THE LANCET
Digital Health

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Lu C, Bera K, Wang X, et al. A prognostic model for overall survival of patients with early-stage non-small cell lung cancer: a multicentre, retrospective study. *Lancet Digit Health* 2020; **2:** e594–606.

Supplementary materials S1: Mathematical description of cellular diversity computation and image analysis methods

A. Local Nuclear Graph Construction

A set of *n* segmented nuclei is denoted as *N*, where $N = \{\eta_i, p \in \{1, 2, \ldots, n\}\}$. The intuition behind Local Nuclear Graph (LNG)(1,2) is to group nuclei in local clusters for better characterizing the interactions between nuclei, locally, and to extract nuclear properties that can quantify tumor morphology efficiently. Formally, a LNG is defined as a graph $G = (N_G, E_G)$, where N_G represents the vertices of the graph (essentially the nuclei centroids), and E_G represents the set of edges connecting the nuclei within *G*. Construction of LNG can be achieved by linking nearby nuclei based on vicinity criteria as follows:

$$P(\eta_u, \eta_v) = d(\eta_u, \eta_v)^{-\alpha},$$

where η_u and η_v are two vertices/nuclei in the LNG and $d(\eta_u, \eta_v)$ represents the Euclidean distance between the two nuclei. Parameter α controls the density of the graph. Intuitively, $P(\eta_u, \eta_v)$ is defined as the probability that two nuclei having a pairwise spatial relation, i.e., the probability of two nuclei being connected in a graph. The probability of the nuclei being connected is a decaying function of the relative distance and quantifies the possibility for one of these nuclei to be grown from the other. Since this probability decreases with increase in distance, we probabilistically define an edge set *E*, such that

$$E = \{(\eta_u, \eta_v) : r < d(\eta_u, \eta_v)^{-\alpha}, \forall \eta_u, \eta_v \in N\},$$

where $r \in [0,1]$ is an empirically determined parameter (normally set r =0.2). In establishing the edges of LNG, we use a decaying probability function with an exponent of -α with $\alpha \geq 0$. The value of α determines the density of the edges in a LNG. Consequently, larger value of α produces sparser graphs. As α approaches 0, the graphs become densely connected and approach a complete graph. We fixed r=0.2, and tuned parameter α to create the subgraph. For a digitized histology image at 40x magnification, we suggest to set α=0.44 or 0.46, the configurations of different α for an example are shown in Fig. S1. As may be observed, if α is large (α>0.46), the LNG will be small and contain just a few nuclei; on the other hand, if the α is low (α<0.44), the LNG is too dense.

B. Nuclear Sub-class Co-occurrence Matrix Construction

Calculation of Nuclear features: In order to quantify the nuclei morphology, several measurements, e.g., shape and texture, based on the pre-segmented nuclei are calculated. We denote each feature extracted for a nucleus η_i as $m_j(\eta_i)$. For each nucleus, we have a set of nuclear features, $M = \{m_j(\eta_i), j \in \{1 \ldots k\}\}$. In this work, 11 nuclear morphologic features were considered, (the considered features are listed in Table

1, note that the nuclei features are not limited to these 11 features, any nuclear morphologic measurement can be used).

Co-occurring nuclear morphology matrix construction: In order to quantify the local cellular diversity between nuclear sub-groups in terms of local nuclear morphology, the co-occurrence nuclear morphology matrix is constructed for each LNG that enables the capturing the frequency of co-occurrence of diverse features for different nuclei within a LNG. Intuitively, if all the nuclei are identical in appearance, the co-occurrence matrix is a 1 x 1 matrix. On the other hand, the greater the diversity and range of attributes, the larger the co-occurrence matrix. To compute the co-occurring nuclear morphology, we discretized the nuclear morphological features $m_j(\eta_i)$ along each feature dimension such that $\overline{m_j}(\eta_i) = \omega * \left\lfloor \frac{m_j(\eta_i)}{\omega} \right\rfloor$, where ω is a quantifying factor. Intuitively, the discretization operation categorizes the nuclei into sub-classes in terms of a certain morphological feature m_j. For example, if we consider the nuclear size as a morphological feature, with ω=3, we can now categorize nuclei into three sub-classes: nuclei with large size, medium size, and small size. Fig. S2(d) and (f) show examples, in which a set of nuclei belonging to a LNG (shown in Fig. S2(b)), are categorized into three groups, in terms of nuclear size and nuclear solidity, respectively (the group numbers are indicated in the parenthesis). Note that different nuclear feature results in different nuclei sub-class groups. The element of co-occurrence matrices records the frequency of co-occurring nuclear sub-group (we set ω=5 for all the experiments).

While constructing the CM, all nuclei that are within an LNG are considered. We denote the LNG in a histology image as $G_k, k \in \{1 \ldots q\}$, where q is the total number of LNGs in the image. For each G_k in conjunction with a particular nuclear morphologic feature, we construct a c×c CM, denoted as $\mathrm{C}_{G_k}^{m_j}$. $\mathrm{C}_{G_k}^{m_j}$ which captures the co-occurrence frequency of nuclear sub-classes, in turn regularized by morphological feature m_j can be expressed as follows:

$$\mathrm{C}_{G_k}^{m_j} = \frac{1}{\sum_{a,b=1}^{\omega} \mathrm{C}_{G_k}^{m_j}(a,b)} \begin{bmatrix} \mathrm{C}_{G_k}^{m_j}(1,1) & \mathrm{C}_{G_k}^{m_j}(1,2) & \ldots & \mathrm{C}_{G_k}^{m_j}(1,\omega) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{C}_{G_k}^{m_j}(\omega,1) & \mathrm{C}_{G_k}^{m_j}(\omega,2) & \ldots & \mathrm{C}_{G_k}^{m_j}(\omega,\omega) \end{bmatrix}$$

where $\mathrm{C}_{G_k}^{m_j}(a,b) = \sum_{\eta_u,\eta_u}^{G_k} \begin{cases} 1, & \text{if } m_j(\eta_v) = a \text{ and } m_j(\eta_v) = b \\ 0, & otherwise \end{cases}$.

Fig. S2(c) and (e) show the CMs based on nuclear size and solidity of a LNG, respectively. One may observe that, by using different nuclei features, the CM is different even though they are from the same nuclear cluster. This may help to mine sub-visual information in different nuclear feature spaces.

C. Description of Cellular diversity feature Extraction

We extract a set of higher order statistics features from LNG to quantify tumor morphology. The overall flowchart for calculating cellular diversity is presented in Fig.S1 and the computational steps are summarized as follows:

Step 1: Given pre-segmented nuclei in a TMA core, we calculate a set of *K* nuclear features for each nucleus (see Table 1 for the specific features we computed).

Step 2: Calculate LNG based on nuclei locations, this is control by parameter α (set to 0.44 and 0.46, to cover different size of LNG, one example is shown in Supplementary Figure S3 with α=0.44).

Step 3: For each LNG and each nuclear feature, we compute a 2D co-occurrence matrix (CM) *C*. For example, for the nuclear feature *Area* in the *i*th LNG, we form a matrix $C_{i,area}$. The element c_{ij} in matrix $C_{i,area}$ reflects the percentage of feature quantification level *i* and *j* co-occurrence in the current LNG. Note that for a LNG we have K co-occurrence matrix, since we have K nuclear features.

Step 4: Based on each co-occurrence matrix, we calculate a set of 13 Haralick measurement (high order statistics summarized in Table 2)(3). For now, a NG has Kx13 features.

Step 5: For each TMA core, we may have N NGs, we then calculate 5 commonly used statistics, i.e., median, standard derivation, range, kurtosis, skewness, along each nuclear feature. We now have Kx13x5 features for a TMA core. In this study, K= 11, so that we have 715 features for a TMA core.

D. Classification of KRAS status from H&E stained TMAs images

A set of N=236 patients with available KRAS mutational status were identified from the patients considered in this study. This included n=60 KRAS+ and n=176 KRAS- patients. The ability of the CellDiv features to distinguish KRAS+ vs. KRAS- patients was evaluated in a classification setting using 5-fold cross-validation over 100 iterations while ensuring balanced classes distribution (4). At each run, a training fold was determined by including 80% samples from minority class, 40 samples from KRAS+ in this case, and an equal number from the majority class, i.e., 40 randomly selected samples from KRAS-. The remaining samples were used to form the testing fold. At each run, a set of top 6 discriminative CellDiv features were selected using Wilcoxon Rank Sum Test method (5) from the training fold. A Random Forests (RF) classifier was then constructed using the top features and validated on the samples in testing fold. Fig.S4C

illustrates two representative cases of KRAS+ and KRAS-, and corresponding CellDiv feature maps. One may observe that in the case of KRAS+, the CellDiv expression in terms of nuclear intensity was higher than in the KRAS- case. All the statistical analysis was performed on MATLAB R2018b platform (MathWorks, Natick, MA, USA). We got a AUC=0.63 ± 0.02 for this classification (top features shown in Table S5).

E. Scanning details of different cohorts

D_1 was scanned and digitized using an Aperio Scanscope CS whole slide imager at 20x magnification. D_2 was scanned and digitized at 20x using a Ventana iScan HT Scanner (serial #: BI15N7205). Scanner information about D_3 obtained from TCGA included Aperio files in .svs format. Finally, five 2000 x 2000 pixels image blocks at 20x magnification were extracted and the average value of features was used to represent each patient in D_3. D_4 was scanned and digitized at 40x using a Panoramic 250 Scanner. The images were down-sampled to 20x magnification for image analysis in this study.

F. Comparing the CellDiv-based model with existing models

We evaluated the CellDiv-based model with the existing hand-crafted models and deep learning (DL) based model. All comparisons and evaluations were done on the training sets to maintain consistency between the models.

Dataset Description

The early stage non-small cell lung carcinomas (NSCLC) modeling cohort comprises a total of 486 patients in the form of digitized TMA image (scanned at 20X magnification digitally) from Cleveland Clinic (D1) and Yale Medical School (D2). Long term clinical out-come was available for all patients in this cohort. Because most of existing methods in the literature involve construction of prognostic models for a binary classification setting, in this work we used a threshold of 3 years to dichotomize patients into short-term survival (<=3 years) and long-tern survival (>3 years).

Machine learning classifier construction and evaluation

A machine learning classifier - Quadratic Discriminant Analysis (QDA), was coupled with Wilcoxon rank sum test (WRST) as a feature selection method for constructing the model under 3-fold nested cross-validation (CV)(6) with 100 runs. The mean and standard deviation of area under the receiver operating characteristic (ROC) curve (AUC), accuracy, specificity, sensitivity across all iterations/runs were reported for the QDA classifier.

Comparing CellDiv with States of the Art Hand-crafted features and Deep Learning Approaches

We compared the efficacy of CellDiv features with four previously published histomorphometric feature approaches(1,7–11) describing both cell morphology and cellular architecture. In total, we investigated the performance of 5 feature families: (1) 100 features describing nuclear shape (7), (2) 51 features describing global cell architectures(8), (3) 72 features describing cell orientation entropy by COrE (9) (24 features with three cell sub-graph setups), (4) 105 Cell Cluster Graph (CCG) features describing local cell cluster arrangement (1) (35 features with three cell sub-graph setups), and (5) 715 CellDiv features, characterizing the complexity of cell sub-graphs. Details of feature families are summarized in Table S4.

We also compared the CellDiv features with a deep learning method (DLM). The DLM was implemented using the DenseNet style Convolutional Neural Network (12). Specifically, a DL architecture comprising 1 input layer, 5 dense block layers, 3 and 1 output layer was constructed. The input layer accepts an image patch of 256 x 256 pixels, and the output layer is a soft-max function which outputs the class probability of being positive or negative. In the DLM, we split each TMA spot image into smaller patches of 256 x 256 pixels, the class labels for these image patches being assigned the same class label as that of the corresponding TMA spot image it was derived from. The average image size of the TMA spot was about 3800x3800 pixels at 20x magnification, which in turn resulted in a total of 23,049 patches. We performed the training and validation using a 3-fold cross-validation approach with 30 runs across each fold, all training and testing being done at the patient and not at the individual image-level. Once each of the individual image patches corresponding to a single patient have been assigned a class label, majority voting was employed to aggregate all the individual predictions to generate a patient-level prediction.

Table S5 shows the classification performance of different hand-crafted feature-based models and the DLM in the modeling cohort (n=486). The CellDiv based classifier achieved the highest AUC value of 0.68 ± 0.01 in differentiating short-term vs long-term survival. Xang et al.(11) and Yu et al.(13) previously presented nuclear histomorphometric based classifiers to predict tumor recurrence. Their approaches largely relied on nuclear shape, texture, and spatial arrangement features. In our comparative study, we found that these features only yielded a weak signal, AUC ≤ 0.55, in separating short-term and long-term survival early stage NSCLC patients. The combination of nuclear shape and spatial arrangement features yielded a mean AUC of 0.64.

While deep learning algorithms have become very popular for a variety of image processing and computer vision applications in digital pathology (14–17), these approaches are typically data-hungry and are dependent on large cohorts of training exemplars to learn the most discriminating representation. Most deep learning models are based on convolutional neural networks, networks in which a cascade of multiple layers comprise nonlinear processing units for feature extraction. Building a good deep learning model however requires a large amount of well-annotated training cases. In the modeling cohort, the deep learning approach

we employed was constrained by the fact that we had an unbalanced dataset. It is likely that the relatively few negative samples, coupled with the class imbalance resulted in a sub-optimally trained deep learning network, with a mean AUC of 0.62. In addition, the local nuclear architecture and cellular diversity information may not be captured by filters and convolutional operations which hinder the convolutional network to yield superior performance than hand-crafted biological inspired features.

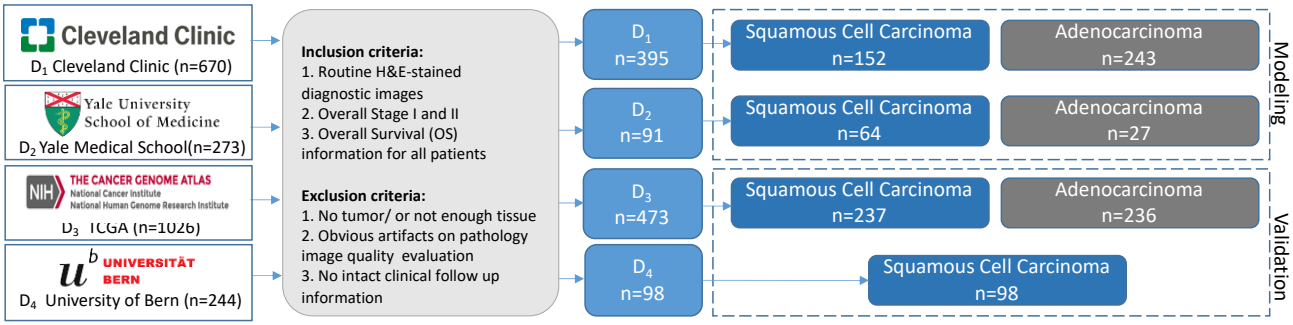

Figure S1: Data preparation and demographics of all cohorts

Patient enrolment, inclusion and exclusion criteria for the CellDiv Pathomic analysis in our study.

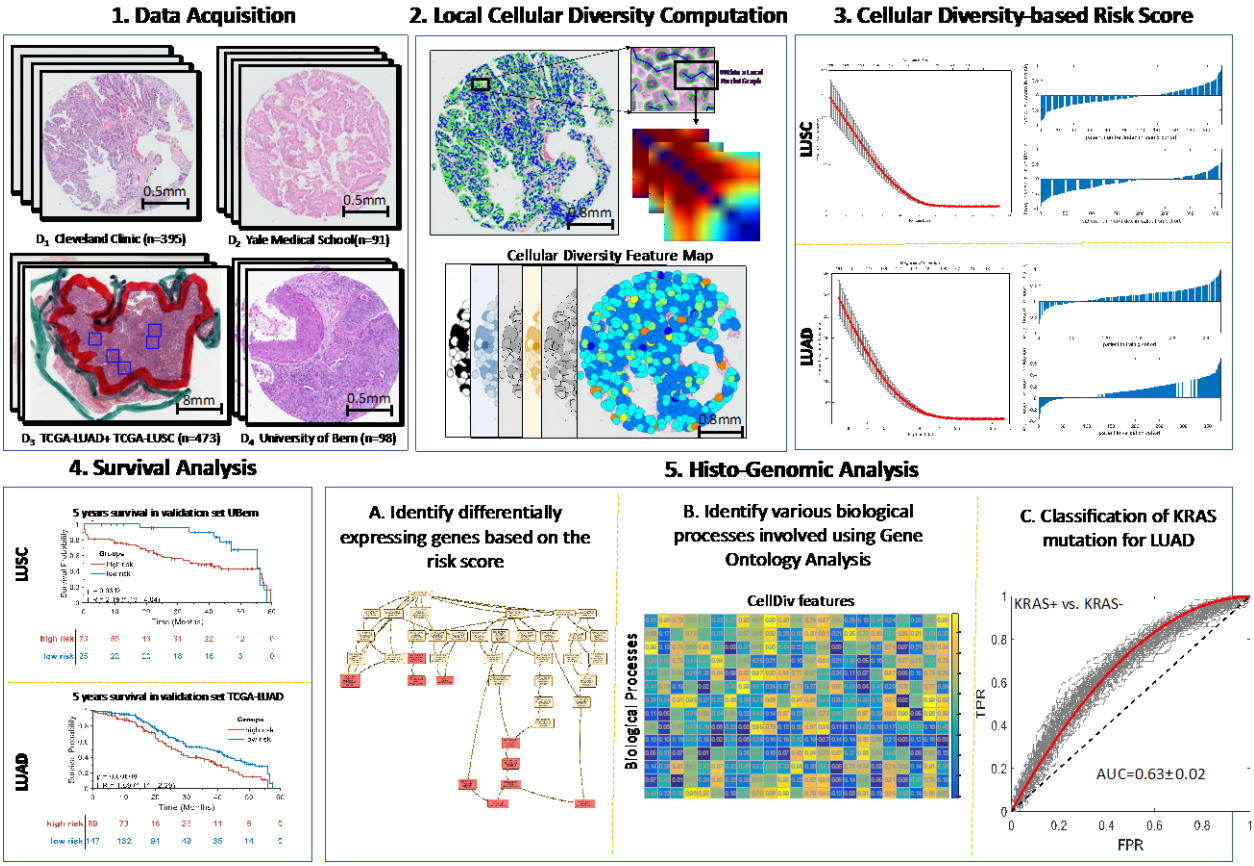

Figure S2: Experimental Design

The flowchart illustrates the complete experimental design. 1. Data Acquisition: Digitized TMAs and WSIs were collected from three different institutes and TCGA, in which cohorts D_1 and D_2 are the modeling cohorts, and D_3 and D_4 are the independent validation cohorts. 2. Local Cellular Diversity Computation: the nuclei were segmented by an automatic method and a local nuclear graph was constructed based on nuclear proximity. CellDiv features were then extracted from each local nuclear graph. 3. Cellular Diversity-based Risk Score: In the training phase, LASSO method was used to discover the top features for constructing risk score, for LUAD and LUSC specifically, using Cox proportional hazard model on modeling cohorts. 4. Survival Analysis: In the test phase, for the TCGA cohort, the same features were extracted from randomly selected five virtue TMA image blocks (2000x2000 pixels) from the whole slide diagnostic image of cohort D_3. The average feature value of the five virtue TMAs was used to represent each patient and feed into the pre-constructed Cox model. A risk score was generated for each patient in D_3 and D_4, and survival analysis was performed to evaluate the pre-trained Cox model. 5. Histogenomic

Analysis: A. Identification of differentially expressing genes. B. Identification of biological processes implicated using Gene Ontology and implementation of single-sample Gene Set Enrichment Analysis (ssGSEA) to understand the relationship between prognostic CellDiv features. C. Classification KRAS mutation vs. wildtype using CellDiv features for LUAD.

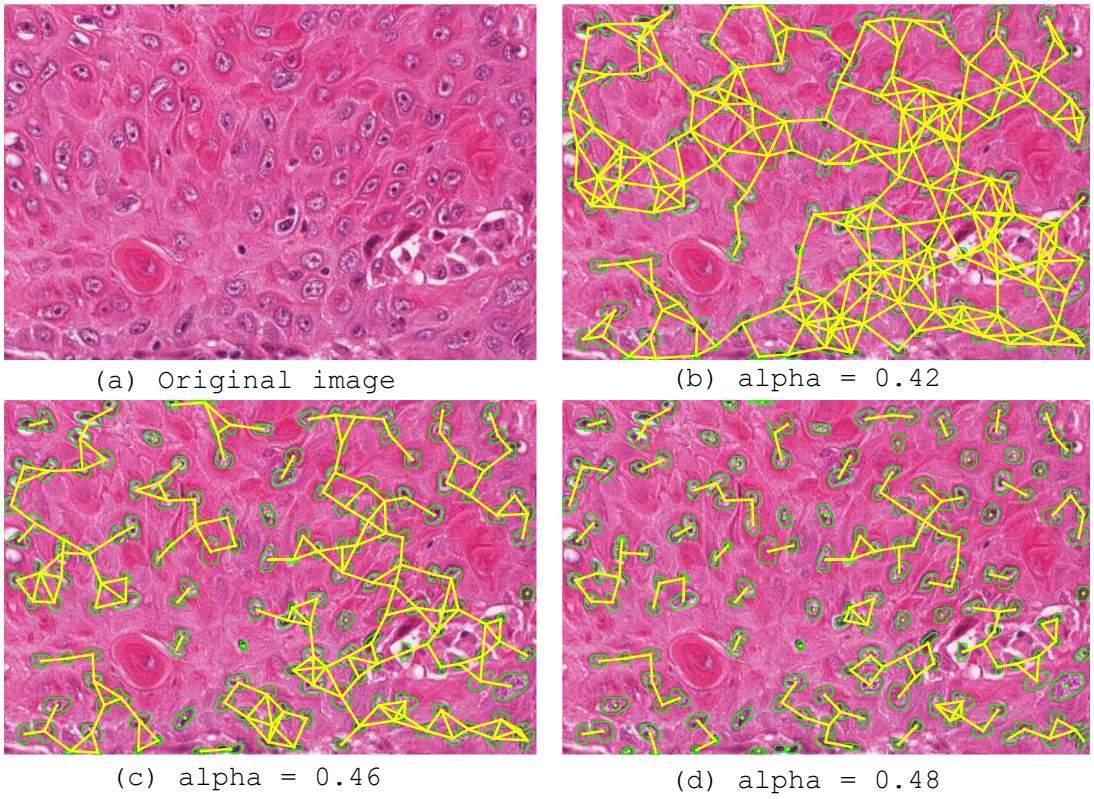

Figure S3: An example of an LNG configuration

(a) Original H&E image; We set r=0.2, and tuned α to (b) 0.42, (c) 0.46, and (d) 0.48 respectively to obtain the different LNG configurations.

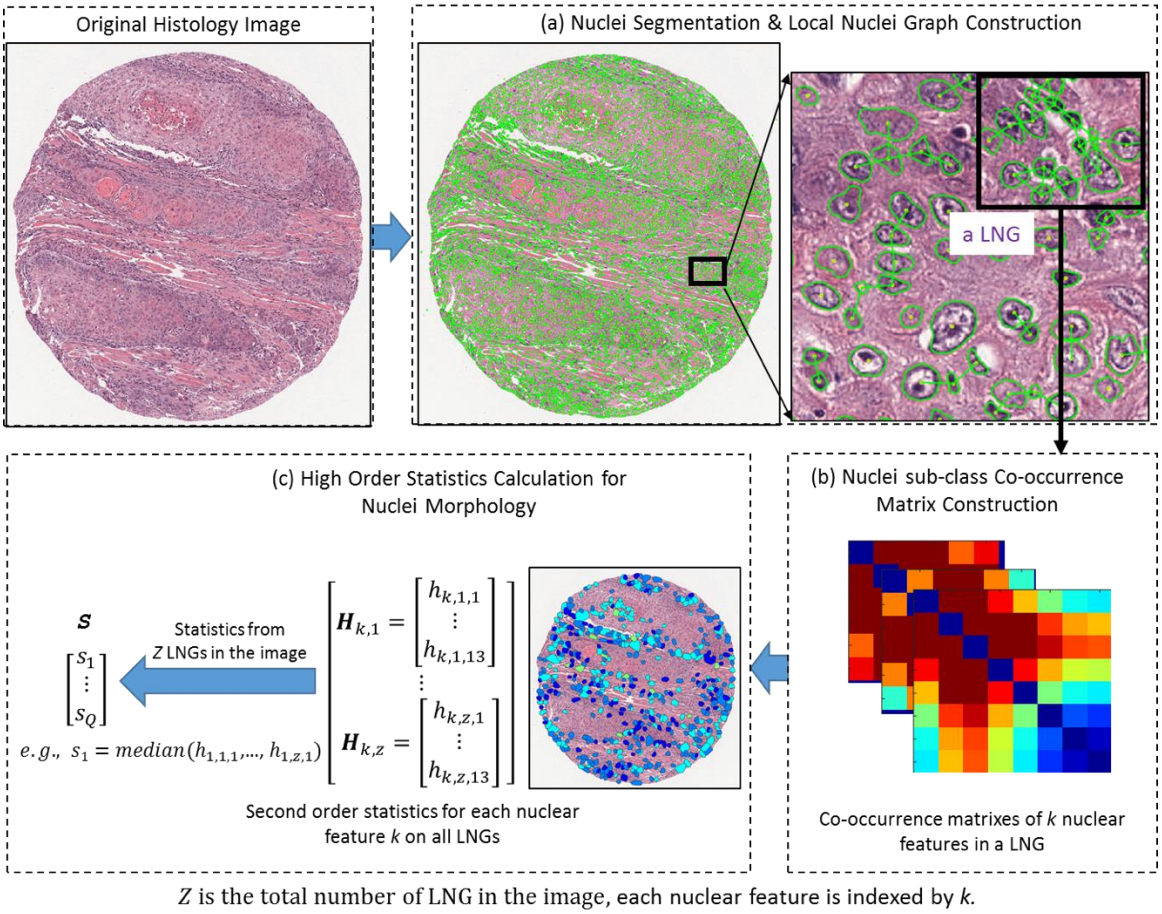

Figure S4: Flowchart for cellular diversity computation

(a) Green contours/lines indicate the nuclear boundaries and the adjacent panel shows the local nuclei graphs (LNG) with edges between proximally located nuclei. (b) Shape, size, and texture features are then extracted for each of the nodes in the LNG and co-occurrence matrices are constructed. From these co-occurrence matrices, high order statistics such as entropy are extracted and used to construct the cellular diversity feature vector ***S***.

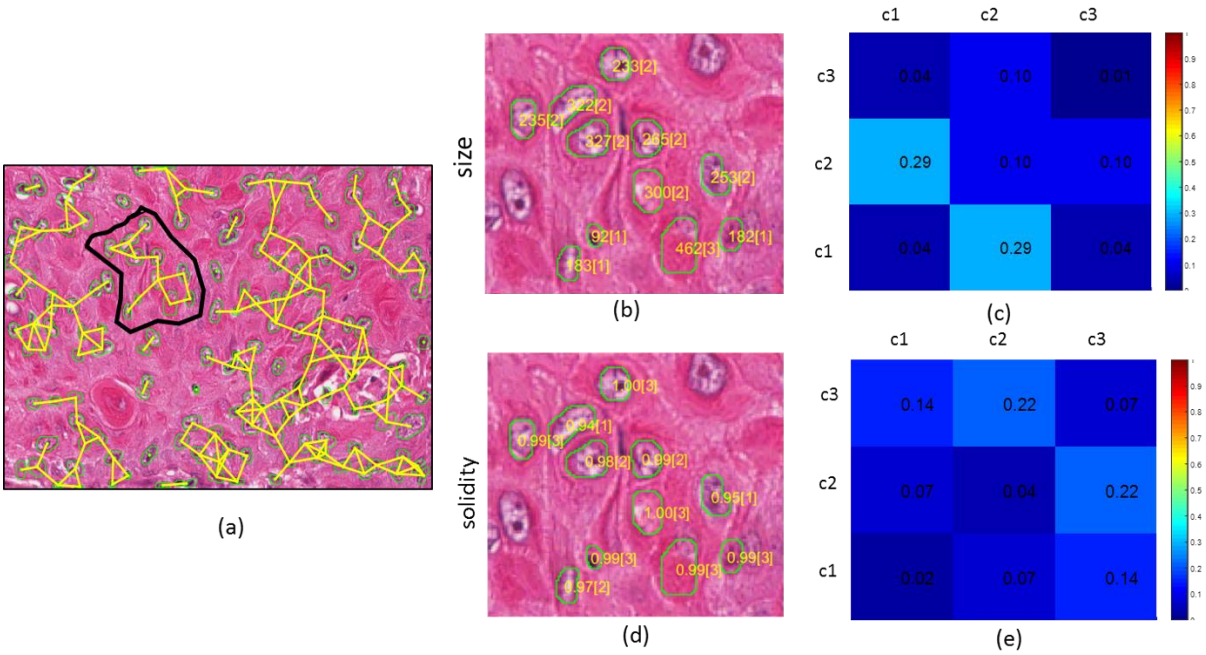

Figure S5: An example of cellular diversity computation

(a) Original H&E image; (b) magnified region of the LNG, α=0.44, marked by black contour shown in (a) with nuclear size and its category (shown in the parenthesis) displayed. (c) corresponding co-occurrence matrix of nuclear size; (d) magnified region of the LNG marked by black contour shown in (a) with nuclear solidity and its category (shown in the parenthesis) displayed (e) corresponding co-occurrence matrix of nuclear solidity.

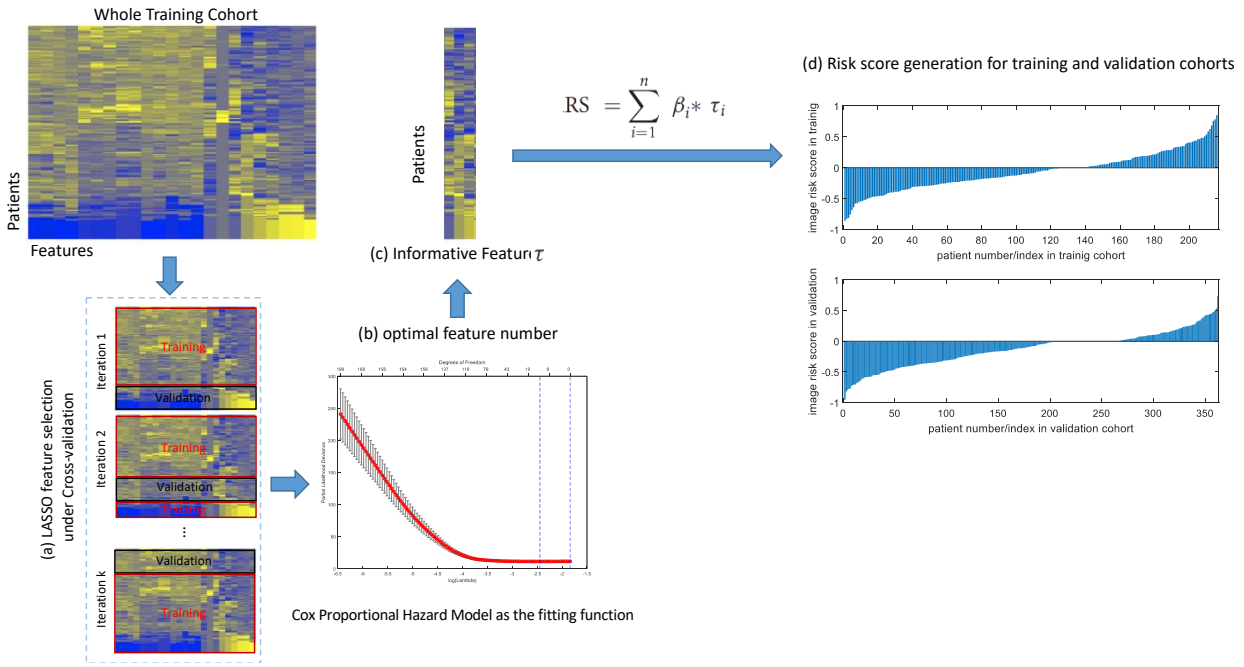

Figure S6: Machine-learning methodology flowchart

After extracting the CellDiv features from all images in the training cohort, we performed (a) feature selection using Least Absolute Shrinkage and Selection Operator (LASSO) with Cox Proportional Hazard Model as the cost function under a 10-fold cross-validation scheme with 100 runs. LASSO L1 regularization technique iteratively shrinks the feature coefficient estimates toward zero, and results in identification of an optimal tuning parameter lambda (18) that increases in a cross-validation setup until features with only nonzero coefficients are retained. Therefore, in cases with a very large number of features, a LASSO model can help both shrink and find the sparse model that involves a small subset of the most informative features (shown in (c)). Thus, features picked by the LASSO models within the training cohort are then pooled in a linear combination and multiplied with their respective coefficients to construct a risk score (RS). For all the images in the training and independent test/validation cohorts, a risk score was then generated for each patient (shown in (d)).

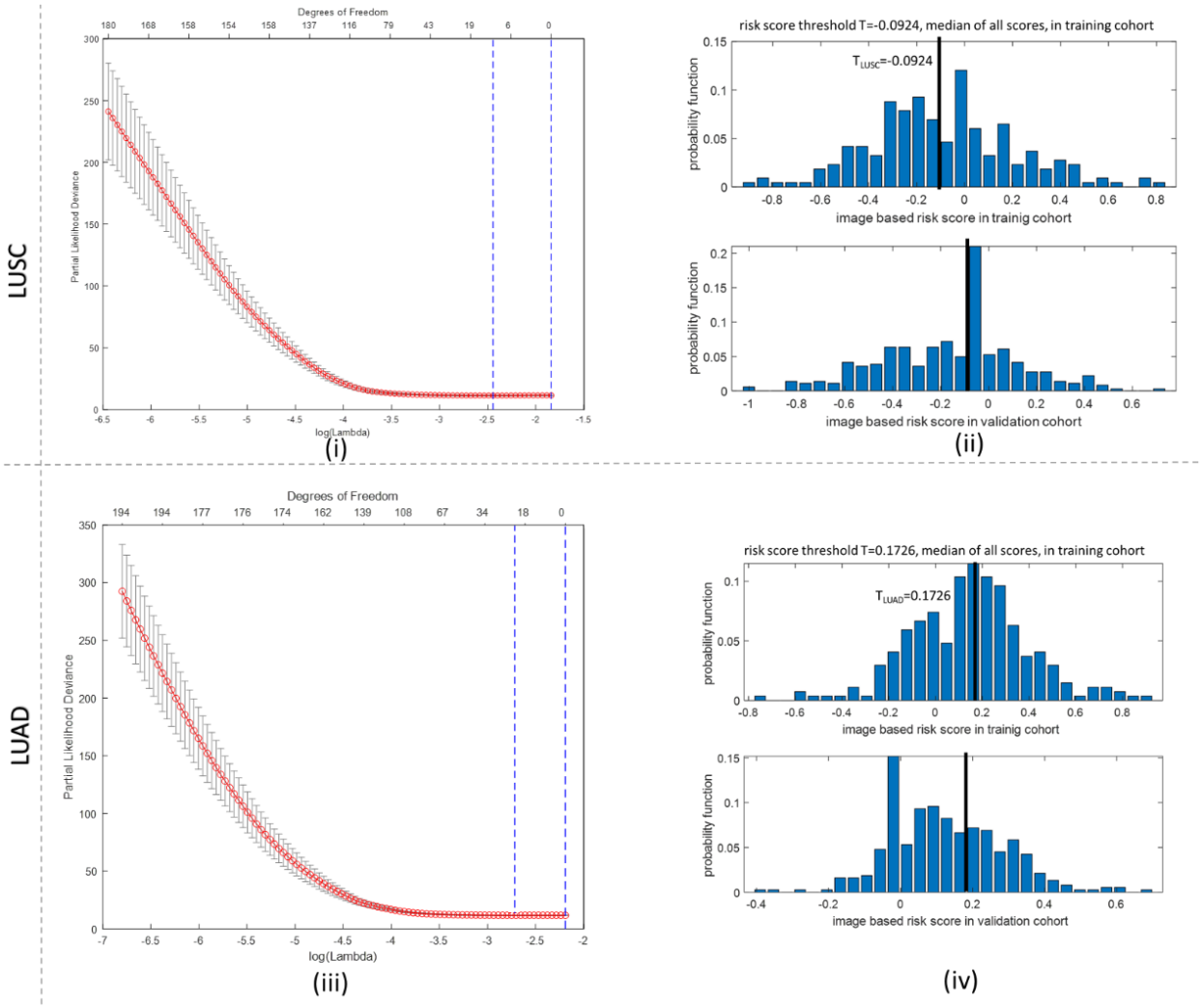

Figure S7: Feature selection using LASSO for (i) LUSC and (iii) LUAD

A set of 11 CellDiv features and a set of 23 CellDiv features were selected to construct the Cox Proportional Hazard Regression model for LUSC and LUAD, respectively. The risk score distribution of LUSC and LUAD are shown in (ii) and (iv) for LUSC and LUAD, respectively. The median value was selected as the threshold from the training cohort to separate the high vs. low risk patients. This threshold was then applied to the validation cohort for the high vs. low risk patients.

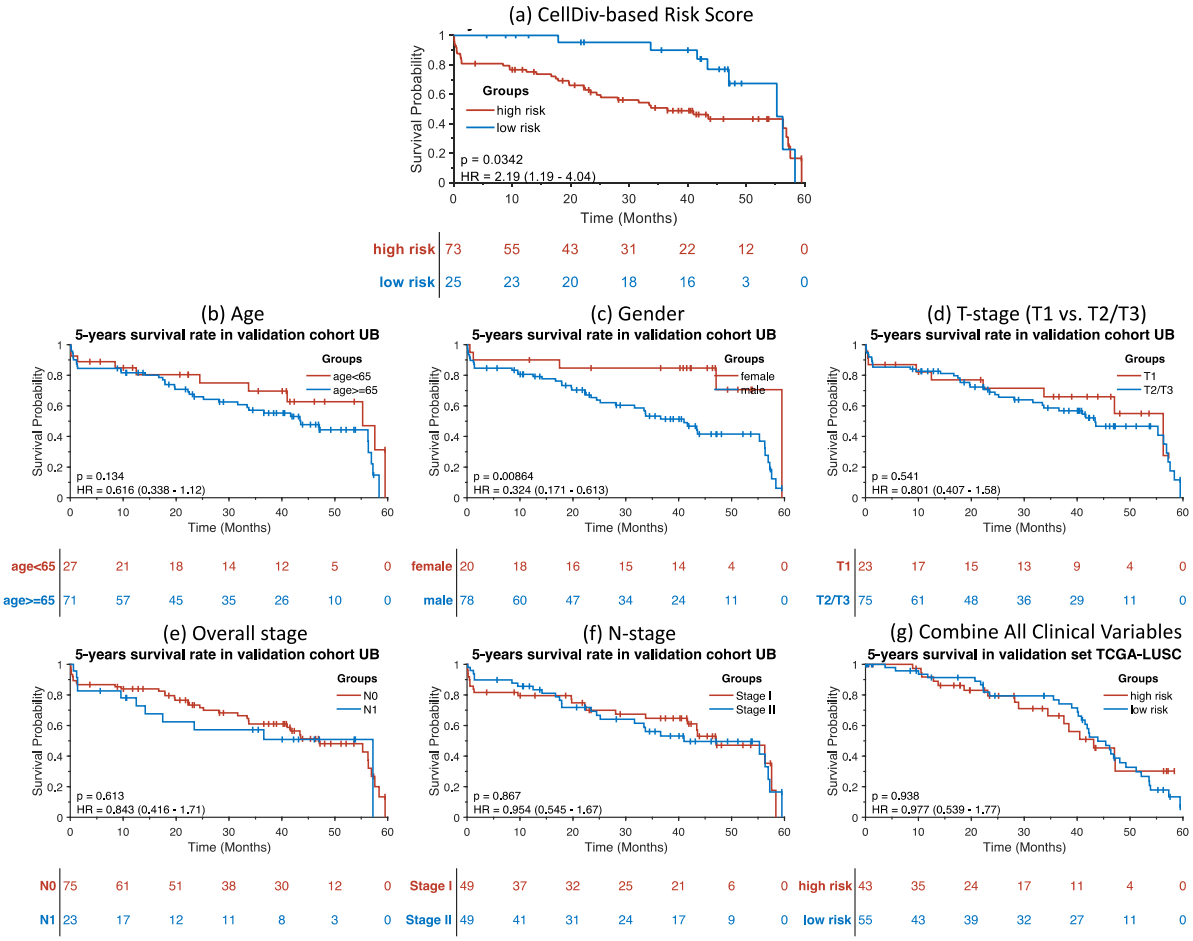

Figure S8: The KM curves for CellDiv-based risk score (a) and all available clinical variables (b)-(g) in validation/test cohort LUSC-UBern

One may observe that in this cohort, apart from the gender, no clinical variable was able to sperate the patients in terms of 5-year survival.

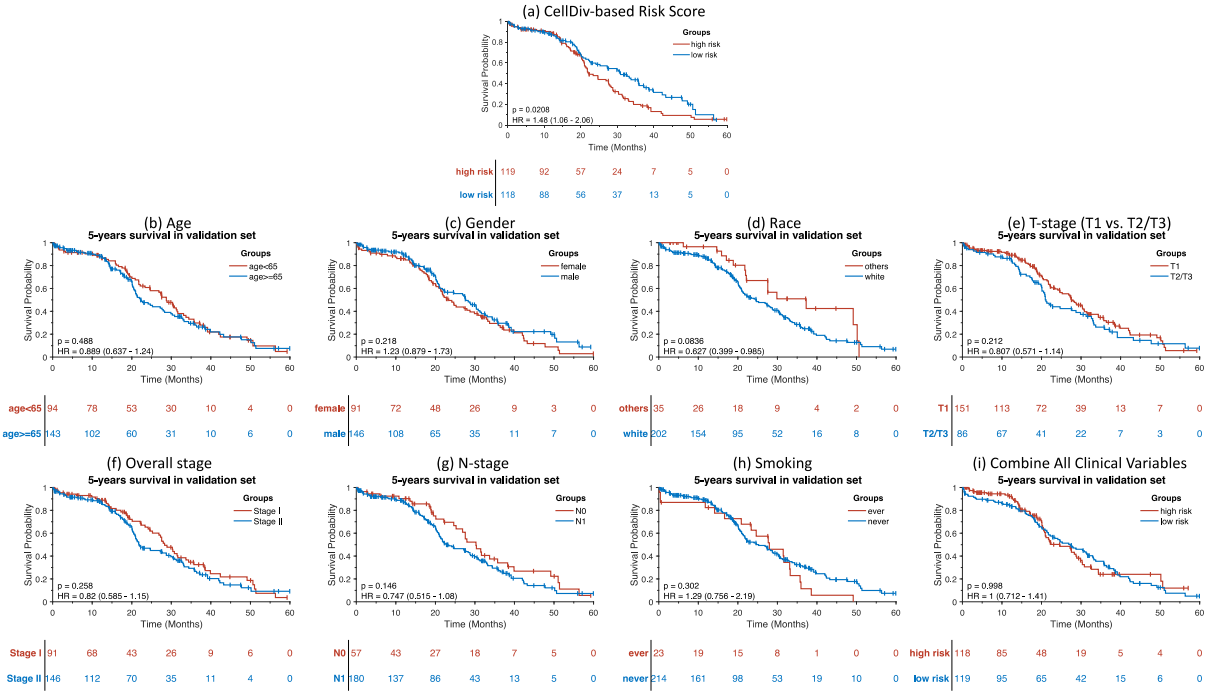

Figure S9: The KM curves for CellDiv-based risk score (a) and all available clinical variables (b)-(i) in validation/test cohort LUSC-TCGA

One may observe that in this cohort, no clinical variable was able to sperate the patients in terms of 5-year survival.

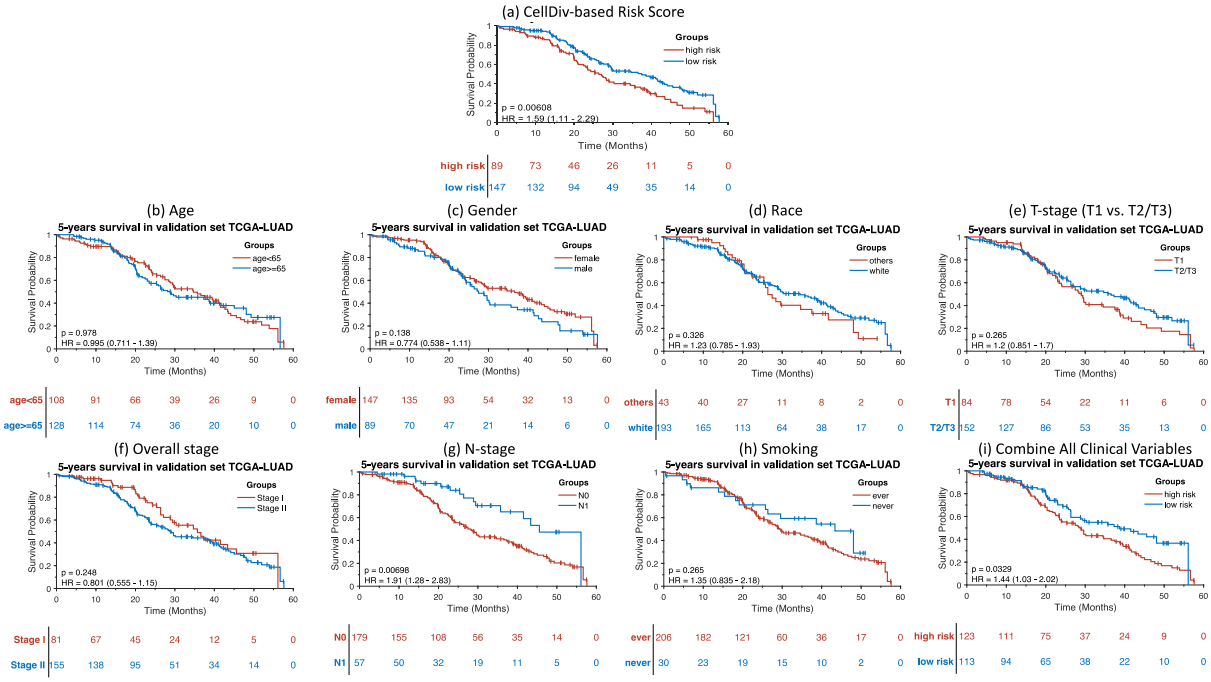

Figure S10: The KM curves for CellDiv-based risk score (a) and all available clinical variables (b)-(i) in validation/test cohort LUAD-TCGA

One may observe that in this cohort, N-stage was prognostic of survival; also, the model that combines all clinical variables was prognostic in terms of 5-years survival. While analyzing differences between the N-stage and CellDiv-based models, we found that these models made errors on different patients, which implies that the combination of the clinical and CellDiv-based model could potentially outperform any individual model (result shown in Table S3-last row, and Table 1 in the main manuscript corroborate this hypothesis).

Table S1: Nuclear features considered in image analysis

Feature Class	No.	Specific Attributes
Nuclear shape	6	Area, Eccentricity, Solidity, Circularity, Major/minor axis length of best fit eclipse
Nuclear appearance	5	Mean intensity, Intensity range, Mean inside/outside boundary intensity, Boundary Saliency

Table S2: 13 Haralick measurements of the co-occurrence matrix (CM)

The "intensity" refers to the quantification levels of nuclear morphology, for example, a 3-levels was used in Fig.2 to quantify nuclear size and solidity. $P_{i,j}$represents the element in CM, where *i* and *j* represents the indices of the quantification level. $P_x \; and \; P_y$ are the partial probability density functions. P_{x+y} is the probability of CM coordinates summing to *x* + *y*, respectively. N_g is the quantification level. $\mu_x, \mu_y, \sigma_x, and \; \sigma_y$are the means and standard deviations of $P_x \; and \; P_x$, respectively.

Descriptor	Intuitive Description	Equation
Entropy	Measure of randomness of CM values High entropy: large variations in CM values extracted from an image Low entropy: increasingly homogenous CM values.	$\sum_{i,j} P_{i,j} \log P_{i,j}$
Energy	Measure of homogeneity of CM values High energy: increasingly uniform distribution of CM values from a ROI Low energy: increasingly heterogeneous CM values.	$\sum_{i,j} {P^2}_{i,j}$
Contrast	Measure of variations present in local regions of contrast. High value: higher spatial frequencies corresponding to large differences in a contiguous set of intensities. Low value: lower spatial frequencies.	$\sum_i \sum_j (i-j)^2 \; \left(P_{i,j}\right)$
Information Measure of Correlation 1	Measure of linear dependency with respect to directional entropy. Roughly inversely varies with correlation, but provides a more "natural" measure of correlation due to its invariance under logarithmic transformation. High IMC1: Greater presence of heterogeneous CM values in linear directions. Low IMC1: Uniformity of CM values in linear directions.	$\frac{entropy + \sum_i \sum_j P(i,j) \log\{P_x(i) P_y(j)\}}{\max(entropy_x, entropy_y)}$

Information Measure of Correlation 2	Measure of linear dependency with respect to randomness of spatial dependency. Roughly directly varies with correlation, but provides a more "natural" measure of correlation due to its invariance under logarithmic transformation. High IMC1: Higher presence of repeating patterns in linear directions. Low IMC2: Lack of uniformity or repeating patterns in linear directions.	$\sqrt{1-\exp\left[-2.0\left[-\sum_i\sum_j P_x(i)P_y(j)\log\{P_x(i)P_y(j)\}\right]\right]}$
Intensity Average	Measure of CM distribution relationship to mean intensity accumulations. High sum average: indicative of higher presence of punctate regions of high intensity. Low sum average: lack of presence of such punctate regions.	$\sum_{i=2}^{2N_g} iP_{x+y}(i)$
Intensity Entropy	Measure of CM relationship to distribution of intensity with respect to entropy. High and low values correspond similarly to entropy values.	$-\sum_{i=2}^{2N_g} P_{x+y}(i)\log\{P_{x+y}(i)\}$
Contrast Entropy	Measure of CM relationship to intensity differences with respect to entropy. Inversely varies with intensity entropy.	$\sum_{i=0}^{N_{g-1}} P_{x-y}(i)\log\{P_{x-y}(i)\}$
Contrast Average	Measure of CM relationship to mean intensity differences. Inversely varies with intensity average.	$\sum_{i=0}^{N_{g-1}} iP_{x-y}(i)$
Correlation	Measure of intensity linear dependency. High Correlation: scale of local pattern is larger than the distance. Low Correlation: scale of local pattern is smaller than the distance.	$\sum_{i,j}\frac{(ij)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$
Intensity Variance	Measure of CM relationship to distribution of intensity with respect to variance. High sum variance: greater standard deviation of sum average.	$\sum_{i=2}^{2N_g}(i-\text{Intensity Entropy})^2 P_{x+y}(i)$

	Low sum variance: low standard deviation of sum average.	
Contrast Inverse Moment	Measure of local regions of homogeneity High value: Higher presence of locally uniform windows in CM. Low value: Higher presence of locally heterogeneous windows in CM.	$\sum_{i,j} \frac{P_{i,j}}{1+(i-j)^2}$
Contrast Variance	Measure of CM relationship to intensity differences with respect to variance Inversely varies with sum variance.	$\sum_{i=0}^{N_{g-1}} (i - \text{Sum Entropy})^2 \, P_{x-y}(i)$

Table S3: Features used for constructing Cox risk model of LUSC

Feature names	Weights
Features related to nuclear shape	
MajorAxisLength:range(intensity-ave)	-0.033375317
MajorAxisLength:skewness(correlation)	-0.049320017
MinorAxisLength:std(contrast-energy)	-0.226360529
Features related to nuclear intensity	
MeanIntensity:range(intensity-var)	0.45547197
IntensityDeviation:std(contrast-energy)	0.091109774
IntensityDeviation:skewness(intensity-ent)	0.061559712
IntensityRange:skewness(contrast-ent)	0.034147727
IntensityRange:skewness(intensity-ent)	0.013221452
MeanInsideBoundaryIntensity:median(energy)	-2.315750963
MeanInsideBoundaryIntensity:skewness(contrast-inverse-moment)	0.248258524
InsideBoundaryIntensityDeviation:skewness(contrast-var)	0.004300509

Table S4: Features used for constructing Cox risk model of LUAD

Feature names	Weights
Features related to nuclear shape	
MajorAxisLength:skewness(contrast-energy)	0.199763
MinorAxisLength:kurtosis(correlation)	0.007342
MinorAxisLength:skewness(info-measure2)	-0.04511
Eccentricity:std(contrast-inverse-moment)	-3.66735
Solidity:mean(info-measure1)	0.451829
Solidity:std(contrast-ent)	-0.10196
Solidity:kurtosis(contrast-ent)	0.034818
Circularity:range(entropy)	0.231151
Circularity:skewness(intensity-var)	-0.10397
Circularity:skewness(info-measure1)	-0.0503
Features related to nuclear intensity	
MeanIntensity:skewness(contrast-var)	-0.00918
MeanIntensity:skewness(intensity-var)	-0.04159
MeanIntensity:skewness(intensity-ent)	-0.08649
MeanIntensity:skewness(energy)	0.152872
IntensityRange:kurtosis(contrast-inverse-moment)	0.07245
IntensityRange:kurtosis(intensity-var)	-0.00141
MeanInsideBoundaryIntensity:std(contrast-inverse-moment)	-1.08264
MeanInsideBoundaryIntensity:std(info-measure2)	-0.18439
InsideBoundaryIntensityRange:kurtosis(intensity-ave)	-0.00178
OutsideBoundaryIntensityDeviation:mean(correlation)	0.366844
OutsideBoundaryIntensityDeviation:mean(info-measure1)	0.330472
OutsideBoundaryIntensityDeviation:std(correlation)	-0.41944
BoundarySaliency:kurtosis(intensity-ave)	-0.00359

Table S5: Top features used for *KRAS* classification

Feature name
MeanInsideBoundaryIntensity:mean(intensity-ave)
MeanInsideBoundaryIntensity:median(intensity-ave)
MeanInsideBoundaryIntensity:median(intensity-ent)
MeanIntensity:mean(intensity-ave)
MeanInsideBoundaryIntensity:mean(intensity-var)
MeanOutsideBoundaryIntensity:range(info-measure2)

Table S6: The precision-recall(PR)-AUC, F1-score, precision and recall of CellDiv model, Clinical variables model, and CellDiv + Clinical variables model for test cohorts: LUSC-TCGA, LUSC-UB, and LUAD-TCGA

In this scenario, we used a threshold of 3 years to categorize the patients into short-term survival (<=3 years) and long-tern survival (>3 years).

Cohort	Model	PR-AUC (random predictor)	F1-score	Recall	Precision
LUSC-TCGA (203+/34-)	CellDiv	0.89 (0.58)	0.92	1.00	0.86
	Clinical variables	0.84 (0.58)	0.92	1.00	0.86
	CellDiv + Clinical variables	0.85 (0.58)	0.92	1.00	0.86
LUSC-UB (55+/43-)	CellDiv	0.63 (0.47)	0.71	0.73	0.70
	Clinical variables	0.52(0.47)	0.69	0.87	0.57
	CellDiv + Clinical variables	0.64 (0.47)	0.71	0.73	0.70
LUAD-TCGA (177+/59-)	CellDiv	0.72 (0.70)	0.86	0.98	0.76
	Clinical variables	0.77 (0.70)	0.86	0.99	0.76
	CellDiv + Clinical variables	0.80 (0.70)	0.86	1.00	0.75

Table S7: Description of global cell graph, nuclear shape, COrE, and CCG features in quantitative evaluation experiments.

Feature set	#	Description
Global Cell Graph (8)	51	Voronoi Diagram: Polygon area, perimeter, chord length; Delaunay Triangulation: Triangle side length, area; Minimum Spanning Tree: Edge length (compute mean, std. dev., range, skewness, kurtosis, disorder of each); Nearest Neighbors: Density of nuclei, distance to nearest nuclei
Nuclear Shape (7)	100	Area, Mean Intensity/Intensity Range of Nuclei, Mean Intensity/Intensity Range Around Nuclei, Eccentricity, Perimeter, Smoothness, Invariant Moment 1-7, Fractal Dimension, Fourier Descriptor 1-10 (Mean, Std. Dev, Median, range, skewness, kurtosis of each)
cell orientation entropy (CorE) (9)	72	4 haralick measurements computed from nuclear orientation co-occurrence matrix (Mean, Std. Dev, median, range, skewness, kurtosis of each with three cell sub-graph setups)
Cell Cluster Graph (CCG) (1)	105	Clustering Coeff C, Clustering Coeff D, Giant Connected Component, Average Eccentricity, Percent of Isolated Points, Number of Central Points, Skewness of Edge Lengths (Mean, Std. Dev, skewness, kurtosis, range of each with three cell sub-graph setups)

Table S8: Performance comparison of different hand-crafted feature-based models, and the DL method

Features/Methods	AUC	Accuracy	Specificity	Sensitivity
Global Cell Graph (8)	0.55±0.01	0.55±0.05	0.58±0.12	0.53±0.13
Nuclear Shape (7)	0.53±0.02	0.59±0.06	0.39±0.18	0.70±0.18
COrE (9)	0.52±0.02	0.51±0.08	0.62±0.21	0.46±0.21
CCG (1)	0.54±0.01	0.52±0.04	0.63±0.12	0.47±0.11
Wang et al. (11)	0.64±0.02	0.61±0.04	0.63±0.11	0.57±0.09
DL (12)	0.62±0.04	0.60±0.07	0.62±0.18	0.59±0.17
CellDiv	0.68±0.01	0.63±0.04	0.72±0.10	0.58±0.09

References

1. Ali S, Veltri R, Epstein JA, Christudass C, Madabhushi A. Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays. In: Gurcan MN, Madabhushi A, editors. 2013 [cited 2016 Mar 18]. p. 86760H. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2008695

2. Lewis JS, Ali S, Luo J, Thorstad WL, Madabhushi A. A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. Am J Surg Pathol. 2014 Jan;38(1):128–37.

3. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics. 1973 Nov;3(6):610–21.

4. De Santo M, Molinara M, Tortorella F, Vento M. Automatic classification of clustered microcalcifications by a multiple expert system. Pattern Recognition. 2003 Jul 1;36(7):1467–77.

5. Lam FC, Longnecker MT. A modified Wilcoxon rank sum test for paired data. Biometrika. 1983;70(2):510–3.

6. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLOS ONE. 2019 Nov 7;14(11):e0224365.

7. Kothari S, Phan JH, Young AN, Wang MD. Histological image classification using biologically interpretable shape-based features. BMC Medical Imaging [Internet]. 2013 Dec [cited 2018 Jan 11];13(1). Available from: http://bmcmedimaging.biomedcentral.com/articles/10.1186/1471-2342-13-9

8. Shin D, Protano M-A, Polydorides AD, Dawsey SM, Pierce MC, Kim MK, et al. Quantitative Analysis of High-Resolution Microendoscopic Images for Diagnosis of Esophageal Squamous Cell Carcinoma. Clinical Gastroenterology and Hepatology. 2015 Feb;13(2):272-279.e2.

9. Lee G, Ali S, Veltri R, Epstein JI, Christudass C, Madabhushi A. Cell orientation entropy (COrE): predicting biochemical recurrence from prostate cancer tissue microarrays. Med Image Comput Comput Assist Interv. 2013;16(Pt 3):396–403.

10. Lee G, Veltri RW, Zhu G, Ali S, Epstein JI, Madabhushi A. Nuclear Shape and Architecture in Benign Fields Predict Biochemical Recurrence in Prostate Cancer Patients Following Radical Prostatectomy: Preliminary Findings. Eur Urol Focus. 2016 Jun 16;

11. Wang X, Janowczyk A, Zhou Y, Thawani R, Fu P, Schalper K, et al. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. Scientific Reports [Internet]. 2017 Dec [cited 2017 Dec 1];7(1). Available from: http://www.nature.com/articles/s41598-017-13773-7

12. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. arXiv:160806993 [cs] [Internet]. 2018 Jan 28 [cited 2020 Jul 15]; Available from: http://arxiv.org/abs/1608.06993

13. Yu K-H, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun. 2016 16;7:12474.

14. Wang H, Cruz-Roa A, Basavanhally A, Gilmore H, Shih N, Feldman M, et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. J Med Imaging (Bellingham). 2014 Oct;1(3):034003.

15. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Scientific Reports [Internet]. 2016 Sep [cited 2018 Feb 16];6(1). Available from: http://www.nature.com/articles/srep26286

16. Irshad H, Veillard A, Roux L, Racoceanu D. Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential. IEEE Reviews in Biomedical Engineering. 2014;7:97–114.

17. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Sci Rep. 2017 Apr 18;7:46450.

18. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Soft [Internet]. 2011 [cited 2020 Jul 4];39(5). Available from: http://www.jstatsoft.org/v39/i05/