (τ 0.5); Clean FP cells rel (11 clean)	3,245	2,724
probe rule 1 (τ 0.5); Clean-FP ratio vs freshcond rel [95% CI]	0.006 [0.000, 0.015]	0.005 [0.000, 0.013]
probe rule 1 (τ 0.5); Clean FP cells aug	3,245	2,724
probe rule 2 (LOOCV τ); F1 rel (24)	0.8027	0.8027
probe rule 2 (LOOCV τ); ΔF1 vs freshcond rel [95% CI]	+0.1311 [−0.0600, +0.3638]	+0.1287 [−0.0625, +0.3609]
probe rule 2 (LOOCV τ); F1 aug (24)	0.7582	0.7582
probe rule 2 (LOOCV τ); ΔF1 vs freshcond aug [95% CI]	−0.0185 [−0.0988, +0.0782]	−0.0233 [−0.1033, +0.0723]
probe rule 2 (LOOCV τ); Clean FP cells rel (11 clean)	3,245	2,724
probe rule 2 (LOOCV τ); Clean-FP ratio vs freshcond rel [95% CI]	0.006 [0.000, 0.015]	0.005 [0.000, 0.013]
probe rule 2 (LOOCV τ); Clean FP cells aug	3,245	2,724
GrandQC (21 slides)†; F1 rel (24)	0.6905	0.6905
GrandQC (21 slides)†; ΔF1 vs freshcond rel [95% CI]	+0.0097 [−0.1714, +0.2243]	+0.0069 [−0.1752, +0.2213]
GrandQC (21 slides)†; F1 aug (24)	0.7664	0.7664
GrandQC (21 slides)†; ΔF1 vs freshcond aug [95% CI]	−0.0169 [−0.1914, +0.1479]	−0.0203 [−0.1953, +0.1442]
GrandQC (21 slides)†; Clean FP cells rel (11 clean)	330,115	330,174
GrandQC (21 slides)†; Clean-FP ratio vs freshcond rel [95% CI]	1.186	1.227
GrandQC (21 slides)†; Clean FP cells aug	330,115	330,174
dev24 (Table 1 contrasts)		
H2all − K2all; ΔF1 rel [95% CI]	+0.0596 [+0.0175, +0.1171]	+0.0589 [+0.0195, +0.1154]
H2all − K2all; ΔF1 aug [95% CI]	+0.0420 [+0.0061, +0.0803]	+0.0387 [−0.0003, +0.0782]
H2all_s7 − K2all_s7; ΔF1 rel [95% CI]	+0.0601 [+0.0228, +0.1159]	+0.0633 [+0.0261, +0.1194]
H2all_s7 − K2all_s7; ΔF1 aug [95% CI]	+0.0493 [+0.0168, +0.0893]	+0.0447 [+0.0121, +0.0832]
H2all − probe rule 1; ΔF1 rel [95% CI]	−0.0665 [−0.2124, +0.0598]	−0.0621 [−0.2069, +0.0630]
H2all − probe rule 1; ΔF1 aug [95% CI]	+0.0694 [−0.0045, +0.1149]	+0.0694 [−0.0046, +0.1148]
H2all_s7 − probe rule 1; ΔF1 rel [95% CI]	−0.0668 [−0.2101, +0.0592]	−0.0629 [−0.2056, +0.0634]
H2all_s7 − probe rule 1; ΔF1 aug [95% CI]	+0.0727 [−0.0009, +0.1187]	+0.0727 [−0.0007, +0.1185]
probe rule 1 − GrandQC (21)†; ΔF1 rel [95% CI]	+0.1251 [+0.0819, +0.1808]	+0.1251 [+0.0819, +0.1809]
probe rule 1 − GrandQC (21)†; ΔF1 aug [95% CI]	+0.0012 [−0.1063, +0.1326]	+0.0012 [−0.1063, +0.1326]
C Region robustness		
Table S2		
region-robust / region-dependent cells	71 / 17	71 / 17
dev24 H2all-K2all; Δ sensitivity out-of-focus; SAT	−0.0003 [−0.0144, +0.0203] (−, contains 0)	+0.0013 [−0.0093, +0.0204] (+, contains 0)
dev24 H2all_s7-K2all_s7; Δ sensitivity out-of-focus; SAT	−0.0004 [−0.0176, +0.0205] (−, contains 0)	+0.0013 [−0.0124, +0.0237] (+, contains 0)
dev24 H2all_s7-K2all_s7; Δ sensitivity fold; SAT	+0.0238 [−0.0007, +0.0697] (+, contains 0)	+0.0439 [+0.0033, +0.0883] (+, above 0)
dev24 probe-H2all_s7; Δ sensitivity air-bubble; SAT	+0.2670 [−0.0004, +0.6786] (+, contains 0)	+0.2697 [+0.0023, +0.6812] (+, above 0)
dev24 probe-freshcond; Δ sensitivity out-of-focus; LAZ	−0.1130 [−0.1663, +0.0004] (−, contains 0)	−0.1113 [−0.1664, −0.0006] (−, below 0)
dev24 probe-freshcond; Δ sensitivity out-of-focus; LAZT	−0.1130 [−0.1663, +0.0004] (−, contains 0)	−0.1113 [−0.1664, −0.0006] (−, below 0)
dev21 probe-GrandQC; clean FP ratio; SAT	+0.0098 [+0.0000, +5.6223] (−, contains 1)	+0.0083 [+0.0000, +0.0083] (−, below 1)
Table S2c		
H2all − K2all augmented ΔF1 SAT	+0.0420 [+0.0061, +0.0803]	+0.0387 [−0.0003, +0.0782]
H2all − K2all augmented ΔF1 LAZ	+0.0677 [+0.0287, +0.1304]	+0.0642 [+0.0230, +0.1274]
H2all_s7 − K2all_s7 augmented ΔF1 SAT	+0.0493 [+0.0168, +0.0893]	+0.0447 [+0.0121, +0.0832]
H2all_s7 − K2all_s7 augmented ΔF1 LAZ	+0.0708 [+0.0316, +0.1351]	+0.0659 [+0.0291, +0.1267]
probe − H2all augmented ΔF1 SAT	−0.0694 [−0.1149, +0.0045]	−0.0694 [−0.1148, +0.0046]
probe − H2all augmented ΔF1 LAZ	−0.0224 [−0.0526, +0.0185]	−0.0222 [−0.0520, +0.0185]
probe − freshcond augmented ΔF1 SAT	−0.0090 [−0.0838, +0.0877]	−0.0138 [−0.0880, +0.0810]
probe − freshcond augmented ΔF1 LAZ	+0.1633 [+0.0608, +0.3122]	+0.1564 [+0.0572, +0.3029]
D Controlled resampling of the development slides		
UNI2-h		
original; clean FP cells (11 clean)	3,245	2,724

Quantity	~5× read	True 10× read
original; pooled F1 (released GT)	0.8156	0.8157
original; clean FP fraction	0.000330	0.000260
resampled; clean FP cells (11 clean)	548,628	619,578
resampled; pooled F1 (released GT)	0.4920	0.4909
resampled; clean FP fraction	0.055840	0.059209
UNI2-h + scale aug.		
original; clean FP cells (11 clean)	6,242	12,459
original; pooled F1 (released GT)	0.7792	0.7789
original; clean FP fraction	0.000635	0.001191
resampled; clean FP cells (11 clean)	5,508	12,596
resampled; pooled F1 (released GT)	0.7993	0.7990
resampled; clean FP fraction	0.000561	0.001204
Phikon-v2		
original; clean FP cells (11 clean)	14,548	18,443
original; pooled F1 (released GT)	0.7796	0.7795
original; clean FP fraction	0.001481	0.001762
resampled; clean FP cells (11 clean)	46,452	70,207
resampled; pooled F1 (released GT)	0.7187	0.7178
resampled; clean FP fraction	0.004728	0.006709
Phikon-v2 + scale aug.		
original; clean FP cells (11 clean)	13,041	22,935
original; pooled F1 (released GT)	0.7622	0.7617
original; clean FP fraction	0.001327	0.002192
resampled; clean FP cells (11 clean)	10,672	19,704
resampled; pooled F1 (released GT)	0.7932	0.7927
resampled; clean FP fraction	0.001086	0.001883
clean SAT cells on the 11 clean slides (all probes)	9,825,041	10,464,294
Δ-B ratio resampled / original; UNI2-h	169.07	227.45
Δ-B ratio; UNI2-h; replicates with zero original clean FP (no interval)	2,170	6,934
Δ-B ratio resampled / original; Phikon-v2	3.19	3.81
Δ-B ratio; Phikon-v2; replicates with zero original clean FP (no interval)	124	21
Phikon-v2 probe: clean FP cells, resampled − original	+31,904 [+2,120, +79,480]	+51,764 [+4,652, +113,526]
Phikon-v2 + scale aug. − Phikon-v2: clean FP cells, resampled read	−35,780 [−89,838, −1,108]	−50,503 [−111,202, −4,848]
Phikon-v2 + scale aug. − Phikon-v2: pooled F1, original read	−0.0175 [−0.0721, +0.0153]	−0.0177 [−0.0723, +0.0150]
UNI2-h + scale aug. − UNI2-h: clean FP cells, resampled read	−543,120 [−1,136,513, −129,085]	−606,982 [−1,186,422, −188,089]
UNI2-h + scale aug. − UNI2-h: pooled F1, original read	−0.0364 [−0.0904, −0.0022]	−0.0367 [−0.0908, −0.0026]
E Frozen combination rules		
PC (0.95, 0.05)		
K2all; pooled F1 (released GT)	0.3943	0.3935
K2all; precision	0.5791	0.5754
K2all; clean FP cells (11 clean)	400,596	428,406
K2all_s7; pooled F1 (released GT)	0.3939	0.3931
K2all_s7; precision	0.5787	0.5749
K2all_s7; clean FP cells (11 clean)	400,024	428,027
H2all; pooled F1 (released GT)	0.4194	0.4184
H2all; precision	0.5722	0.5686
H2all; clean FP cells (11 clean)	419,690	449,387
H2all_s7; pooled F1 (released GT)	0.4124	0.4114
H2all_s7; precision	0.5762	0.5723
H2all_s7; clean FP cells (11 clean)	403,302	433,754
PR (m 16)		
K2all; pooled F1 (released GT)	0.6012	0.6012
K2all; precision	0.7348	0.7349
K2all; clean FP cells (11 clean)	8,057	7,225
K2all_s7; pooled F1 (released GT)	0.6004	0.6003
K2all_s7; precision	0.7342	0.7343
K2all_s7; clean FP cells (11 clean)	8,290	7,183
H2all; pooled F1 (released GT)	0.6363	0.6363
H2all; precision	0.7449	0.7451
H2all; clean FP cells (11 clean)	7,456	6,568
H2all_s7; pooled F1 (released GT)	0.6324	0.6324
H2all_s7; precision	0.7440	0.7441
H2all_s7; clean FP cells (11 clean)	7,461	6,744
BF (α 0.5)		
K2all; pooled F1 (released GT)	0.8044	0.8044

Quantity	~5× read	True 10× read
K2all; precision	0.8233	0.8233
K2all; clean FP cells (11 clean)	6,287	5,592
K2all_s7; pooled F1 (released GT)	0.8047	0.8047
K2all_s7; precision	0.8237	0.8237
K2all_s7; clean FP cells (11 clean)	6,329	5,693
H2all; pooled F1 (released GT)	0.8078	0.8079
H2all; precision	0.8142	0.8142
H2all; clean FP cells (11 clean)	5,125	4,371
H2all_s7; pooled F1 (released GT)	0.8084	0.8085
H2all_s7; precision	0.8152	0.8153
H2all_s7; clean FP cells (11 clean)	4,853	4,056
PC fallback slides (every arm)	0	0
H2all: pooled F1, BF (leave-one-slide-out) | R0 65/75	0.8078 | 0.7551	0.8079 | 0.7547
H2all: clean FP cells, BF (leave-one-slide-out) | R0 65/75	5,125 | 217,554	4,371 | 229,036
H2all_s7: pooled F1, BF (leave-one-slide-out) | R0 65/75	0.8084 | 0.7590	0.8085 | 0.7587
H2all_s7: clean FP cells, BF (leave-one-slide-out) | R0 65/75	4,853 | 178,965	4,056 | 189,584
BF meets both criteria for H2all and H2all_s7 (F1 at least R0 F1 − 0.03; fewer clean FP than R0)	yes	yes

Table S6 Second encoder and scale augmentation: probe levels at the original read and after ×0.5 → ×2 resampling (τ 0.5). (a) Development benchmark, true 10× read: clean false-positive grid cells and fraction on the 11 clean slides (10,464,294 SAT grid cells), and pooled F1 under the released annotations; points only. (b) Fırat, SAT-F, and (c) Fırat, COV-F: clean false positives and fold sensitivity with 95% intervals (paired slide bootstrap stratified by clean and fold slides, 20,000 replicates). Scale augmentation was trained with the same resampling operator as the one evaluated.

a Development benchmark, true 10× read (original → resampled)

Probe	Clean FP cells, original	Clean FP cells, resampled	Clean FP fraction, original	Clean FP fraction, resampled	Pooled F1, original	Pooled F1, resampled
UNI2-h	2,724	619,578	0.000260	0.059209	0.8157	0.4909
UNI2-h + scale aug.	12,459	12,596	0.001191	0.001204	0.7789	0.7990
Phikon-v2	18,443	70,207	0.001762	0.006709	0.7795	0.7178
Phikon-v2 + scale aug.	22,935	19,704	0.002192	0.001883	0.7617	0.7927

b Fırat, SAT-F (primary)

Probe	Clean FP, original [95% CI]	Clean FP, resampled [95% CI]	Fold sensitivity, original [95% CI]	Fold sensitivity, resampled [95% CI]
UNI2-h	0.0372 [0.0245, 0.0523]	0.1135 [0.0804, 0.1493]	0.533 [0.478, 0.583]	0.642 [0.592, 0.688]
UNI2-h + scale aug.	0.0530 [0.0380, 0.0715]	0.0545 [0.0388, 0.0742]	0.618 [0.564, 0.665]	0.596 [0.544, 0.641]
Phikon-v2	0.1633 [0.1379, 0.1923]	0.1608 [0.1370, 0.1883]	0.403 [0.353, 0.454]	0.418 [0.385, 0.451]
Phikon-v2 + scale aug.	0.2234 [0.1985, 0.2504]	0.1348 [0.1159, 0.1559]	0.409 [0.361, 0.458]	0.304 [0.271, 0.337]

c Fırat, COV-F

Probe	Clean FP, original [95% CI]	Clean FP, resampled [95% CI]	Fold sensitivity, original [95% CI]	Fold sensitivity, resampled [95% CI]
UNI2-h	0.0675 [0.0488, 0.0899]	0.1946 [0.1464, 0.2488]	0.698 [0.643, 0.747]	0.830 [0.786, 0.865]
UNI2-h + scale aug.	0.1314 [0.1027, 0.1638]	0.1436 [0.1117, 0.1799]	0.800 [0.745, 0.843]	0.783 [0.731, 0.825]
Phikon-v2	0.2700 [0.2205, 0.3218]	0.2724 [0.2337, 0.3147]	0.518 [0.460, 0.577]	0.560 [0.524, 0.595]
Phikon-v2 + scale aug.	0.3691 [0.3201, 0.4200]	0.2285 [0.1950, 0.2664]	0.518 [0.462, 0.573]	0.388 [0.347, 0.428]

Table S7 SlideGuard: the seed-7 arms and the K2-based combination-rule arms, as in Table 2. Artifact Dice by cohort on the 47 slides without the six high-quality DHMC slides, and artifact-free Dice on those six; mean over slides [95% slide-bootstrap interval, 20,000 replicates].

Arm	TCGA (25)	MSCNRIO (6)	DHMC non-HQ (4)	Human (35)	Mouse skin (12)	HQ DHMC (6), artifact-free Dice
K2all_s7	0.1547 [0.0753, 0.2511]	0.2450 [0.1366, 0.3468]	0.2138 [0.0091, 0.4185]	0.1769 [0.1103, 0.2524]	0.0972 [0.0419, 0.1621]	0.9283 [0.8752, 0.9666]
H2all_s7	0.1490 [0.0768, 0.2364]	0.2547 [0.1431, 0.3646]	0.1954 [0.0105, 0.3803]	0.1724 [0.1108, 0.2420]	0.1159 [0.0503, 0.1922]	0.9316 [0.8852, 0.9673]
PC-K2all	0.1848 [0.1013, 0.2804]	0.1452 [0.0620, 0.2366]	0.3608 [0.0177, 0.7040]	0.1981 [0.1239, 0.2820]	0.1295 [0.0577, 0.2205]	0.9880 [0.9851, 0.9917]
PC-K2all_s7	0.1830 [0.1004, 0.2778]	0.1523 [0.0614, 0.2531]	0.3580 [0.0174, 0.6986]	0.1977 [0.1242, 0.2811]	0.1227 [0.0536, 0.2077]	0.9881 [0.9854, 0.9916]
PC-H2all_s7	0.1776 [0.0968, 0.2715]	0.1678 [0.1095, 0.2250]	0.3306 [0.0199, 0.6412]	0.1934 [0.1234, 0.2729]	0.1744 [0.0837, 0.2768]	0.9860 [0.9827, 0.9902]
PR-K2all	0.1591 [0.0724, 0.2595]	0.2484 [0.1767, 0.3269]	0.4169 [0.1392, 0.6946]	0.2039 [0.1284, 0.2874]	0.1896 [0.0883, 0.3077]	0.9994 [0.9984, 1.0000]
PR-K2all_s7	0.1536 [0.0688, 0.2537]	0.2500 [0.1812, 0.3245]	0.4174 [0.1402, 0.6946]	0.2003 [0.1255, 0.2841]	0.1870 [0.0886, 0.3005]	0.9994 [0.9984, 1.0000]
PR-H2all_s7	0.1671 [0.0774, 0.2692]	0.2290 [0.1714, 0.2960]	0.3975 [0.1375, 0.6576]	0.2040 [0.1289, 0.2862]	0.2115 [0.0945, 0.3498]	0.9993 [0.9983, 0.9999]
BF-K2all	0.1574 [0.0707, 0.2569]	0.2280 [0.1341, 0.3274]	0.2903 [0.0002, 0.5856]	0.1847 [0.1118, 0.2670]	0.2338 [0.1107, 0.3733]	0.9997 [0.9992, 1.0000]
BF-K2all_s7	0.1581 [0.0712, 0.2577]	0.2300 [0.1359, 0.3296]	0.2903 [0.0002, 0.5883]	0.1855 [0.1126, 0.2678]	0.2330 [0.1103, 0.3713]	0.9997 [0.9992, 1.0000]
BF-H2all_s7	0.1619 [0.0705, 0.2661]	0.2341 [0.1471, 0.3218]	0.3249 [0.0035, 0.6464]	0.1929 [0.1161, 0.2796]	0.2372 [0.1119, 0.3779]	0.9996 [0.9990, 1.0000]