**Supplementary Methods**

Prognostic value of signatures (Collagen component, Immune component, Collagen+Immune components) in the chemotherapy, radiotherapy, and immunotherapy-treated validation cohorts

The signature (Collagen features alone) was not associated with OS in chemotherapy and radiotherapy-treated patients and PFS in chemotherapy-treated patients (chemotherapy CSCC (D1, p=0.63, HR=1.2, 95% CI=0.57-2.52, c-index=0.59), radiotherapy CSCC (D2, p=0.1, HR=1.74, 95% CI=0.91-3.35, c-index=0.59), chemotherapy EC (D3, p=0.38, HR=0.676, 95% CI=0.63-1.71, c-index=0.41), chemotherapy HGSOC (D8, p=0.54, HR=1.59, 95% CI=0.422-5.96, c-index=0.71)). In the cohort treated with no therapy (D4), the signature was associated with OS (p=0.04, HR=3.1, 95% CI=1.02-15.6, c-index=0.82). In the immunotherapy-treated cohorts (D5, D6 and D7), the signature was not associated with PFS (p=0.58, HR=0.823, 95% CI=0.42-1.62, c-index=0.51).

The signature (Immune features alone) was associated with OS in chemotherapy and radiotherapy-treated patients but not with PFS in chemotherapy-treated patients (chemotherapy CSCC (D1, p=0.001, HR=4.3, 95% CI=2.05-9.03, c-index=0.7), radiotherapy CSCC (D2, p=0.007, HR=2.96, 95% CI=1.52-5.73, c-index=0.69), chemotherapy EC (D3, p=0.037, HR=3.38, 95% CI=1.34-8.51, c-index=0.66), chemotherapy HGSOC (D8, p=0.08, HR=3.07, 95% CI=0.9-9.85, c-index=0.57)). In the cohort treated with no therapy (D4), the signature was not associated with OS (p=0.08, HR=3.11, 95% CI=0.727-13.3, c-index=0.75). In the immunotherapy-treated cohorts (D5, D6 and D7), the signature was also not associated with PFS (p=0.122, HR=1.79, 95% CI=0.9-3.54, c-index=0.58).

The CollaTIL signature (both Collagen and Immune features) was associated with OS or PFS in various validation cohorts. Specifically, in the chemotherapy-treated validation cohorts (D1, D3 and D8), it was associated with OS in D1 (p=0.0256, HR=2.54, 95% CI=1.21-5.34, c-index=0.72), D3 (p=0.0213, HR=7.36, 95% CI=2.73-19.8, c-index=0.7), and PFS in D8 (p=0.0434, HR=3.07, 95% CI=1.01-9.85, c-index=0.74). In the radiotherapy-treated validation cohort (D2), it was also associated with OS (p=0.006, HR=2.87, 95% CI=1.49-5.53, c-index=0.72). Moreover, the signature identified high risk patients who had significantly worse PFS than low-risk patients in the immunotherapy-treated validation cohorts (D5, D6, and D7, p=0.0184, HR=2.72, 95% CI=1.36-5.42, c-index=0.65) (**Fig. 4**).

Univariate analysis, multivariable analysis, and integrated nomogram

After performing univariate and multivariable analysis of clinical and pathological variables on validation cohorts (D1-D8) for predicting OS/PFS, we determined the clinical variables that should be incorporated into the construction of a clinical histomorphometric nomogram along with CollaTIL (**Supplementary Table 3**). Results of the multivariable analysis indicated that CollaTIL risk score (obtained during univariate analysis of CollaTIL) and stage were the two variables that should be used for training the Cox regression model on the D0 cohort. We tested the nomogram in the validation cohorts (D1-D8), and found that it was associated with OS in all chemotherapy and radiotherapy-treated cohorts and with PFS in the chemotherapy-treated cohort (D1, p=0.04, HR=2.2, 95% CI=1.05-4.61; D2, p=0.003, HR=3.11, 95% CI=1.62-5.99; D3, p=0.04, HR=5.76, 95% CI=2.02-16.5; D8, p=0.043, HR=3.07, 95% CI=1.01-9.85). Moreover, in the immunotherapy-treated validation cohorts, it was associated with PFS (D5, D6, and D7, p=0.018, HR=2.72, 95% CI=1.36-5.42).

**More detail on the patients and specimens**

For genomic analysis, we employed a log2-transformed fragments per kilobase of mRNA per million reads (FPKM) of the estimated gene expression profiles (level 3, from “IlluminaHiSeq\_RNASeqV2” platform) of tumor samples from the TCGA normalized read counts, which aligned to the GRCh38 (V.28) reference genome and their corresponding H&E WSI and clinical data were available.

**Survival analysis**

OS was defined as the time from diagnosis to death or last follow-up for survivors, while PFS was defined as the time from diagnosis to disease progression or death, whichever came first, and was censored for patients without disease progression at the last follow-up. A Cox proportional hazard model1 (referred to as Cox regression model) with elastic net penalty2 was trained using CollaTIL features on the D0 cohort to predict OS. All features were standardized using MinMax scaling method. A coefficient was assigned to each of the features in the final model and a continuous risk score was obtained by linear combination of top features weighted by corresponding coefficients for each patient. The continuous risk score for each patient reflects an estimated risk for OS or PFS. This risk score was converted to a binary high vs low value using the mean threshold. The CollaTIL signature was subsequently validated on D1-D4 for predicting OS and D5-D8 for predicting PFS, using the same set of feature coefficients. Kaplan-Meier survival analysis with log-rank test was used to examine the differences of time-to event data between the two defined patient groups. The performance of models was summarized by hazard ratios (HRs) along with their 95% confidence intervals (CIs) using Wald test and Harrell’s concordance index (c-index) on D1-D8 validation cohorts. P-values less than 0.05 were considered statistically significant.

**Testing the performance of different survival models**

For this task, survival models such as Cox proportional regression model1 (without elastic net penalty2), Cox proportional regression model1 (with elastic net penalty2), Random survival forest and Survival support vector machine were validated on cohorts (D1-D8) by training on D0 cohort. The c-indexes were used for comparing the performance of these survival models on the validation cohorts (D1-D8) (see **Supplementary Table 4**).

We can conclude that Random survival forest gives similar results (in terms of c-index) as compared to Cox proportional regression model (with elastic net penalty) used for our study. This is another survival model that can be used. However, it is important to note that the choice of the hyperparameter, number of estimators, used for training the random survival forest model has to be decided according to the number of patients in the training cohort to avoid overfitting. For our case, we decided to use the value of 10 as the training cohort (D0) consisted of 95 patients.

**Disease type sensitivity and training size sensitivity analysis of the model performance**

To ensure that our results are consistent irrespective of the number of samples in the training set or histology or type of treatment given to patients after surgery, we performed three experiments listed below. Experiment 1 is the primary model described in the study where the training dataset consisted of HGSOC patients treated with chemotherapy. In this section, we performed additional experiments (experiment 2 and experiment 3) and compared the c-index values across all the validation cohorts. In experiment 2, we replaced the training cohort used for our study with the D2 cohort (CSCC, training size=128) and validated it on cohorts (D0, D1, D3, D4, (D5 and D6 and D7), D8). While for experiment 3, we replaced the training cohort used for our study with the D1 cohort (CSCC, training size=134) and validated it on cohorts (D0, D2, D3, D4, (D5 and D6 and D7), D8). We can infer that our signature was not sensitive to the choice of training cohort used for the study. It shows similar c-index values irrespective of the training cohort used (D0 or D1 or D2) (see **Supplementary Table 5)**.

**Steps involved in extracting features from collagen component**

The steps involved in extracting features from collagen component are as follows:

* Detecting collagen fibers
* Find the orientation of the detected collagen fibers
* Discretize the orientations in 18 bins (0 to 17)
* Then tumor neighborhoods of nine different sizes (200x200-pixel, 250x250-pixel, 300x300-pixel, 350x350-pixel, 400x400-pixel, 450x450-pixel, 500x500-pixel, 550x550-pixel, 600x600-pixel) move across the extracted tiles of whole slide images in a sliding window way, calculate the disorder of collagen fiber orientations by constructing an orientation co-occurrence matrix within each tumor neighborhood
* Quantitative measurement of the disorder in collagen fiber orientations was measured using the orientation co-occurrence matrix using entropy theory

**Step 1: Detecting collagen fibers using derivative of Gaussian model**

* We used the derivative of Gaussian based model in Matlab to detect the linear structures in the image, which are the prevalent morphology phenotype exhibited by collagen fibers on H&E images.

**Step 2: Orientation of detected collagen fibers (regionprops function in Matlab)**

* We use the collagen fibers that are detected by the derivative-of-Gaussian model. Essentially for each pixel we compute the x-gradient, y-gradient and that gives us the orientation (Orientation (θ) = atan2(y-gradient, x-gradient))
* For each collagen fiber, we create the histogram of orientations for each pixel as computed above and find the orientation that occurs more frequently that is considered as the orientation for that collagen fiber.

**Step 3: Discretize the orientations in 18 bins from 0 to 17**

**Step 4: Orientation co-occurrence matrix**

Divide each tile extracted from the whole slide image in tumor neighborhoods of nine different sizes (200x200-pixel, 250x250-pixel and so on). Within each neighborhood, compute the orientation co-occurrence matrix where each row and column of the matrix denote the frequency of orientations having value row or column.

**Step 5: Quantitative measurement of the disorder in collagen fiber orientations**

After obtaining the orientation co-occurrence matrix for each neighborhood, we essentially use this matrix to compute the disorder in collagen fiber orientations using entropy theory.

**Comparison of TIL detection models**

We conducted an experiment to determine which TIL classification model we should use for the present work: SVM3 or Hover-Net4. For this, we randomly selected 80 tiles of size 3000x3000-pixel from the training cohort (D0). We applied both approaches to the tiles and asked two pathologists to select the model that provided the best TIL detection. One of the pathologists stated that the SVM model performed better in 52% of the tiles, Hover-Net in 28%, and there was not a significant difference in the remaining 20%. The second pathologist noted that the SVM was better in 75% of the cases, Hover-Net in 19%, and no significant difference was observed in 6% of the cases. We asked the pathologists for their overall impression of the models, and both of them agreed that while Hover-Net was very precise, it had lower recall. For this reason, we chose to use the SVM model.

**Signature Index**

In Single Sample Gene Set Enrichment Analysis (ssGSEA) enrichment analysis tools, the term "signature index" typically refers to a numerical score or metric that quantifies the extent to which a particular gene expression signature or gene set is enriched or depleted in a single sample or specimen. It summarizes certain characteristics or properties of data. In ssGSEA, the "signature index" assigned to a sample, indicating how closely its gene expression profile aligns with the gene expression pattern of a particular gene set or signature. It quantifies the degree of enrichment (positive index) or depletion (negative index) of the genes within the signature in the sample5.

We used signature index values to gain information about the pathways that are activated or suppressed in individual samples, helping us understand the underlying biology of the high risk and low risk cohorts in the context of the identified gene sets.

**False discovery rate (FDR)**

FDR stands for False Discovery Rate. In multiple hypothesis testing of gene expression data, the FDR addresses a specific challenge that traditional p-values or family-wise error rates (FWER) do not fully account for the control of false positives when testing multiple hypotheses simultaneously6.

**Supplementary Table 1. Quality check results of Epithelium/Stroma segmentation, Nuclei segmentation, Collagen fiber segmentation, and TIL detection methods**

|  |  |  |
| --- | --- | --- |
| **Task** | **Pathologist 1, Dr. Stefanie Avril**  **(% of tiles belonging to**  **good/fair category)** | **Pathologist 2, Dr. Mojgan Mokhtari**  **(% of tiles belonging to**  **good/fair category)** |
| Epithelium/Stroma segmentation | 90% | 94% |
| Nuclei  segmentation | 100% | 100% |
| Collagen fiber segmentation | 90% | 92% |
| TIL detection | 90% | 94% |

**Supplementary Table 2. Top Features contributing to CollaTIL signature**

|  |  |  |
| --- | --- | --- |
| **Feature index** | **Feature description** | **HR (per unit increase)** |
| 1 | Ratio of non-TILs density to the surrounding (20 microns proximity) TILs in the epithelium compartment | 1.17 |
| 2 | Number of epithelial TIL clusters surrounding (20 microns proximity) a non-TIL cluster in the epithelium compartment | 0.95 |
| 3 | Presence percentage (ratio of present clusters to total number of clusters) of stromal non-TIL clusters being around another non-TIL cluster in the stromal compartment | 1.26 |
| 4 | Intersected area of clusters of epithelial TILs and non-TILs in invasive tumor front compartment | 0.51 |
| 5 | Minimum area of stromal TIL clusters in invasive tumor front compartment | 1.78 |
| 6 | Range of area of epithelial non-TIL clusters in invasive tumor front compartment | 2.26 |
| 7 | Mean entropy value of the collagen fiber orientation disorder feature map using 200x200-pixel neighborhood | 0.75 |
| 8 | Minimum entropy value of the collagen fiber orientation disorder feature map using 200x200-pixel neighborhood | 0.5 |
| 9 | Maximum entropy value of the collagen fiber orientation disorder feature map using 250x250-pixel neighborhood | 0.94 |
| 10 | Minimum entropy value of the collagen fiber orientation disorder feature map using 350x350-pixel neighborhood | 0.84 |
| 11 | Minimum entropy value of the collagen fiber orientation disorder feature map using 400x400-pixel neighborhood | 0.64 |
| 12 | Minimum entropy value of the collagen fiber orientation disorder feature map using 450x450-pixel neighborhood | 1.64 |
| 13 | Maximum entropy value of the collagen fiber orientation disorder feature map using 550x550-pixel neighborhood | 1.38 |
| 14 | Maximum entropy value of the collagen fiber orientation disorder feature map using 600x600-pixel neighborhood | 0.36 |

**Supplementary Table 3. Hazard Ratios and P-values from univariate and multivariable analysis of OS (D1-D4) and PFS (D5-D8)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Clinicopathological Variables | **D1**  **HR**  **(p, 95% CI)** | **D2**  **HR**  **(p, 95% CI)** | **D3**  **HR**  **(p, 95% CI)** | **D4**  **HR**  **(p, 95% CI)** | **D5, D6, D7**  **HR**  **(p, 95% CI)** | **D8**  **HR**  **(p, 95% CI)** |
| **Univariate Analysis** |  |  |  |  |  |  |
| Age  (>60 years vs. <=60 years) | **HR = 3.21**  **(p=0.001, CI=1.19-8.63)** | HR=1.38  (p=0.401, CI=0.604-3.14) | **HR=3.57 (p=0.0295, CI=1.42-8.93)** | HR=1.38 (p=0.681, CI=0.311-6.13) | HR=0.879 (p=0.735, CI=0.401-1.92) | HR=1.7 (p=0.355, CI=0.547-5.3) |
| FIGO stage  (stage III/IV vs. stage I/II) | **HR=2.66 (p=0.0085, CI=1.04-6.75)** | **HR=2.24 (p=0.019, CI=1.01-5.21)** | HR=1.59 (p=0.295, CI=0.646-3.92) | HR=1  (p=1, CI=0-Inf) | HR=1.66 (p=0.153, CI=0.848-3.23) | HR=1  (p=1, CI=0-Inf) |
| Tumor grade  (high vs. low) | \*no data | \*no data | HR=2.26  (p=0.182, CI=0.85-6.04) | HR=3.13 (p=0.08, CI=0.62-18) | HR=0.985 (p=0.972, CI=0.5-2.26) | \*no data |
| CollaTIL  (high vs. low) | **HR=2.54 (p=0.0256, CI=1.21-5.34)** | **HR=2.87 (p=0.006, CI=1.49-5.53)** | **HR=7.36 (p=0.0213, CI=2.73-19.8)** | **HR=4.39 (p=0.0259, CI=1.07-18)** | **HR=2.72 (p=0.0184, CI=1.36-5.42)** | **HR=3.07 (p=0.0434, CI=1.01-9.85)** |
| Molecular subtypes  (Cnhigh vs. Cnlow) | **-** | **-** | HR=1.58 (p=0.3, CI=0.64-3.89) | HR=1.06 (p=0.9, CI=0.2-5.4) | **-** | **-** |
| HPV dependency data  (HPV+ vs. HPV-) | HR=1.18  (p=0.38, CI=1.1-6.9) | HR=1.47  (p=0.7,  CI=1.1-7.73) | - | - | **-** | **-** |
| **Multivariable Analysis** |  |  |  |  |  |  |
| Clinicopathological Variables | **D1**  **HR**  **(p, 95% CI)** | **D2**  **HR**  **(p, 95% CI)** | **D3**  **HR**  **(p, 95% CI)** | **D4**  **HR**  **(p, 95% CI)** | **D5, D6, D7**  **HR**  **(p, 95% CI)** | **D8**  **HR**  **(p, 95% CI)** |
| Age  (>60 years vs. <=60 years) | **HR=2.62**  **(p=0.02, CI=1.18-5.8)** | HR=1.28  (p=0.53, CI=0.59-2.8) | HR=3.39  (p=0.06, CI=0.95-12) | HR=1.08 (p=1, CI=0-Inf) | HR=1.09 (p=0.83, CI=0.5-2.4) | HR=1.43 (p=0.58, CI=0.4-5.2) |
| FIGO stage  (FIGO IV vs. FIGO III vs. FIGO II vs. FIGO I) | **HR=1.46**  **(p=0.04, CI=1.03-2.1)** | **HR=1.58**  **(p=0.01, CI=1.15-2.2)** | **HR=1.56**  **(p=0.048, CI=1.01-2.5)** | HR=1 (p=1, CI=0-Inf) | HR=1.21 (p=0.2, CI=0.9-1.6) | HR=1.09 (p=0.9, CI=0.3-4) |
| CollaTIL  (high vs. low) | **HR=2.58**  **(p=0.03, CI=1.09-6.1)** | **HR=3.25**  **(p=0.0001, CI=1.45-7.3)** | HR=6.16  (p=0.08, CI=0.79-48) | HR=9.35 (p=1, CI=0-Inf) | **HR=2.74 (p=0.03, CI=1.13-6.7)** | HR=1 (p=1, CI=0-Inf) |

\*HR, Hazard Ratio; CI, Confidence Interval

\*Bolded values indicate significant Hazard Ratios and P-values.

\*FIGO, International Federation of Gynecology and Obstetrics.

**Supplementary Table 4. Performance of different survival models on validation cohorts (D1-D8) by training on D0 cohort using CollaTIL features**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Survival Models** | **D1**  **(c-index)** | **D2**  **(c-index)** | **D3**  **(c-index)** | **D4**  **(c-index)** | **D5, D6, D7**  **(c-index)** | **D8**  **(c-index)** |
| CPRM + Elastic Penalty | 0.7 | 0.67 | 0.68 | 0.75 | 0.58 | 0.64 |
| CPRM | 0.58 | 0.65 | 0.65 | 0.74 | 0.58 | 0.66 |
| Random survival forest | 0.58 | 0.55 | 0.6 | 0.6 | 0.58 | 0.62 |
| Survival SVM | 0.68 | 0.67 | 0.65 | 0.81 | 0.6 | 0.73 |

\*CPRM, Cox proportional regression model; SVM, Support vector machine

**Supplementary Table 5. Disease type sensitivity analysis of the model performance using different training cohorts**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Experiment** | **Training cohort used** | **D0**  **(c-index)** | **D1**  **(c-index)** | **D2**  **(c-index)** | **D3**  **(c-index)** | **D4**  **(c-index)** | **D5, D6, D7 (c-index)** | **D8**  **(c-index)** |
| Experiment 1 | D0 | - | 0.7 | 0.67 | 0.68 | 0.75 | 0.58 | 0.64 |
| Experiment 2 | D2 | 0.58 | 0.67 | - | 0.61 | 0.74 | 0.55 | 0.55 |
| Experiment 3 | D1 | 0.6 | - | 0.71 | 0.64 | 0.75 | 0.61 | 0.55 |

\*Experiment 1: original training cohort used for study (D0); Experiment 2: training cohort used was D2 cohort; Experiment 3: training cohort used was D1 cohort).

**Supplementary Table 6. List of gene sets for Macrophage gene signature and Amino acid signature**

|  |  |
| --- | --- |
| Macrophages\_Gene Signature | Amino\_acid\_Gene Signature |
| |  | | --- | | ADORA3 | | ATP8B4 | | C1QB | | C3AR1 | | C5AR1 | | CD163 | | CD300A | | FCGR2A | | LIPA | | LY96 | | MSR1 | | SLCO2B1 | | VSIG4 | | ADAMDEC1 | | AOAH | | ARRB2 | | BCL2A1 | | C1ORF54 | | C1QA | | C2 | | CCR1 | | CCRL2 | | CD4 | | CD68 | | CD74 | | CD86 | | ADA2 | | CLEC7A | | CMKLR1 | | CSF1R | | CTSB | | CTSS | | CYBB | | CYTH4 | | DPYD | | ADGRE2 | | FCER1G | | FCGR1A | | FCGR1B | | FCGR3B | | FPR3 | | GPNMB | | HK3 | | HLA-DRB6 | | IFI30 | | IGSF6 | | ITGAM | | ITGAX | | ITGB2 | | LAIR1 | | LAPTM5 | | LILRB4 | | MAN2B1 | | MFSD1 | | MNDA | | MS4A4A | | MS4A7 | | MYO1F | | NCKAP1L | | NPL | | NR1H3 | | PLA2G7 | | PLEKHO2 | | SCPEP1 | | SLAMF8 | | SLC15A3 | | SLC31A2 | | SNX10 | | SPI1 | | TBXAS1 | | TLR8 | | TMEM140 | | TNFAIP2 | | TNFRSF1B | | TNFSF13B | | TRPV2 | | TYMP | | TYROBP | | |  | | --- | | CARNMT1 | | DUOX1 | | HIBADH | | OAT | | PSMA7 | | PSMC2 | | PSMC4 | | PSMD11 | | PSMD8 | | RPL28 | | RPS16 | | RPS19 | | SERINC4 | | SLC6A11 | | AASS | | AIMP2 | | AADAT | | AANAT | | ACAD8 | | ACADSB | | ACAT1 | | ACMSD | | ADI1 | | ADO | | AFMID | | AGMAT | | AGXT | | AGXT2 | | AHCY | | AIMP1 | | ALDH18A1 | | ALDH4A1 | | ALDH6A1 | | ALDH7A1 | | ALDH9A1 | | AMD1 | | AMDHD1 | | AMT | | APIP | | ARG1 | | ARG2 | | ASL | | ASMT | | ASNS | | ASPA | | ASPG | | ASRGL1 | | ASS1 | | AUH | | AZIN1 | | AZIN2 | | BBOX1 | | BCAT1 | | BCAT2 | | BCKDHA | | BCKDHB | | BCKDK | | BHMT | | BHMT2 | | CARNS1 | | CBS | | CDO1 | | CGA | | CKB | | CKM | | CKMT1A | | CKMT2 | | CNDP2 | | CPS1 | | CRYM | | CSAD | | CTH | | DAO | | DARS1 | | DBH | | DBT | | DCT | | DDC | | DDO | | DHTKD1 | | DIO1 | | DIO2 | | DIO3 | | DLAT | | DLD | | DLST | | DUOX2 | | EEF1E1 | | EEFSEC | | ENOPH1 | | ETHE1 | | FAH | | FAU | | FOLH1 | | FTCD | | GADL1 | | GAMT | | GATM | | GCAT | | GCDH | | GCLC | | GCLM | | GCSH | | GLDC | | GLS | | GLS2 | | GLUD1 | | GLUD2 | | GLUL | | GNMT | | GOT1 | | GOT2 | | GPT | | GPT2 | | GRHPR | | GSR | | GSTZ1 | | HAL | | HAO1 | | HDC | | HGD | | HIBCH | | HNMT | | HPD | | HSD17B10 | | IARS1 | | IDO1 | | IDO2 | | IL4I1 | | INMT | | IVD | | IYD | | KARS1 | | KMO | | KYAT1 | | KYAT3 | | KYNU | | LARS1 | | LIAS | | LIPT1 | | LIPT2 | | MARS1 | | MAT1A | | MCCC1 | | MCCC2 | | MRI1 | | MTAP | | MTR | | MTRR | | NAALAD2 | | NAGS | | NDUFAB1 | | NMRAL1 | | NNMT | | NQO1 | | OAZ1 | | OAZ2 | | OAZ3 | | OCA2 | | ODC1 | | OGDH | | OTC | | PAH | | PAOX | | PAPSS1 | | PAPSS2 | | PCBD1 | | PDHA1 | | PDHB | | PDHX | | PHGDH | | PIPOX | | PNMT | | PPM1K | | PRODH | | PSAT1 | | PSMA1 | | PSMA2 | | PSMA3 | | PSMA4 | | PSMA5 | | PSMA6 | | PSMA8 | | PSMB1 | | PSMB10 | | PSMB11 | | PSMB2 | | PSMB3 | | PSMB4 | | PSMB5 | | PSMB6 | | PSMB7 | | PSMB8 | | PSMB9 | | PSMC1 | | PSMC3 | | PSMC5 | | PSMC6 | | PSMD1 | | PSMD10 | | PSMD12 | | PSMD13 | | PSMD14 | | PSMD2 | | PSMD3 | | PSMD4 | | PSMD5 | | PSMD6 | | PSMD7 | | PSMD9 | | PSME1 | | PSME2 | | PSME3 | | PSME4 | | PSMF1 | | PSPH | | PSTK | | PYCR1 | | PYCR2 | | PYCR3 | | QDPR | | RARS1 | | RPL10 | | RPL10A | | RPL10L | | RPL11 | | RPL12 | | RPL13 | | RPL13A | | RPL14 | | RPL15 | | RPL17 | | RPL18 | | RPL18A | | RPL19 | | RPL21 | | RPL22 | | RPL22L1 | | RPL23 | | RPL23A | | RPL24 | | RPL26 | | RPL26L1 | | RPL27 | | RPL27A | | RPL29 | | RPL3 | | RPL30 | | RPL31 | | RPL32 | | RPL34 | | RPL35 | | RPL35A | | RPL36 | | RPL36A | | RPL36AL | | RPL37 | | RPL37A | | RPL38 | | RPL39 | | RPL39L | | RPL3L | | RPL4 | | RPL41 | | RPL5 | | RPL6 | | RPL7 | | RPL7A | | RPL8 | | RPL9 | | RPLP0 | | RPLP1 | | RPLP2 | | RPS10 | | RPS11 | | RPS12 | | RPS13 | | RPS14 | | RPS15 | | RPS15A | | RPS17 | | RPS18 | | RPS2 | | RPS20 | | RPS21 | | RPS23 | | RPS24 | | RPS25 | | RPS26 | | RPS27 | | RPS27A | | RPS27L | | RPS28 | | RPS29 | | RPS3 | | RPS3A | | RPS4X | | RPS4Y1 | | RPS4Y2 | | RPS5 | | RPS6 | | RPS7 | | RPS8 | | RPS9 | | RPSA | | SARS1 | | SAT1 | | SCLY | | SECISBP2 | | SEM1 | | SEPSECS | | SERINC1 | | SERINC2 | | SERINC3 | | SERINC5 | | SHMT1 | | SLC25A10 | | SLC25A15 | | SLC25A2 | | SLC25A21 | | SLC3A2 | | SLC45A2 | | SLC5A5 | | SLC6A12 | | SLC6A7 | | SLC6A8 | | SLC7A5 | | SMOX | | SMS | | SQOR | | SRM | | SUOX | | TAT | | TDO2 | | TH | | TMLHE | | TPH1 | | TPH2 | | TPO | | TSHB | | TST | | TXNRD1 | | TYR | | TYRP1 | | UBA52 | | UROC1 | |

**Supplementary Table 7.** Features from TIL component

|  |  |
| --- | --- |
| **Feature index** | **Feature description** |
| 1 | Ratio of non-TILs density to the surrounding (20 microns proximity) TILs in the epithelium compartment |
| 2 | Number of epithelial TIL clusters surrounding (20 microns proximity) a non-TIL cluster in the epithelium compartment |
| 3 | Presence percentage (ratio of present clusters to total number of clusters) of stromal non-TIL clusters being around another non-TIL cluster in the stromal compartment |
| 4 | Intersected area of clusters of epithelial TILs and non-TILs in invasive tumor front compartment |
| 5 | Minimum area of stromal TIL clusters in invasive tumor front compartment |
| 6 | Range of area of epithelial non-TIL clusters in invasive tumor front compartment |
| 7 | Range of density of TIL clusters to the surrounding non-TIL ones in stroma |

**Supplementary Table 8.** Features from collagen component

|  |  |
| --- | --- |
| **Feature index** | **Feature description** |
| 1 | Mean entropy value of the collagen fiber orientation disorder feature map using 200x200-pixel neighborhood |
| 2 | Minimum entropy value of the collagen fiber orientation disorder feature map using 200x200-pixel neighborhood |
| 3 | Maximum entropy value of the collagen fiber orientation disorder feature map using 200x200-pixel neighborhood |
| 4 | Mean entropy value of the collagen fiber orientation disorder feature map using 250x250-pixel neighborhood |
| 5 | Minimum entropy value of the collagen fiber orientation disorder feature map using 250x250-pixel neighborhood |
| 6 | Maximum entropy value of the collagen fiber orientation disorder feature map using 250x250-pixel neighborhood |
| 7 | Mean entropy value of the collagen fiber orientation disorder feature map using 300x300-pixel neighborhood |
| 8 | Minimum entropy value of the collagen fiber orientation disorder feature map using 300x300-pixel neighborhood |
| 9 | Maximum entropy value of the collagen fiber orientation disorder feature map using 300x300-pixel neighborhood |
| 10 | Mean entropy value of the collagen fiber orientation disorder feature map using 350x350-pixel neighborhood |
| 11 | Minimum entropy value of the collagen fiber orientation disorder feature map using 350x350-pixel neighborhood |
| 12 | Maximum entropy value of the collagen fiber orientation disorder feature map using 350x350-pixel neighborhood |
| 13 | Mean entropy value of the collagen fiber orientation disorder feature map using 400x400-pixel neighborhood |
| 14 | Minimum entropy value of the collagen fiber orientation disorder feature map using 400x400-pixel neighborhood |
| 15 | Maximum entropy value of the collagen fiber orientation disorder feature map using 400x400-pixel neighborhood |
| 16 | Mean entropy value of the collagen fiber orientation disorder feature map using 450x450-pixel neighborhood |
| 17 | Minimum entropy value of the collagen fiber orientation disorder feature map using 450x450-pixel neighborhood |
| 18 | Maximum entropy value of the collagen fiber orientation disorder feature map using 450x450-pixel neighborhood |
| 19 | Mean entropy value of the collagen fiber orientation disorder feature map using 500x500-pixel neighborhood |
| 20 | Minimum entropy value of the collagen fiber orientation disorder feature map using 500x500-pixel neighborhood |
| 21 | Maximum entropy value of the collagen fiber orientation disorder feature map using 500x500-pixel neighborhood |
| 22 | Mean entropy value of the collagen fiber orientation disorder feature map using 550x550-pixel neighborhood |
| 23 | Minimum entropy value of the collagen fiber orientation disorder feature map using 550x550-pixel neighborhood |
| 24 | Maximum entropy value of the collagen fiber orientation disorder feature map using 550x550-pixel neighborhood |
| 25 | Mean entropy value of the collagen fiber orientation disorder feature map using 600x600-pixel neighborhood |
| 26 | Minimum entropy value of the collagen fiber orientation disorder feature map using 600x600-pixel neighborhood |
| 27 | Maximum entropy value of the collagen fiber orientation disorder feature map using 600x600-pixel neighborhood |

A picture containing text, screenshot, purple, font

Description automatically generated

**Supplementary Fig. 1: Expression of the genes that are strongly correlated with the risk scores predicted based on CollaTIL.** Y axis indicates log2 FPKM values of the mRNA expression of the top correlated genes on TCGA dataset. X axis indicates the risk-scores predicted based on CollaTIL. R indicates Pearson correlation coefficients.

A picture containing text, diagram, map, screenshot

Description automatically generated

**Supplementary Fig. 2: Kaplan-Meier curves of TCA-Cycle related genes, *NDUFA1, NDUFA12, COX7B* and *CD8A* that are significantly upregulated and associated with a better prognosis in CSCC patients.** The statistical significance of differences in survival rates between high expressed and low expressed categories was determined using the LogRank test (P).

**References**

1. Cox, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34**, 187–202 (1972).

2. Wu, Y. Elastic net for Cox’s proportional hazards model with a solution path algorithm. *Stat Sin* **22**, 271–294 (2012).

3. Romero Castro, E. *et al.* A watershed and feature-based approach for automated detection of lymphocytes on lung cancer images. in 26 (SPIE-Intl Soc Optical Eng, 2018). doi:10.1117/12.2293147.

4. Graham, S. *et al.* Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* **58**, (2019).

5. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).

6. Benjamini, Y. & Hochberg, Y. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. *Source: Journal of the Royal Statistical Society. Series B (Methodological)* vol. 57 (1995).