

A novel topology-guided adaptive strategy for automated nuclear-cytoplasmic H-Score quantification in immunohistochemistry

Weiyu Wang

Quanzhou Orthopedic-traumatological Hospital, Fujian University of Traditional Chinese Medicine

Mengting Lin

Fuzhou Hospital of Traditional Chinese Medicine

Fuxian Zou

Quanzhou Orthopedic-traumatological Hospital, Fujian University of Traditional Chinese Medicine

Zhenzhen Yang

Quanzhou Orthopedic-traumatological Hospital, Fujian University of Traditional Chinese Medicine

Mingqing Huang

hmq1115@126.com

College of Pharmacy, Fujian-Hong Kong-Macau-Taiwan Collaborative Laboratory for the Inheritance and Innovation of Traditional Chinese Medicine, Fujian University of Traditional Chinese Medicine

Changxian Chen

Quanzhou Orthopedic-traumatological Hospital, Fujian University of Traditional Chinese Medicine

Article

Keywords: Digital pathology, Immunohistochemistry (IHC), Deep learning segmentation, Batch effects, H-Score quantification, ImageJ/Fiji plugin

Posted Date: September 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10678817/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A novel topology-guided adaptive strategy for automated nuclear-cytoplasmic H-Score quantification in immunohistochemistry

Weiyu Wang ^{a, 1}, Mengting Lin ^{b, 1}, Fuxian Zou ^{a, 1}, Zhenzhen Yang ^{a, *}, Mingqing Huang ^{c, *},
Changxian Chen ^{a, *}

^a Quanzhou Orthopedic-traumatological Hospital, Fujian University of Traditional Chinese Medicine, Quanzhou 362000, Fujian, China

^b Fuzhou Hospital of Traditional Chinese Medicine, Fuzhou 350001, China

^c College of Pharmacy, Fujian-Hong Kong-Macau-Taiwan Collaborative Laboratory for the Inheritance and Innovation of Traditional Chinese Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou 350108, China

¹ These authors contributed equally to this work.

* Corresponding authors.

E-mail addresses: 88131506@qq.com (Z. Yang), hmq1115@126.com (M. Huang), ccx6070@qq.com (C. Chen)

Abstract

Automated quantification of immunohistochemistry (IHC) supports reproducible biomarker assessment, but routine use remains limited by cellular crowding and staining variability. We present an ImageJ/Fiji-native workflow for batch quantification of nuclear and cytoplasmic H-scores without task-specific model training. It integrates StarDist nuclear detection, a Voronoi-

based Cell Distance (VCD) topology-guided segmentation strategy and user-guided adaptive calibration (Mode 2). VCD constructs Skeleton by Influence Zones (SKIZ)-derived boundaries from detected nuclei, reducing boundary fusion and spatial dilution during morphological dilation. Mode 2 uses reference regions from a representative positive-control image to derive batch-specific intensity thresholds and reduce staining-related bias. In stress tests, 99.7% of detected cell boundaries were within 2 pixels of manual ground-truth annotations. Across 752 IHC images representing diverse staining conditions, Mode 2 closely agreed with clinician-assigned H-scores (CCC = 0.997), with a mean Bland–Altman bias of −2.35 H-score units. Across two Human Protein Atlas pan-cancer cohorts, automated H-scores correlated with expert annotations for nuclear Timeless ($n = 330$; Spearman’s $\rho = 0.760$) and cytoplasmic CDK1 ($n = 290$; $\rho = 0.660$). This open-source ImageJ/Fiji plugin provides an accessible, reproducible workflow for batch IHC quantification.

Keywords: Digital pathology; Immunohistochemistry (IHC); Deep learning segmentation; Batch effects; H-Score quantification; ImageJ/Fiji plugin.

1. Introduction

Immunohistochemistry (IHC) is central to clinical histopathology, enabling visualization of protein expression for cancer diagnosis, prognostic stratification, and therapeutic decision-making[1]. Traditionally, IHC biomarker quantification, particularly for the widely used H-Score, relies on manual estimation of staining intensity and positive cell fractions[2]. However, manual evaluation remains labor-intensive, difficult to scale to large image cohorts, and subject to inter-observer variability[3,4]. Although digital pathology tools have advanced automated

analysis, single-cell quantification in dense clinical tissues remains challenging[5]. Moreover, many existing analysis workflows require substantial manual intervention for image-by-image processing and parameter adjustment, limiting their scalability for high-throughput analysis of large clinical cohorts and experimental datasets.

These analytical challenges stem from structural and technical constraints in current workflows. Traditional segmentation methods, including watershed algorithms, frequently fail in crowded cell regions where indistinct boundaries lead to cell merging artifacts[6]. This limitation is amplified during the cell expansion required for cytoplasmic marker evaluation, as unconstrained spatial growth incorporates background tissue and distorts intensity measurements. Furthermore, batch-to-batch variations in staining intensity, driven by differences in tissue processing and scanner optics, present a practical barrier[7,8]. Because conventional pipelines depend on fixed intensity thresholds, they fail when applied across variable clinical cohorts, requiring manual recalibration.

To address these limitations, we developed an open-source ImageJ/Fiji plugin[9] combining deep-learning nuclear detection (StarDist)[10] with a Voronoi-based topological strategy for IHC quantification. While deep-learning frameworks such as Cellpose[11] have demonstrated high segmentation accuracy, their deployment often requires dedicated computing environments or GPU hardware. In contrast, standard pathology software such as QuPath commonly relies on watershed-based cell expansion[6]. Our ImageJ-native pipeline offers an accessible, graphical user interface that maintains accurate cell boundary delineation without requiring programming or specialized hardware. In parallel, an adaptive calibration workflow

adjusts thresholds using batch-specific reference intensity profiles, mitigating batch effects without rigid cutoffs and enabling consistent batch analysis across large image datasets.

Here, we systematically evaluated the technical performance and practical utility of this framework. In a clinician-validation cohort comprising 752 IHC images representing diverse staining conditions, the adaptive Mode 2 workflow showed close agreement with clinician-assigned H-scores (Lin’s concordance correlation coefficient, 0.997). We further evaluated the pipeline using 620 tissue specimens from the Human Protein Atlas (HPA)[12], comprising the nuclear marker Timeless ($n = 330$) and the cytoplasmic marker CDK1 ($n = 290$) across multiple tumor types. Automated H-scores correlated with HPA expert annotations for both nuclear (Spearman’s $\rho = 0.760$) and cytoplasmic ($\rho = 0.660$) targets. Together, these results support the framework as an accessible and reproducible tool for batch quantitative IHC analysis in digital pathology.

2 Methods

2.1 Dataset overview and ground truth establishment

IHC datasets used for algorithm development, optimization, and validation were retrieved from the HPA database[12,13]. Different evaluation cohorts were established for hyperparameter optimization, segmentation benchmarking, contrast calibration, boundary stability analysis, clinical validation, and pan-cancer validation. The composition, experimental purpose, and sample size of each cohort are summarized in **Table 1**.

Table 1. Summary of evaluation cohorts used for algorithm development, optimization, and

validation.

Cohort	Purpose	Sample Size
Hyperparameter optimization	Threshold optimization	25 FOVs
Segmentation benchmark	GT evaluation	30 FOVs
Contrast calibration	Stain variability	10 FOVs
Boundary stability	Dilation analysis	10 FOVs
Staining-batch robustness	Phenotypic and H-score consistency	5 IHC images
Clinician-validation cohort	Agreement assessment across diverse staining conditions	752 IHC images
Pan-cancer validation	Timeless/CDK1	620 specimens

Ground truth (GT) segmentation masks and cell annotations for the segmentation benchmark cohort were independently established by a panel of three board-certified clinical pathologists through consensus review. For pan-cancer validation, a cohort of 620 tissue specimens was compiled from the HPA, comprising the nuclear marker Timeless ($n = 330$) and the cytoplasmic marker CDK1 ($n = 290$) across diverse tumor types. Official HPA categorical expression grades (Negative, Low, Medium, and High) served as the reference standard for downstream correlation analyses.

2.2 Workflow architecture and adaptive segmentation strategy

An ImageJ/Fiji-native computational pipeline was developed for automated batch analysis of IHC images. The workflow integrates color deconvolution, deep-learning-based nuclear

detection, Voronoi-constrained cell segmentation, adaptive intensity thresholding, and automated H-Score calculation.

Input RGB images first undergo color deconvolution[14,15] to separate hematoxylin-stained nuclei (Channel 1) from DAB-stained biomarker signals (Channel 2). Nuclear instances are detected from Channel 1 using StarDist[10], followed by area-based filtering (5th – 95th percentiles) to remove non-cellular objects and segmentation artifacts.

Cell boundaries are subsequently generated on Channel 2 using a Voronoi-based Cell Distance (VCD) algorithm[16]. The resulting Voronoi partitions are intersected with dilated nuclear masks to constrain cell expansion at the midpoints between adjacent nuclei, thereby preventing cell merging and minimizing signal contamination from neighboring cells.

Following cell boundary reconstruction, the mean grayscale intensity of the DAB channel within each segmented cell was quantified and classified into four staining categories (0, 1+, 2+, and 3+). Mode 1 applied fixed grayscale thresholds of 60, 120, and 180[17]. In Mode 2, the user first selected a representative positive-control image from the staining batch and interactively sampled regions representing negative/background, weakly positive (1+), moderately positive (2+), and strongly positive (3+) staining. The mean grayscale intensity was calculated for each reference category. The three classification thresholds were defined as the midpoints between the mean intensities of adjacent categories: $T_{3+ / 2+} = (I_3^+ + I_2^+)/2$, $T_{2+ / 1+} = (I_2^+ + I_1^+)/2$, and $T_{1+ / 0} = (I_1^+ + I_0)/2$. The resulting thresholds were subsequently applied to all images within the corresponding batch.

Automated H-Scores were calculated according to the following equation:

$$\text{H-Score} = 1 \times P_1 + 2 \times P_2 + 3 \times P_3$$

where P_1 , P_2 , and P_3 represent the percentages of weakly (1+), moderately (2+), and strongly (3+) stained cells, respectively.

2.3 Evaluation of segmentation performance and cellular density stratification

2.3.1 Pixel-level evaluation metrics and density stratification

Segmentation accuracy was evaluated by comparing automated segmentation masks with manual GT masks using a 2-pixel boundary tolerance. Segmented regions were classified as true positives (TP), false positives (FP), or false negatives (FN) according to their overlap with the GT masks. Performance was quantified using five commonly adopted metrics: Precision [$TP / (TP + FP)$], Recall [$TP / (TP + FN)$], F1 score [$2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$], Dice coefficient [$2 \times TP / (2 \times TP + FP + FN)$], and Intersection over Union (IoU) [$TP / (TP + FP + FN)$]

To evaluate segmentation robustness across different cellular densities, the segmentation benchmark cohort (**Table 1**) was stratified into sparse-, moderate-, and dense groups based on nuclear density. Automated cell counts were subsequently compared with the corresponding GT annotations within each density category.

2.3.2 Morphological dilation and boundary collision modeling

To evaluate segmentation stability during cell boundary expansion, the same StarDist-derived nuclear masks were used as the initial input for both workflows. Cytoplasmic regions were simulated by incrementally dilating the nuclear masks from 0 to 10 pixels in 1-pixel increments. In the conventional workflow, touching regions were processed using ImageJ’s binary watershed operation without the VCD constraint. In the proposed workflow, the dilated masks were intersected with the SKIZ-derived VCD partitions to restrict expansion at the boundaries between adjacent nuclei. Identical initial masks, dilation radii, and downstream measurements were used for both workflows. Changes in single-cell area, Precision, Recall, and F1 score were evaluated at each dilation radius.

2.4 Hyperparameter optimization and computational performance

2.4.1 Dual-parameter grid search for detection thresholds

Detection probability thresholds and non-maximum suppression (NMS) Intersection over Union (IoU) cutoffs were systematically optimized using the hyperparameter optimization cohort (**Table 1**) through a two-dimensional grid search[5]. Candidate values ranging from 0.10 to 0.50, with 0.05 increments, were evaluated against expert-annotated GT nuclear coordinates. For each parameter combination, automated cell-counting F1 scores and per-frame inference time were recorded. The resulting bivariate performance heatmaps were used to identify parameter combinations that achieved favorable trade-offs between detection accuracy and computational efficiency.

2.4.2 Spatial resolution scaling and morphological nuclear filtering

To evaluate algorithm scalability across different image resolutions, the hyperparameter optimization cohort (**Table 1**) was systematically resampled into square image patches ranging from 500×500 to 2500×2500 pixels. Segmentation accuracy and per-patch processing time were evaluated at each resolution to determine the optimal balance between computational efficiency and segmentation performance. To reduce non-cellular background objects and exclude abnormal nuclear clusters, a morphological filtering step was applied based on nuclear area. Objects with areas outside the 5th and 95th percentiles of the annotated nuclear size distribution were removed prior to downstream analysis[18].

2.4.3 Dynamic contrast calibration for inter-sample stain variability

To evaluate the robustness of the proposed calibration strategy under varying staining conditions, the contrast calibration cohort (**Table 1**) was stratified into normal (baseline), dark (under-stained), and bright (over-stained) groups according to mean image luminance[14]. A secondary grid search was performed over the minimum (0–40) and maximum (210–255) pixel intensity thresholds for linear contrast stretching. For each staining group, combinations of minimum and maximum intensity thresholds were evaluated to determine the optimal parameters for adaptive contrast calibration prior to nuclear detection and cell segmentation.

2.5 Clinical validation of phenotypic subtyping and robustness

2.5.1 Phenotypic subtyping and automated H-Score quantification

Biomarker expression was quantified within the segmented cell boundaries based on DAB optical density[19]. Individual cells were classified into four staining categories: 0 (negative),

1+ (weakly positive), 2+ (moderately positive), and 3+ (strongly positive)[2]. The proportions of cells in each category were subsequently used to calculate the H-Score for each field of view according to the predefined scoring formula, with H-Scores ranging from 0 to 300.

2.5.2 Assessment of batch consistency and boundary disturbance sensitivity

To evaluate H-Score stability under different staining conditions, five source IHC images were each evaluated under Light, Normal, and Dark staining conditions, resulting in three paired measurements per source image. Single-cell phenotypic subtyping and H-Score quantification were performed using Mode 1 and Mode 2. Because the three staining-condition measurements originated from the same source images, the data were treated as paired repeated measurements.

To evaluate the sensitivity of H-Score quantification to boundary segmentation errors, images from the boundary stability cohort (**Table 1**) were analyzed by progressively expanding nuclear masks using dilation radii ranging from 0 to 10 pixels in 1-pixel increments. Changes in cell subtype proportions (0, 1+, 2+, and 3+) together with corresponding H-Score trajectories were analyzed under both analytical frameworks.

2.6 Expert agreement assessment and pan-cancer validation

2.6.1 Expert agreement and quantitative bias evaluation

Agreement between automated and clinician-assigned H-scores was evaluated using 752 IHC images representing diverse staining conditions from the clinician-validation cohort (**Table 1**). Each IHC image was treated as the analytical unit. Complete specimen- and patient-level

identifiers were unavailable for this cohort; therefore, potential dependence among images originating from the same specimen could not be fully evaluated. Reference H-scores were independently assigned by three clinicians without access to the computational predictions. The clinician-derived reference score for each image was calculated as the mean of the three individual scores. Agreement between Mode 1 or Mode 2 measurements and the clinician-derived reference scores was quantified using Lin's concordance correlation coefficient (CCC). Bland–Altman analysis was used to assess systematic bias and limits of agreement across the H-score range. Differences were calculated as the automated H-score minus the clinician-assigned H-score. The mean difference represented systematic bias, and the 95% limits of agreement were calculated as the mean difference $\pm$ 1.96 standard deviations.

2.6.2 Pan-cancer cohort acquisition, data preprocessing, and correlation analysis

To evaluate performance across diverse cancer types and subcellular targets, the pan-cancer validation cohort (**Table 1**) was constructed using two tissue microarray (TMA) cohorts retrieved from the HPA database[20]. Target biomarkers were selected to represent distinct subcellular localization patterns: Timeless ($n = 330$, predominantly nuclear)[21] and CDK1 ($n = 290$, predominantly cytoplasmic)[22].

Non-evaluable specimens and damaged image artifacts were excluded prior to analysis. HPA expert annotations were converted into ordinal scores (0 = Negative, 1 = Low, 2 = Medium, and 3 = High) as clinical reference labels. Data processing and statistical analyses were performed using Python (v3.9; pandas and scipy.stats packages). Differences between adjacent HPA expression categories were evaluated using two-sided Mann–Whitney U tests. P values

were adjusted separately within the Timeless and CDK1 cohorts using the Holm method. The monotonic association between continuous automated H-Scores and ordinal HPA expression scores was assessed using Spearman's rank correlation coefficient (ρ).

2.7 Software architecture, implementation, and code availability

To improve accessibility and reproducibility for non-programming users, the automated pipeline was implemented natively within the ImageJ/Fiji (v1.54p) image processing framework using the ImageJ Macro (IJM) language[9]. Automated nuclear detection was performed using the StarDist model integrated through its official ImageJ plugin[10]. The complete workflow was packaged as a standalone Java Archive (.jar) plugin to enable execution without additional Python environments or command-line operations [23]. The plugin supports automated batch processing of multiple images, minimizing repetitive manual operations during large-scale analysis.

The plugin can be deployed by placing the compiled .jar file into the standard ImageJ/Fiji plugins directory. To promote algorithmic transparency and computational reproducibility, the executable plugin, associated IJM macro scripts, and user documentation are publicly available through a GitHub repository (<https://github.com/wangweiyu-bit/AutoIHC-Plugin>).

2.8 Statistical analysis

All statistical analyses and data visualizations were performed using Python (v3.9; scipy.stats, seaborn, and matplotlib packages). Continuous variables were presented as mean $\pm$ standard deviation (SD) for normally distributed data or median with IQR for non-normally distributed

data. Two-sided Student's t-tests and Mann–Whitney U tests were used for comparisons between two groups of parametric and non-parametric data, respectively. For comparisons among multiple groups, one-way analysis of variance (ANOVA) was performed with appropriate post hoc testing.

Agreement and correlation analyses were performed using Lin's Concordance Correlation Coefficient (CCC), Bland–Altman analysis, and Spearman's rank correlation coefficient (ρ). Statistical significance was defined as $P < 0.05$, with significance levels indicated as $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and ns (not significant).

3 Results

3.1 Development of the adaptive segmentation and analysis workflow

To overcome the limitations of fixed-parameter IHC quantification, we developed an end-to-end computational pipeline comprising four sequential modules (**Fig. 1**).

Step 1 (dual-mode calibration) processed batches of IHC images by allowing users to select either predefined empirical thresholds or interactive adaptive boundaries (**Fig. 1a**).

Step 2 comprised image preprocessing, nuclear detection, comparator visualization, and morphological filtering (**Fig. 1b–f**). Color deconvolution separated the input image into hematoxylin (Colour-1) and DAB (Colour-2) channels, after which StarDist generated the initial nuclear instance masks (**Fig. 1b**). A conventional watershed result was included to illustrate boundary-fusion artifacts in crowded regions (**Fig. 1c**). Users could adjust image resolution, contrast, detection probability, and non-maximum suppression parameters through

the preprocessing controls (**Fig. 1d**). Nuclear-area thresholds were subsequently applied to remove small debris and abnormally large objects, producing the refined nuclear masks used for downstream analysis (**Fig. 1e, f**).

Step 3 (adaptive cellular segmentation) reconstructed cell boundaries using a Colour-2-guided strategy designed for densely packed tissues (**Fig. 1g, h**). A VCD algorithm generated Voronoi partitions, which were intersected with dilated nuclear masks to produce non-overlapping single-cell boundaries (**Fig. 1g**).

Step 4 (single-cell phenotypic subtyping and quantitative visualization) quantified DAB optical density within each segmented cell and classified cells into four expression categories. The pipeline automatically generated color-coded subtype maps together with quantitative measurements, enabling downstream analyses of cell area, cell subtype proportions, and H-Score distributions (**Fig. 1i, j**).

3.2 Superior pixel-level accuracy of the adaptive segmentation method

3.2.1 Performance comparison under varying cellular densities

The adaptive segmentation framework was compared with a conventional watershed-based approach across tissue sections with varying cellular densities (**Fig. 2a–f**). In moderately and highly dense tissue regions, the conventional watershed algorithm produced boundary fusion artifacts that merged adjacent nuclei into single structures (**Fig. 2a**). Quantitative cell counting showed that the adaptive framework generated cell counts consistent with manual GT annotations, whereas the conventional approach underestimated cell numbers in dense tissues

because of boundary fusion (**Fig. 2b**).

Segmentation quality was further evaluated using pixel-level metrics, including Recall, Precision, F1 score, Dice coefficient, and IoU, by comparison with manual GT masks (**Fig. 2c**). The adaptive framework consistently achieved higher Recall, Precision, and F1 scores than the conventional approach (**Fig. 2d**), together with improved Dice and IoU values (**Fig. 2e**). Cumulative boundary error analysis showed that 99.7% of cell borders detected by the adaptive framework fell within a 2-pixel tolerance of the manual GT, whereas considerably larger boundary deviations were observed for the conventional watershed method (**Fig. 2f**).

3.2.2 Robustness against morphological dilation and boundary preservation

The robustness of the segmentation framework during morphological expansion was evaluated using a boundary collision model with nuclear dilation radii ranging from 0 to 10 pixels (**Fig. 2g–k**). The Voronoi-guided algorithm constructed a SKIZ boundary equidistant from adjacent nuclei, thereby reducing boundary fusion during cell expansion (**Fig. 2g**).

Visual inspection showed that merging artifacts appeared in the conventional watershed algorithm after only 1 pixel of dilation and progressively increased until individual cell boundaries were no longer distinguishable at a dilation radius of 10 pixels (**Fig. 2h**). In contrast, the VCD method preserved cell morphology throughout the entire dilation range. Quantitatively, boundary fusion in the watershed method increased individual cell areas to more than five times their initial values (**Fig. 2i**), whereas cell area distributions produced by the VCD method remained largely unchanged. As the dilation radius increased from 0 to 10

pixels, Recall, Precision, and F1 scores for the watershed method progressively declined toward zero, whereas the VCD method maintained values close to 1.0 (**Fig. 2j**). Correlation analysis further showed that Dice coefficients remained consistently high across the full range of cell areas for the VCD method (**Fig. 2k**).

3.3 Hyperparameter optimization and computational robustness

3.3.1 Grid search optimization of core detection parameters

Hyperparameter sensitivity and computational efficiency of the StarDist detection module were evaluated using a dual-parameter grid search (**Fig. 3a–e**). Low probability thresholds (e.g., `probThresh` = 0.1) resulted in false-positive over-segmentation artifacts (**Fig. 3a**). Two-dimensional heatmap analysis identified an optimal accuracy range for `probThresh` between 0.3 and 0.5 (**Fig. 3b**), while processing speed peaked between 0.35 and 0.45 (**Fig. 3c**). Increasing `probThresh` from 0.1 to 0.5 progressively reduced false-positive detections (**Fig. 3d**) and shortened per-frame inference time (**Fig. 3e**). In contrast, varying the non-maximum suppression threshold (`nmsThresh`) from 0.1 to 0.5 produced minimal changes in either segmentation accuracy or computational efficiency (**Fig. 3d, e**).

3.3.2 Balancing spatial resolution and processing efficiency

The effects of image resolution on segmentation performance and computational efficiency were evaluated using image patches of 500×500 , 1000×1000 , and 2000×2000 pixels (**Fig. 3f–j**). Increasing image resolution to 2000×2000 pixels amplified background noise, resulting in false-positive detections and localized over-segmentation (**Fig. 3f, i**). Processing time

increased with image size, exceeding 15 s per patch at the highest resolution (**Fig. 3h**). By comparison, 1000×1000 pixels provided high segmentation accuracy while requiring only ~ 2.5 s per patch (**Fig. 3h, i**). Applying a morphological size-exclusion filter at this resolution removed 2.5% of microscopic debris and 2.5% of non-cellular aggregates while retaining 95% of target nuclei (**Fig. 3j**).

3.3.3 Robustness across variable illumination conditions

The robustness of the segmentation framework under different staining conditions was evaluated using Dark, Normal, and Bright image sets (**Fig. 3g, k**). Applying the default contrast settings (0/255) to uncalibrated images reduced nuclear detection performance (**Fig. 3g**). Adaptive contrast calibration restored segmentation accuracy across all staining conditions (**Fig. 3k**). Using intensity bounds of 0/255 for Normal images achieved an accuracy of $98.9 \pm 9.4\%$, whereas optimized bounds of 0/210 and 30/255 yielded accuracies of $99.5 \pm 2.2\%$ for Dark images and $99.0 \pm 6.9\%$ for Bright images, respectively (**Fig. 3k**).

3.4 Clinical robustness against staining variations and morphological scaling

3.4.1 Mitigation of batch-specific staining variations via interactive calibration

The stability of quantitative image-level scoring under different staining conditions was evaluated using Light, Normal, and Dark staining cohorts (**Fig. 4a–e**). Global intensity distributions differed among the three staining groups (**Fig. 4b**). Applying fixed intensity thresholds (60, 120, and 180) in the empirical mode (Mode 1) resulted in systematic misclassification in non-standard batches, with most cells classified as negative in Light images

and as strongly positive (3+) in Dark images (**Fig. 4a**).

Cell subtype distributions reflected these staining-dependent shifts (**Fig. 4d**). Mode 1 showed substantial variation in cell subtype proportions across the three staining conditions (**Fig. 4d**). By contrast, the adaptive calibration mode (Mode 2) produced consistent cell subtype assignments (**Fig. 4a**) and comparable subtype distributions across all staining groups (**Fig. 4e**). Correspondingly, H-Scores generated by Mode 1 decreased to approximately 50 in the Light cohort and increased to approximately 200 in the Dark cohort (**Fig. 4c**), whereas Mode 2 maintained values close to 150 across all staining conditions (**Fig. 4c**).

3.4.2 Morphological robustness of quantitative scoring

The influence of spatial mask expansion on quantitative scoring was evaluated by progressively increasing the nuclear dilation radius from 0 to 10 pixels (**Fig. 4f–i**). In Mode 1, increasing dilation progressively reduced mean DAB optical density within segmented cells, shifting strongly positive cells toward weaker staining categories (**Fig. 4f**). Consequently, the proportion of negative cells increased, whereas the proportions of positive subtypes decreased (**Fig. 4g**), resulting in a progressive decline in H-Scores with increasing dilation radius (**Fig. 4i**).

By comparison, the adaptive calibration strategy (Mode 2) maintained stable cell subtype distributions throughout the entire dilation range (**Fig. 4h**). Corresponding H-Scores also remained largely unchanged from 0 to 10 pixels of dilation (**Fig. 4i**).

3.5 Pan-cancer validation and clinical concordance

3.5.1 High concordance with clinician reference assessment

Agreement between automated and clinician-assigned H-scores was evaluated across 752 IHC images representing diverse staining conditions (**Fig. 5a–c**). Mode 2 showed close agreement with the clinician-derived reference scores, with a Lin’s concordance correlation coefficient (CCC) of 0.997, whereas Mode 1 showed lower concordance (CCC = 0.438, **Fig. 5a**).

Bland–Altman analysis further showed that Mode 2 had a small mean bias of −2.35 H-score units, indicating slight underestimation relative to the clinician-derived reference scores. By comparison, Mode 1 showed a larger negative bias of −29.68 H-score units, indicating more pronounced systematic underestimation (**Fig. 5b**). Analysis of phenotypic subpopulation distributions showed that Mode 2 more closely matched clinician annotations, whereas Mode 1 produced a higher proportion of cells classified as negative (**Fig. 5c**).

3.5.2 Pan-cancer validation across distinct subcellular biomarkers

To further evaluate pan-cancer generalizability, the pipeline was applied to an independent pan-cancer cohort from the HPA, including biomarkers with distinct subcellular localization patterns: Timeless (predominantly nuclear) and CDK1 (predominantly cytoplasmic) (**Fig. 5d–i**).

For the nuclear biomarker Timeless, automated H-Scores demonstrated progressive increases across the predefined HPA expression categories (Negative, Low, Medium, and High), indicating separation among ordinal expression groups (**Fig. 5d, e**). Across the Timeless cohort ($n = 330$), automated H-Scores showed a positive correlation with HPA expert annotations,

achieving a Spearman's rank correlation coefficient (ρ) of 0.760 (**Fig. 5f**). Similar correlation trends were observed across multiple tumor types, including melanoma, pancreatic, gastric, and lung carcinomas.

For the cytoplasmic biomarker CDK1, the pipeline maintained quantitative performance despite differences in subcellular localization and cellular signal distribution (**Fig. 5g, h**). Across the CDK1 cohort ($n = 290$), automated H-Scores correlated with HPA reference categories, with a Spearman's rank correlation coefficient (ρ) of 0.660 (**Fig. 5i**). Comparable associations were observed across several cancer types, including thyroid, gastric, and pancreatic carcinomas.

4. Discussion

Automated quantification of IHC represents an important component of digital pathology and precision oncology. However, the translation of computational approaches into practical quantitative workflows remains challenging due to two major technical limitations: inaccurate cell boundary reconstruction in densely packed tissues and signal instability caused by inter-sample staining variability[24]. In this study, we developed an ImageJ-native computational pipeline that integrates StarDist-based nuclear detection, a Voronoi-based Cell Distance (VCD) topological segmentation strategy, and an interactive adaptive calibration model (Mode 2) for reproducible single-cell phenotyping and H-Score quantification across heterogeneous IHC datasets.

A key advantage of this framework is its ability to preserve cellular boundaries during spatial

expansion in dense tissue environments. Conventional segmentation approaches frequently encounter difficulties in crowded regions, where indistinct cell borders result in boundary fusion and inaccurate quantification[6]. This limitation becomes more prominent during nuclear dilation-based cytoplasmic signal extraction, as uncontrolled expansion may incorporate surrounding unstained regions and dilute biomarker intensity. The proposed VCD strategy addresses this challenge by introducing a Voronoi-derived SKIZ boundary as a topological constraint, maintaining spatial separation between neighboring cells during expansion. In synthetic dilation experiments up to 10 pixels, the VCD-based approach preserved cellular morphology, maintained segmentation accuracy, and prevented progressive H-Score distortion caused by boundary collision. These properties are particularly relevant for quantitative analysis of cytoplasmic biomarkers, where accurate reconstruction of cellular territories is essential.

Beyond morphological stability, controlling staining variability is another critical requirement for robust IHC quantification. Differences in tissue processing, staining procedures and digital scanning conditions can introduce substantial intensity variation across experimental batches[25,26]. Consistent with this challenge, fixed intensity thresholds (Mode 1) were sensitive to such variation, resulting in systematic shifts in cell classification and H-score estimation. Mode 1 showed only moderate concordance with clinician-assigned scores (CCC = 0.438) and a mean bias of -29.68 H-score units. By contrast, the adaptive calibration strategy (Mode 2) adjusted intensity thresholds using batch-specific reference signals and achieved substantially closer agreement with clinician assessment across 752 IHC images representing diverse staining conditions (CCC = 0.997; mean bias, -2.35 H-score units). These findings

indicate that adaptive calibration can reduce the systematic underestimation associated with fixed thresholds under variable staining conditions. In the independent pan-cancer validation cohort, the framework also performed consistently across biomarkers with distinct subcellular localizations. Automated H-scores correlated with HPA expert annotations for both nuclear Timeless ($n = 330$; Spearman's $\rho = 0.760$) and cytoplasmic CDK1 ($n = 290$; Spearman's $\rho = 0.660$). Together, these results support the reproducibility of the framework across heterogeneous tissue images and its applicability to quantitative analysis of both nuclear and cytoplasmic IHC biomarkers.

Another important feature of this work is the emphasis on accessibility and reproducibility. Recent studies have also demonstrated the value of open-source computational workflows for scalable quantitative IHC analysis, although their applications have generally remained disease- or tissue-specific[27]. Although deep-learning-based pathology analysis methods have achieved impressive performance, their implementation often requires specialized computational environments, programming expertise, or dedicated hardware resources. By integrating the workflow into the ImageJ/Fiji ecosystem and packaging the complete pipeline as an open-source Java plugin, our framework enables researchers and pathology laboratories to perform automated IHC analysis using a widely adopted image-processing platform. This design reduces technical barriers and facilitates reproducible quantitative analysis without requiring complex computational infrastructure.

Several limitations should be acknowledged. First, although Timeless and CDK1 were selected to represent nuclear and cytoplasmic staining patterns, respectively, validation was limited to

two representative biomarkers. Future studies should expand evaluation to additional biomarkers with diverse expression levels, subcellular distributions, and clinical applications to further establish the generalizability of the framework. Second, the current validation cohort was primarily derived from publicly available cancer datasets, which limits assessment in specialized disease contexts. As our research focus gradually extends from tumor-related pathology toward bone tumors and other bone-associated diseases, future validation will incorporate orthopedic-related specimens, including bone tumor and musculoskeletal disease cohorts, to evaluate clinical applicability in these settings. In addition, while the current framework focuses on brightfield chromogenic IHC, adaptation to multiplex immunofluorescence and spatial omics platforms will require additional optimization of spectral unmixing, cell segmentation, phenotyping, and data-management procedures[28,29]. Future developments may further integrate advanced computational approaches, including graph neural networks (GNNs), to model spatial cellular interactions and extract higher-order tissue organization features[30]. In addition, agreement was assessed at the image level because complete specimen-level identifiers were unavailable for the clinician-validation cohort. The reported estimates may therefore not fully account for potential correlations among images originating from the same specimen.

In conclusion, this ImageJ-native computational framework provides an accessible and reproducible solution for automated IHC quantification by addressing two major sources of analytical variability: spatial boundary distortion and staining-dependent intensity fluctuation. By integrating deep-learning-assisted nuclear detection, topology-guided cellular segmentation, and adaptive signal calibration, the proposed pipeline facilitates standardized biomarker

evaluation, high-throughput analysis, and future translational applications in digital pathology research.

5. Conclusions

We developed an ImageJ-native plugin for automated IHC quantification that integrates StarDist-based nuclear detection, a VCD topology-guided segmentation strategy and an interactive adaptive calibration model. The resulting pipeline enabled single-cell phenotypic classification and H-score quantification across varying tissue densities and staining conditions, while supporting automated batch processing of large-scale IHC image datasets within ImageJ/Fiji. In the clinician-validation cohort comprising 752 IHC images representing diverse staining conditions, Mode 2 showed close agreement with clinician-assigned H-scores (CCC = 0.997), with a mean Bland–Altman bias of -2.35 H-score units. Independent pan-cancer validation further showed significant correlations between automated H-scores and HPA expert annotations for both nuclear Timeless ($n = 330$; Spearman's $\rho = 0.760$) and cytoplasmic CDK1 ($n = 290$; Spearman's $\rho = 0.660$). Together, these results demonstrate the performance of the plugin in batch-enabled quantitative analysis of nuclear and cytoplasmic IHC biomarkers.

Acknowledgements

We acknowledge the Human Protein Atlas (HPA, <https://www.proteinatlas.org>) for providing publicly available immunohistochemistry image datasets used in the pan-cancer validation of this study.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 82274080], the Natural Science Foundation of Fujian Province [grant number 2025J01306], the Joint Funds for the Innovation of Science and Technology of Fujian Province [grant numbers 2025Y9519 and 2024Y9588], and the Fujian Provincial Health Commission Science and Technology Plan Project [grant number 2025QNA097]. The funding sources had no involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Data availability statement

The executable ImageJ plugin, source macros, and user documentation are publicly available at <https://github.com/wangweiyu-bit/AutoIHC-Plugin>. The immunohistochemistry images analyzed during the pan-cancer validation were retrieved from the publicly accessible HPA database (<https://www.proteinatlas.org>).

Data and Image Source Transparency Statement

The authors explicitly declare that all clinical images utilized in this study were exclusively sourced from the aforementioned public repository. Because the HPA database is a publicly accessible resource, the same or similar images may appear in independent studies using this database. Such overlap reflects the nature of shared public datasets and does not indicate image duplication or manipulation.

Ethics statement

This study focused exclusively on computational software development and the retrospective analysis of publicly available, fully anonymized image datasets from the Human Protein Atlas. According to institutional guidelines, this study did not require ethical approval because it involved only publicly available, fully anonymized datasets without direct involvement of human participants or identifiable information.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

W.W. contributed to conceptualization, methodology, software development, algorithm architecture, formal analysis, investigation, data curation, visualization, writing the original draft, reviewing and editing the manuscript, and funding acquisition. M.L. contributed to validation, resources, data curation, and visualization. F.Z. contributed to validation and data curation. Z.Y. contributed to resources and funding acquisition. M.H. contributed to supervision, project administration, and funding acquisition. C.C. contributed to supervision and funding acquisition. All authors reviewed and approved the final manuscript.

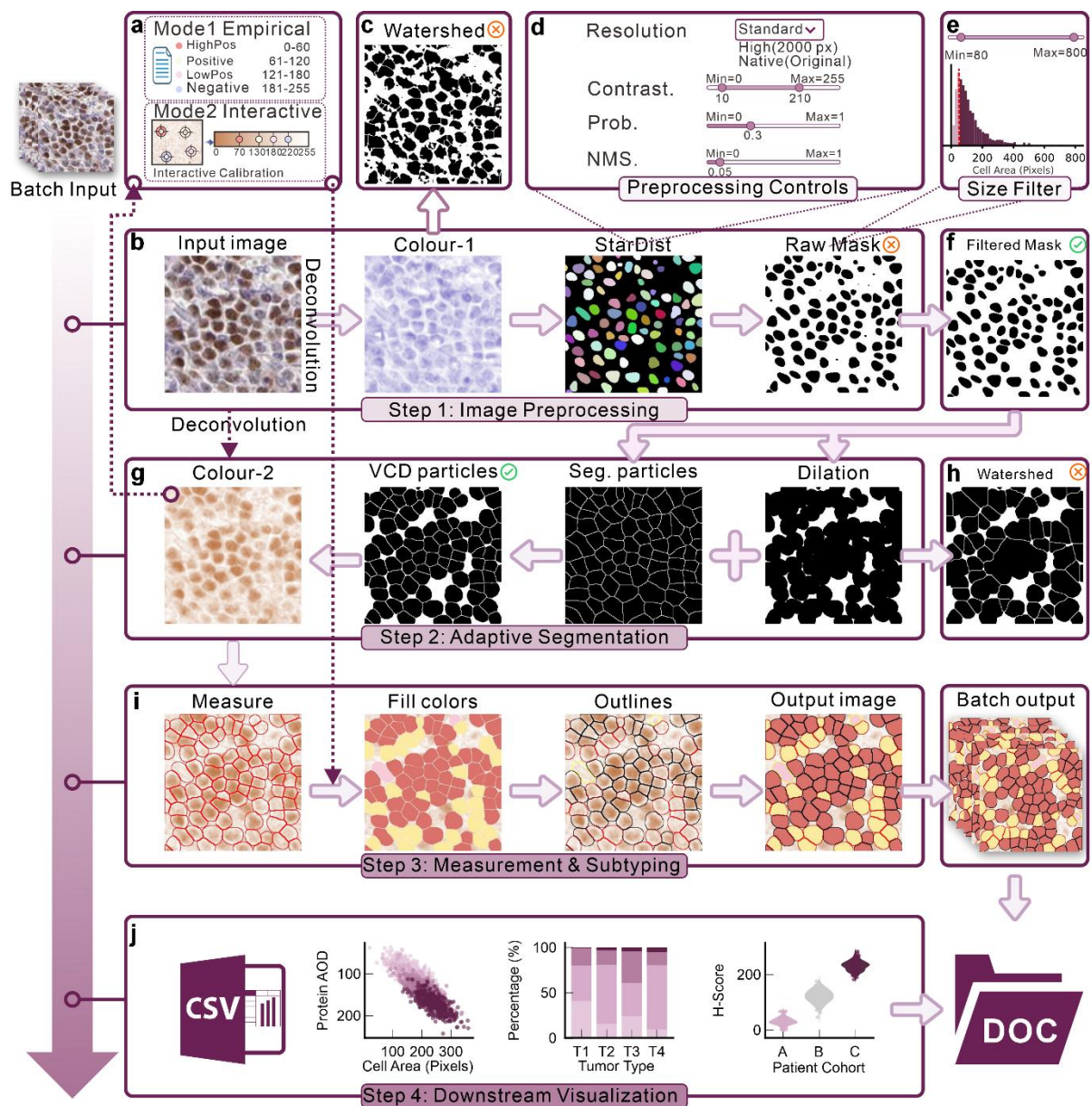
References

- 1 Matos, L. L., Trufelli, D. C., de Matos, M. G. & da Silva Pinhal, M. A.

- Immunohistochemistry as an important tool in biomarkers detection and clinical practice. *Biomark Insights* **5**, 9-20, doi:10.4137/bmi.s2185 (2010).
- 2 Fedchenko, N. & Reifenrath, J. Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue - a review. *Diagn Pathol* **9**, 221, doi:10.1186/s13000-014-0221-9 (2014).
- 3 Ram, S. *et al.* Pixelwise H-score: A novel digital image analysis-based metric to quantify membrane biomarker expression from immunohistochemistry images. *PLoS One* **16**, e0245638, doi:10.1371/journal.pone.0245638 (2021).
- 4 Kirkegaard, T. *et al.* Observer variation in immunohistochemical analysis of protein expression, time for a change? *Histopathology* **48**, 787-794, doi:10.1111/j.1365-2559.2006.02412.x (2006).
- 5 Graham, S. *et al.* Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* **58**, 101563, doi:10.1016/j.media.2019.101563 (2019).
- 6 Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878, doi:10.1038/s41598-017-17204-5 (2017).
- 7 Tellez, D. *et al.* Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med Image Anal* **58**, 101544, doi:10.1016/j.media.2019.101544 (2019).
- 8 Howard, F. M. *et al.* The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nat Commun* **12**, 4423, doi:10.1038/s41467-021-24698-1 (2021).
- 9 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
- 10 Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-Convex Polygons. *Springer, Cham* **11071**, 265-273, doi:10.1007/978-3-030-00934-2_30 (2018).
- 11 Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100-106, doi:10.1038/s41592-020-01018-x (2021).
- 12 Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).
- 13 Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, doi:10.1126/science.aan2507 (2017).
- 14 Vahadane, A. *et al.* Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Trans Med Imaging* **35**, 1962-1971, doi:10.1109/TMI.2016.2529665 (2016).
- 15 Ruifrok, A. C. & Johnston, D. A. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol* **23**, 291-299 (2001).
- 16 Legland, D., Arganda-Carreras, I. & Andrey, P. MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics* **32**, 3532-3534, doi:10.1093/bioinformatics/btw413 (2016).
- 17 Varghese, F., Bukhari, A. B., Malhotra, R. & De, A. IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. *PLoS One* **9**, e96801, doi:10.1371/journal.pone.0096801

(2014).

- 18 McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol* **16**, e2005970, doi:10.1371/journal.pbio.2005970 (2018).
- 19 Wen, Z. *et al.* Deep Learning-Based H-Score Quantification of Immunohistochemistry-Stained Images. *Mod Pathol* **37**, 100398, doi:10.1016/j.modpat.2023.100398 (2024).
- 20 Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, doi:10.1126/science.aal3321 (2017).
- 21 Bianco, J. N. *et al.* Overexpression of Claspin and Timeless protects cancer cells from replication stress in a checkpoint-independent manner. *Nat Commun* **10**, 910, doi:10.1038/s41467-019-08886-8 (2019).
- 22 Asghar, U., Witkiewicz, A. K., Turner, N. C. & Knudsen, E. S. The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nat Rev Drug Discov* **14**, 130-146, doi:10.1038/nrd4504 (2015).
- 23 Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529, doi:10.1186/s12859-017-1934-z (2017).
- 24 Niazi, M. K. K., Parwani, A. V. & Gurcan, M. N. Digital pathology and artificial intelligence. *Lancet Oncol* **20**, e253-e261, doi:10.1016/S1470-2045(19)30154-8 (2019).
- 25 van der Laak, J., Litjens, G. & Ciompi, F. Deep learning in histopathology: the path to the clinic. *Nat Med* **27**, 775-784, doi:10.1038/s41591-021-01343-4 (2021).
- 26 Zarella, M. D. *et al.* A Practical Guide to Whole Slide Imaging: A White Paper From the Digital Pathology Association. *Arch Pathol Lab Med* **143**, 222-234, doi:10.5858/arpa.2018-0343-RA (2019).
- 27 Ding, Y. *et al.* An open source pipeline for quantitative immunohistochemistry image analysis of inflammatory skin disease using artificial intelligence. *J Eur Acad Dermatol Venereol* **37**, 605-614, doi:10.1111/jdv.18726 (2023).
- 28 Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat Methods* **18**, 9-14, doi:10.1038/s41592-020-01033-y (2021).
- 29 Taube, J. M. *et al.* Society for Immunotherapy of Cancer: updates and best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) image analysis and data sharing. *J Immunother Cancer* **13**, doi:10.1136/jitc-2024-008875 (2025).
- 30 Brussee, S., Buzzanca, G., Schrader, A. M. R. & Kers, J. Graph neural networks in histopathology: Emerging trends and future directions. *Med Image Anal* **101**, 103444, doi:10.1016/j.media.2024.103444 (2025).



594 **Fig. 1 | Schematic overview of the adaptive segmentation and analysis pipeline.**

595 **a**, Dual-mode calibration module enabling predefined fixed thresholds (Mode 1) or batch-
596 specific adaptive threshold calibration (Mode 2). **b**, Color deconvolution separating input
597 images into nuclear (Color-1) and biomarker (Color-2) channels. **c**, Representative watershed-
598 based nuclear segmentation on Color-1, illustrating boundary merging artifacts in densely
599 packed regions. **d**, Optimization of preprocessing parameters following StarDist-based nuclear
600 detection. **e**, Morphological size-based filtering for removal of non-cellular debris and

601 abnormal objects. **f**, Refined nuclear masks generated after artifact removal. **g**, Adaptive
602 cellular segmentation strategy combining dilated nuclear masks with VCD partitions through
603 logical intersection. **h**, Comparison demonstrating cell boundary merging artifacts generated
604 by watershed-based expansion. **i**, DAB intensity extraction within Color-2 regions of interest
605 (ROIs) and color-coded single-cell phenotypic subtyping based on calibrated thresholds. **j**,
606 Automated export of quantitative measurements and downstream visualization of cell
607 morphology, phenotypic proportions, and H-Score distributions.

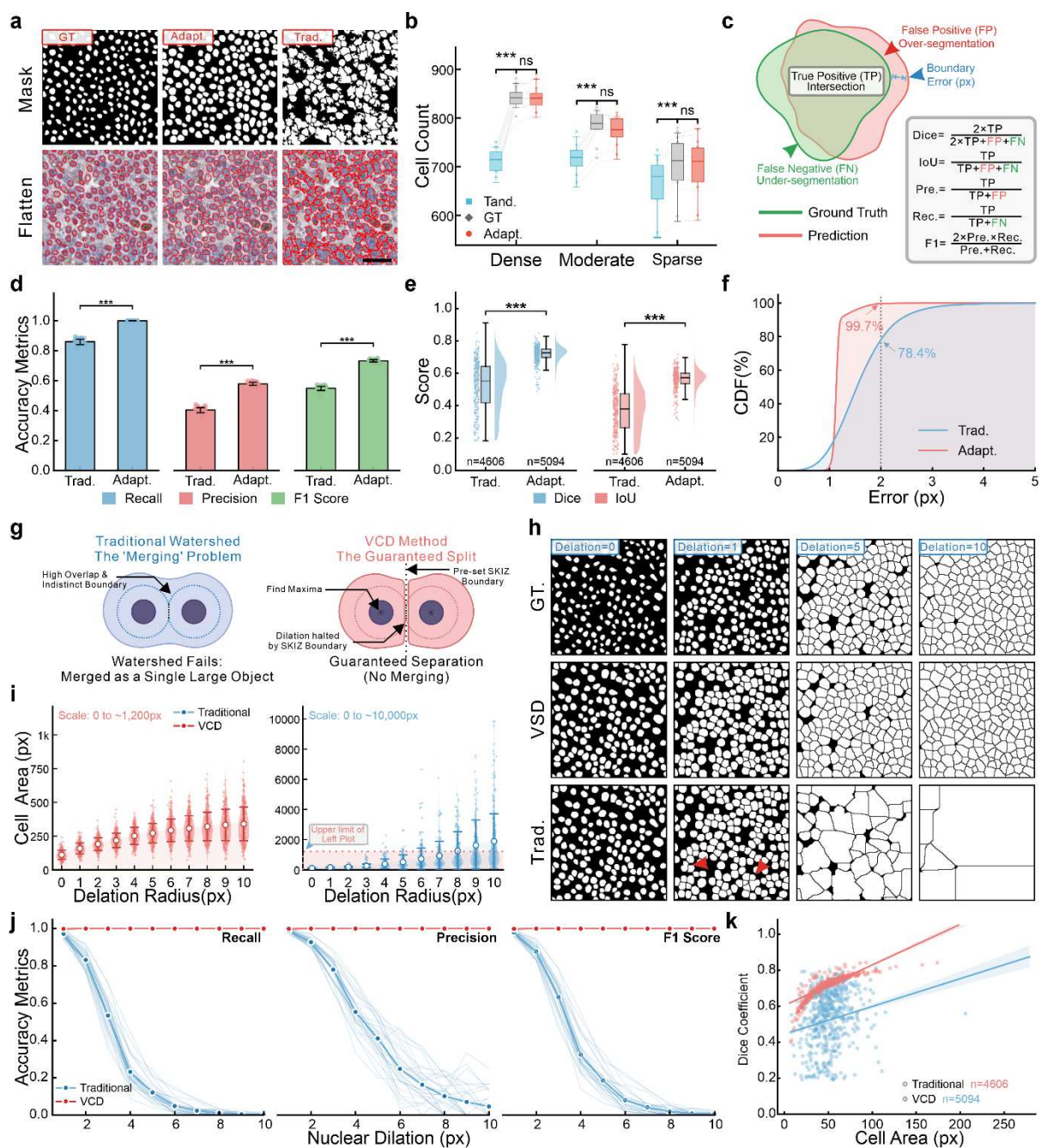


Fig. 2 | Pixel-level segmentation accuracy and robustness under morphological expansion.

a, Representative segmentation masks generated from manual GT, the adaptive VCD-based strategy, and the conventional watershed method overlaid on tissue micrographs (scale bar, 50 μ m). **b**, Automated cell counts compared across sparse, moderate, and dense cellular density cohorts. **c**, Schematic definitions of five segmentation evaluation metrics. **d**, **e**, Quantitative comparison of Recall, Precision, and F1 score (**d**), together with Dice coefficient and IoU

615 values (**e**). **f**, Cumulative distribution function (CDF) of boundary pixel localization errors. **g**,
616 Schematic comparison of watershed-induced boundary fusion and VCD-based topological
617 boundary constraints. **h**, Segmentation outcomes under progressive nuclear dilation (0–10
618 pixels). Red arrowheads indicate boundary merging artifacts. **i**, Changes in segmented cell area
619 during progressive nuclear dilation, comparing VCD stability (left) with watershed-induced
620 area expansion (right). **j**, Changes in Recall, Precision, and F1 scores across increasing nuclear
621 dilation radii. **k**, Correlation between intrinsic cell area and Dice coefficient. Data are presented
622 as mean $\pm$ SD ($n = 30$ FOVs for the segmentation benchmark in **a–f**; $n = 10$ images for the
623 boundary-stability analysis in **g–k**). Statistical significance is indicated as *** $P < 0.001$; ns,
624 not significant.

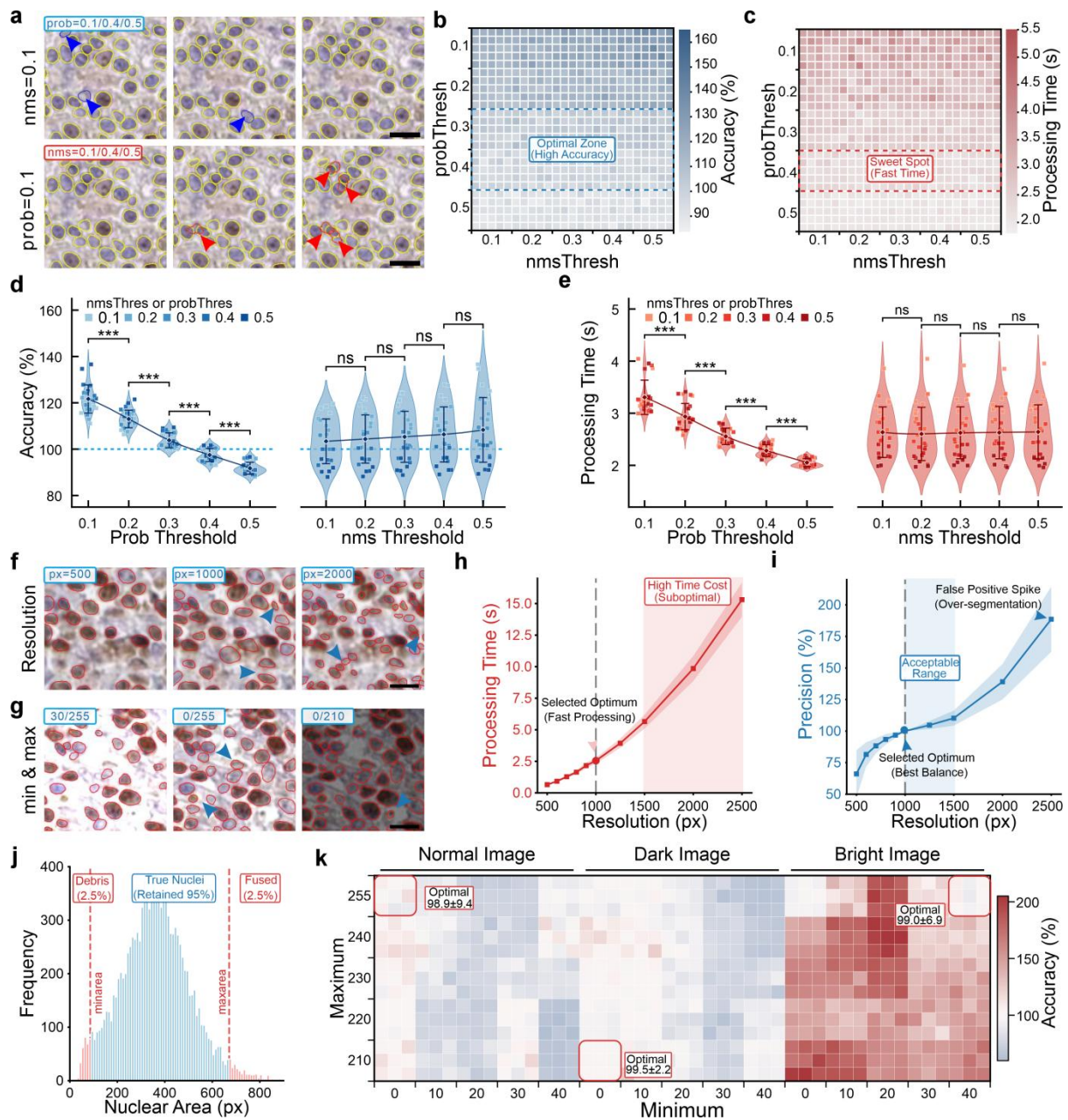


Fig. 3 | Hyperparameter optimization and computational robustness under variable imaging conditions.

a, Representative nuclear segmentation results obtained using different probability thresholds (probThresh; top row) and non-maximum suppression thresholds (nmsThresh; bottom row). Blue arrowheads indicate missed nuclear detections, and red arrowheads indicate segmentation errors associated with inappropriate threshold settings. **b**, Grid-search heatmap showing segmentation accuracy across combinations of probThresh and nmsThresh. **c**, Grid-search

heatmap showing processing time across the same parameter combinations. **d**, Distributions of
 segmentation accuracy across different probThresh and nmsThresh values. **e**, Distributions of
 processing time across different probThresh and nmsThresh values. **f**, Representative nuclear
 segmentation results at image resolutions ranging from 500×500 to 2000×2000 pixels. Blue
 arrowheads indicate false-positive detections at higher resolutions. **g**, Representative
 segmentation results obtained using different minimum and maximum intensity bounds for
 contrast adjustment. **h**, Processing time as a function of image resolution. **i**, Segmentation
 precision as a function of image resolution. **j**, Nuclear area distribution used to define the lower
 and upper size-exclusion thresholds, retaining the central 95% of detected nuclei. **k**, Grid-
 search heatmaps showing segmentation accuracy across minimum and maximum contrast
 bounds under Normal, Dark, and Bright imaging conditions. Scale bars, $25 \mu\text{m}$ in **a**, **f**, and **g**.
 Data are presented as mean $\pm$ SD ($n = 25$ images for hyperparameter and spatial-resolution
 analyses; $n = 10$ images for contrast-calibration analysis). Statistical significance is indicated
 as *** $P < 0.001$; ns, not significant.

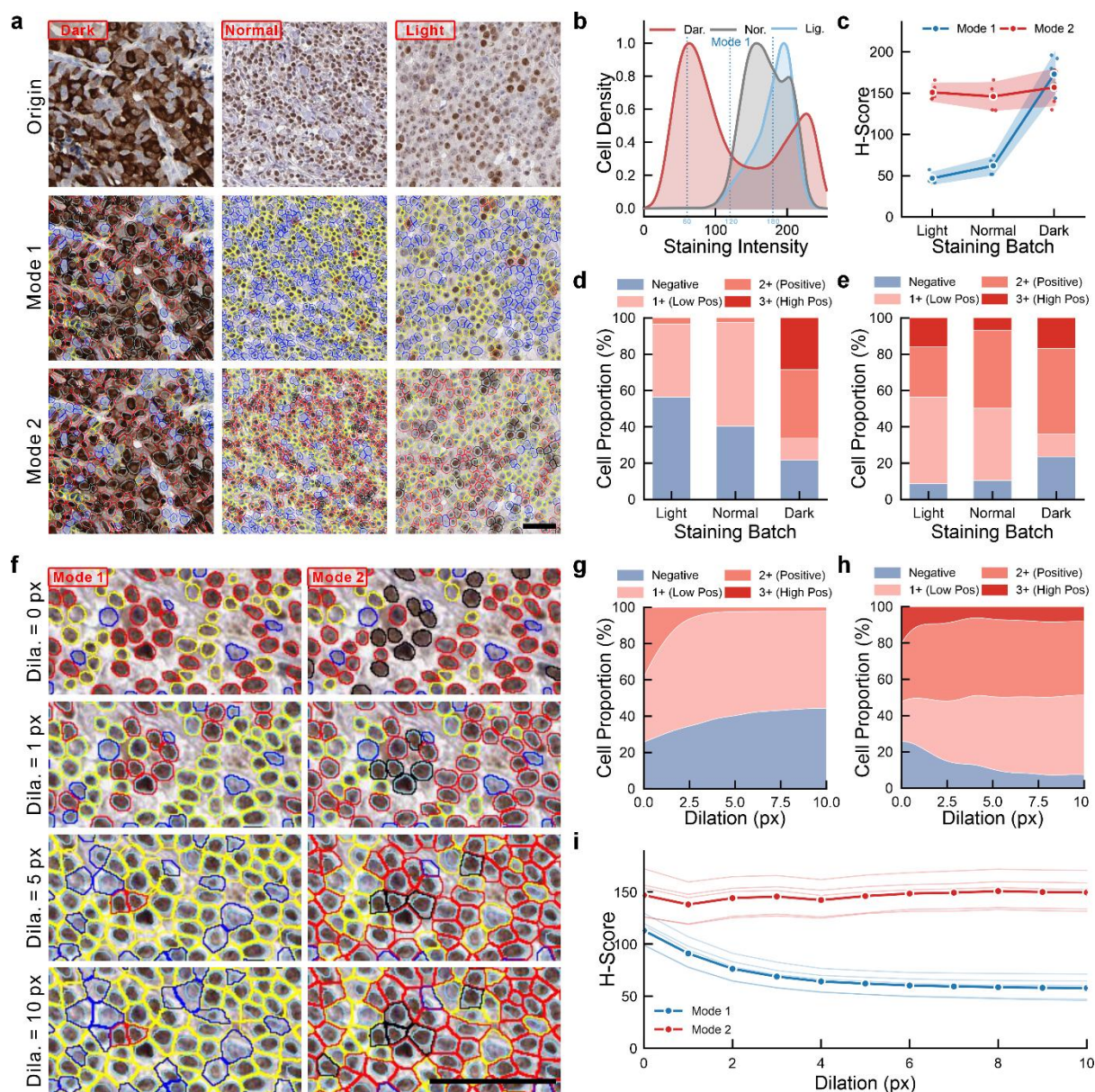


Fig. 4 | Robustness of adaptive phenotypic subtyping under staining variations and morphological boundary expansion.

a, Representative comparison of cellular phenotypic subtyping between Mode 1 (empirical fixed thresholds) and Mode 2 (interactive adaptive calibration) under different staining conditions (Light, Normal, and Dark). **b**, Density distribution curves illustrating DAB intensity variation across three staining batches, all derived from clinically confirmed positive tissue samples. Predefined static thresholds for Mode 1 (60, 120, and 180) are indicated by dashed lines. **c**, Variation in automated H-Scores across staining batches comparing Mode 1 and Mode

656 2. **d, e**, Stacked bar charts showing phenotypic subtype proportions across staining batches
 657 generated by Mode 1 (**d**) and Mode 2 (**e**). **f**, Visualization of phenotypic classification changes
 658 caused by increasing nuclear dilation radii (0, 1, 5, and 10 pixels). Mode 1 resulted in
 659 misclassification of positive cells due to incorporation of surrounding background pixels. **g, h**,
 660 Quantitative changes in phenotypic subtype proportions as a function of dilation radius for
 661 Mode 1 (**g**) and Mode 2 (**h**). **i**, H-Score trajectories as a function of dilation radius. Mode 1
 662 exhibits progressive score attenuation due to spatial dilution, whereas Mode 2 maintains stable
 663 H-Score measurements. Scale bars, 50 μm in a and f. Data are presented as mean $\pm$ SD ($n = 5$
 664 source IHC images, each evaluated under Light, Normal, and Dark conditions, for **a–e**; $n = 10$
 665 IHC images for the boundary-stability analysis in **f–i**).

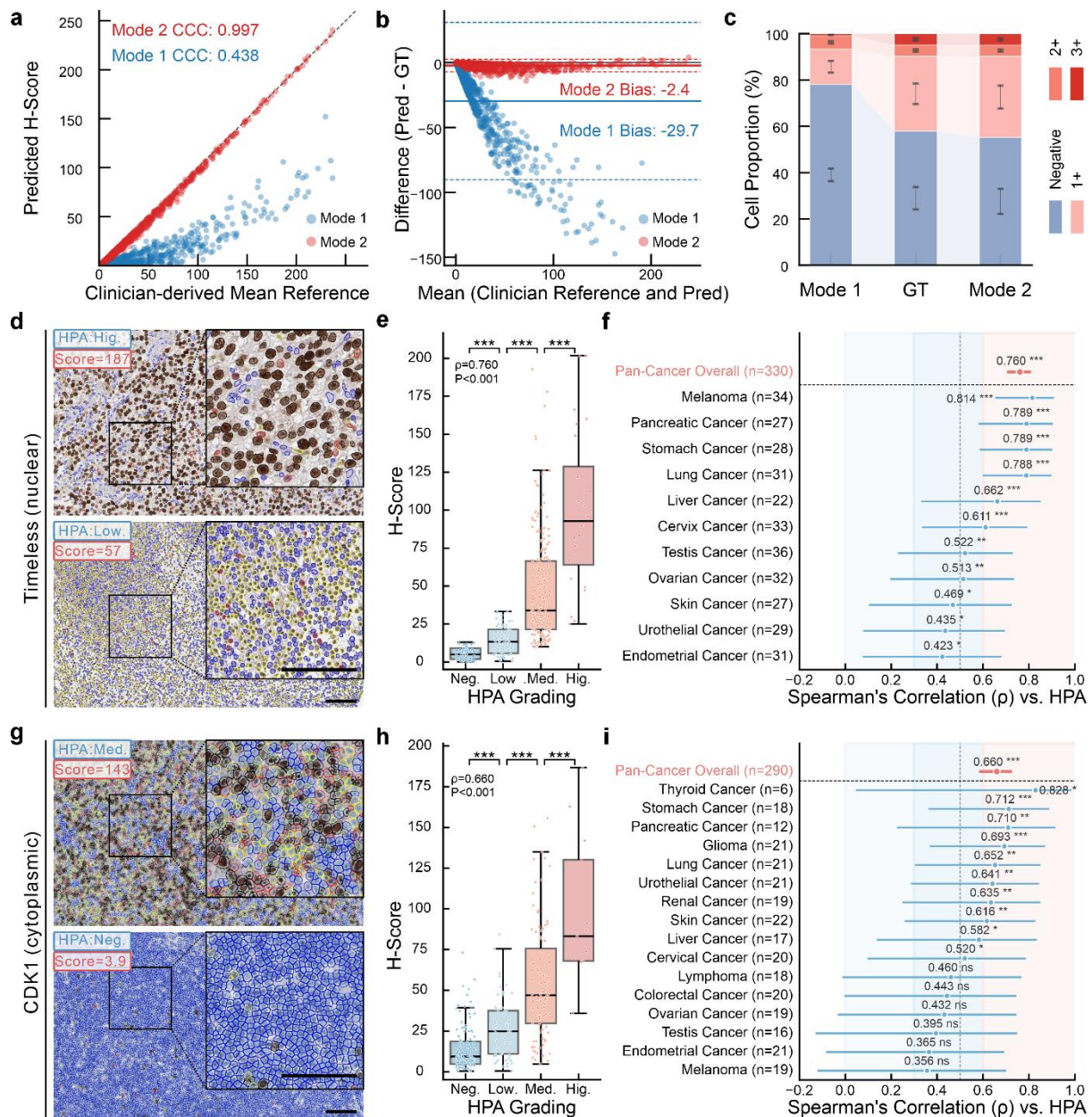


Fig. 5 | Clinical validation and pan-cancer evaluation of automated IHC quantification.

a, CCC analysis evaluating agreement between H-scores generated by Mode 1 or Mode 2 and clinician-derived mean reference scores. **b**, Bland–Altman analysis evaluating systematic bias and 95% limits of agreement between automated H-scores and clinician-derived mean reference scores. **c**, Stacked bar charts illustrating phenotypic subpopulation proportions obtained using Mode 1, Mode 2, and clinician-derived reference annotations. **d**, Representative immunohistochemistry images of the nuclear biomarker Timeless from the HPA, showing High

674 and Low expression categories. **e**, Boxplots showing automated H-Scores for Timeless across
 675 increasing HPA expression categories. **f**, Forest plot showing Spearman's rank correlation
 676 coefficients (ρ) between automated H-Scores and HPA expression annotations across the
 677 Timeless pan-cancer cohort. **g**, Representative immunohistochemistry images of the
 678 cytoplasmic biomarker CDK1 from the HPA, showing Medium and Negative expression
 679 categories. **h**, Boxplots showing automated H-Scores for CDK1 across corresponding HPA
 680 expression categories. **i**, Forest plot showing Spearman's rank correlation coefficients (ρ)
 681 between automated H-Scores and HPA expression annotations across the CDK1 pan-cancer
 682 cohort. Scale bars, 100 μ m in d and g. Data in **a–c** were obtained from 752 IHC images
 683 representing diverse staining conditions. Pan-cancer cohort evaluations incorporate $n = 330$
 684 and $n = 290$ cases for Timeless (**e, f**) and CDK1 (**h, i**), respectively. For **e** and **h**, adjacent HPA
 685 expression categories were compared using two-sided Mann–Whitney U tests with Holm
 686 correction; $***P < 0.001$. For **f** and **i**, symbols indicate two-sided P values for Spearman's
 687 correlation: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; ns, not significant.