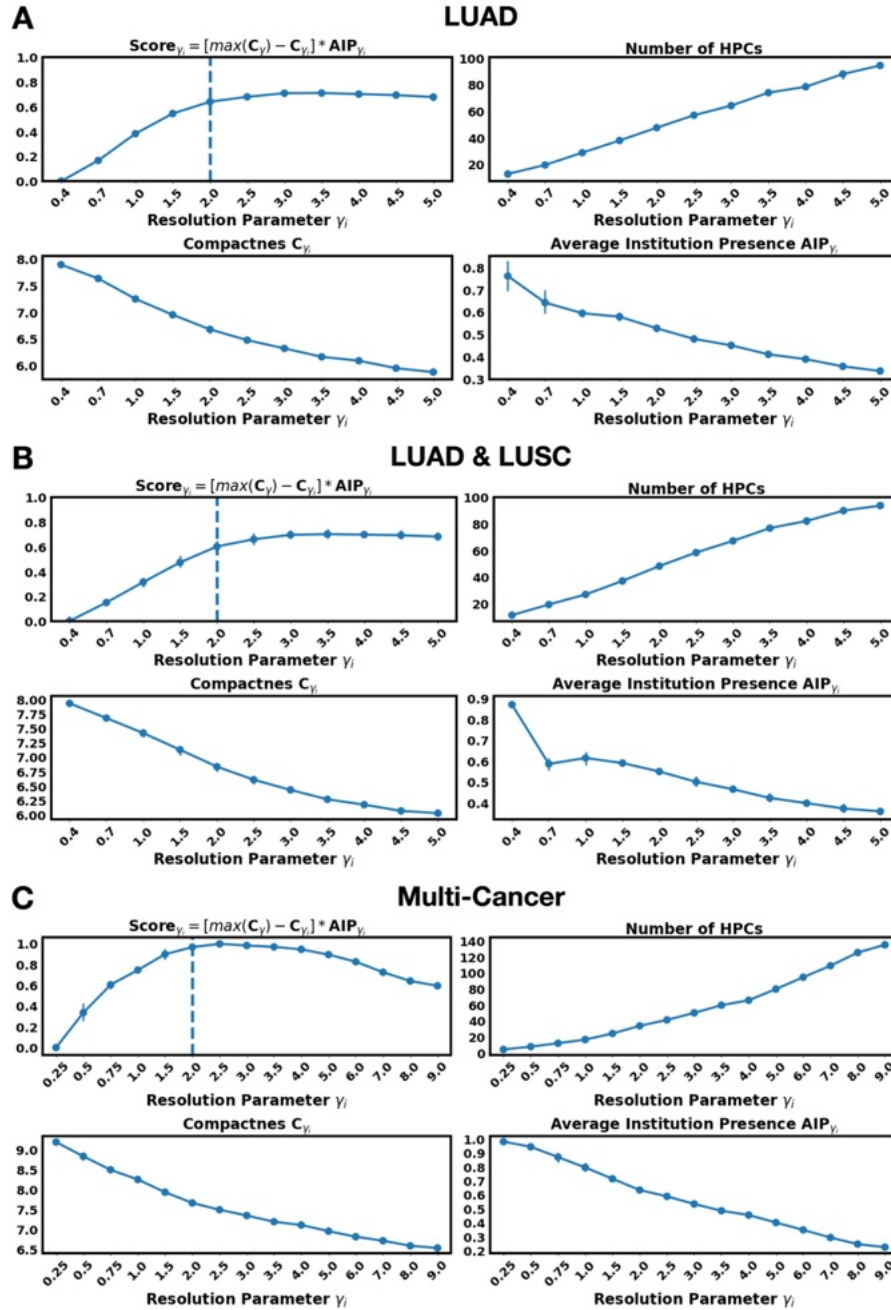
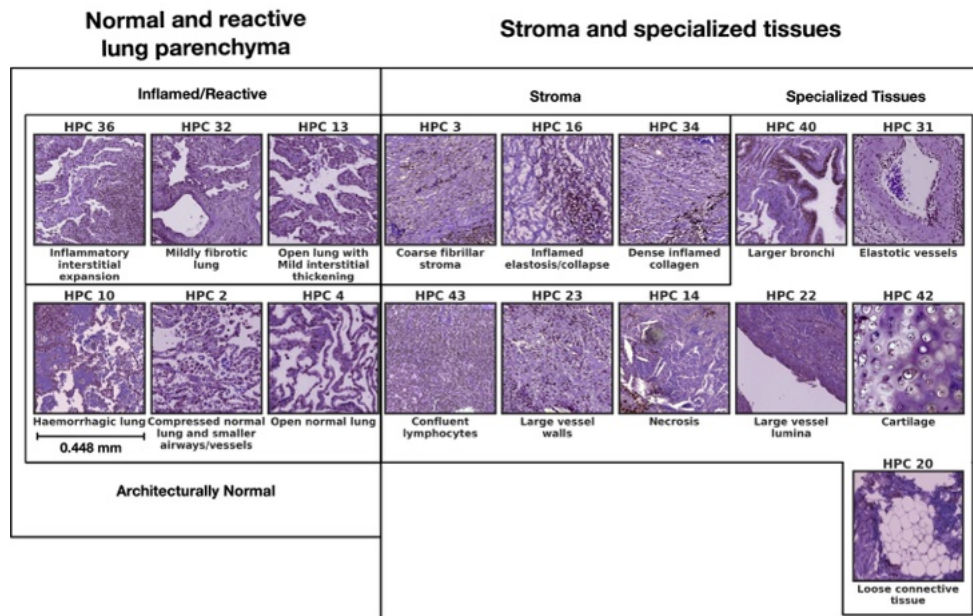


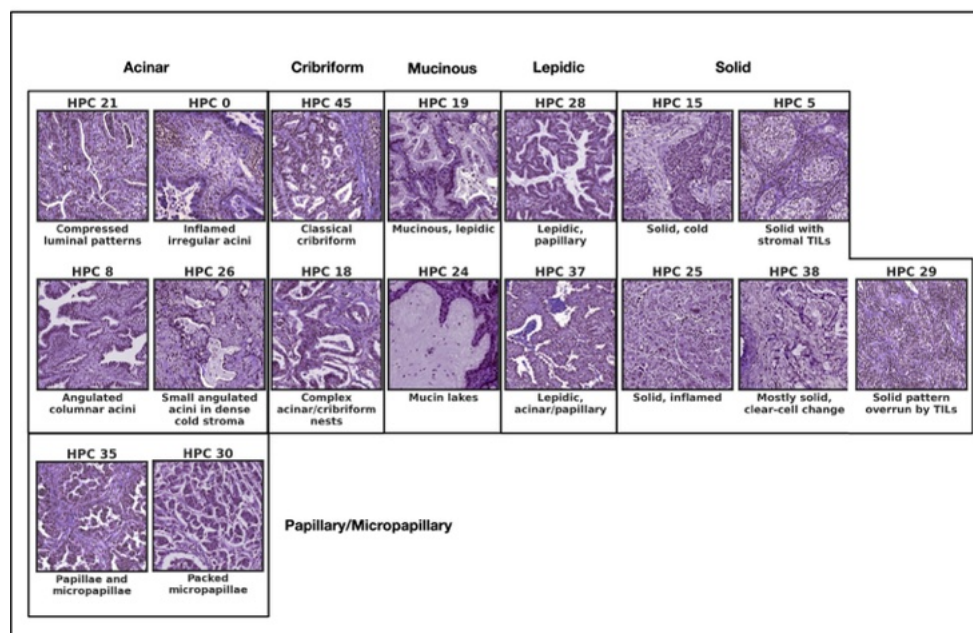
Supplementary Information



Supplementary Figure 1. Selection of number of HPCs. Graphs are shown for **A** the LUAD study, **B** the LUAD vs LUSC classification task and **C** the multi-cancer analysis. For each of them, we show the final score corresponding to **Equation 6** (top left panel), the resulting number of HPCs for each Leiden resolution (top right), and the two trade-off elements forming the final score: the compactness parameters associated with **Equation 4** (bottom left) and the average presence of institutions per HPC from **Equation 5**. The dotted line shows the optimal Leiden resolution. We selected the optimal Leiden resolution where the score function shows a larger trend change and begins saturation (elbow method or knee of a curve). Source data are provided as a Source Data file.

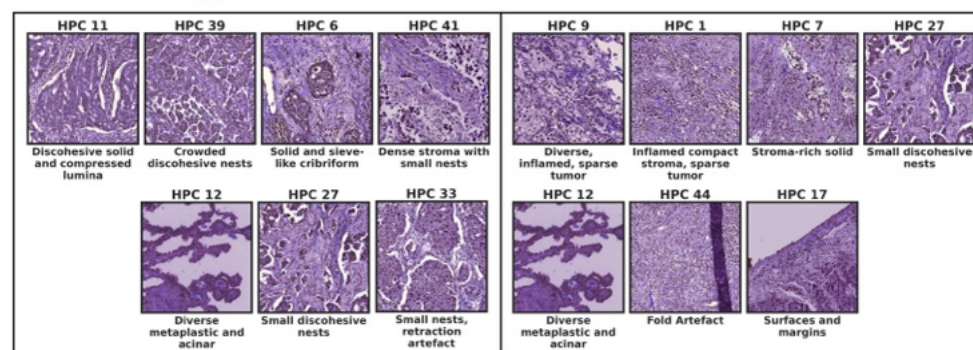


Classical adenocarcinoma appearances

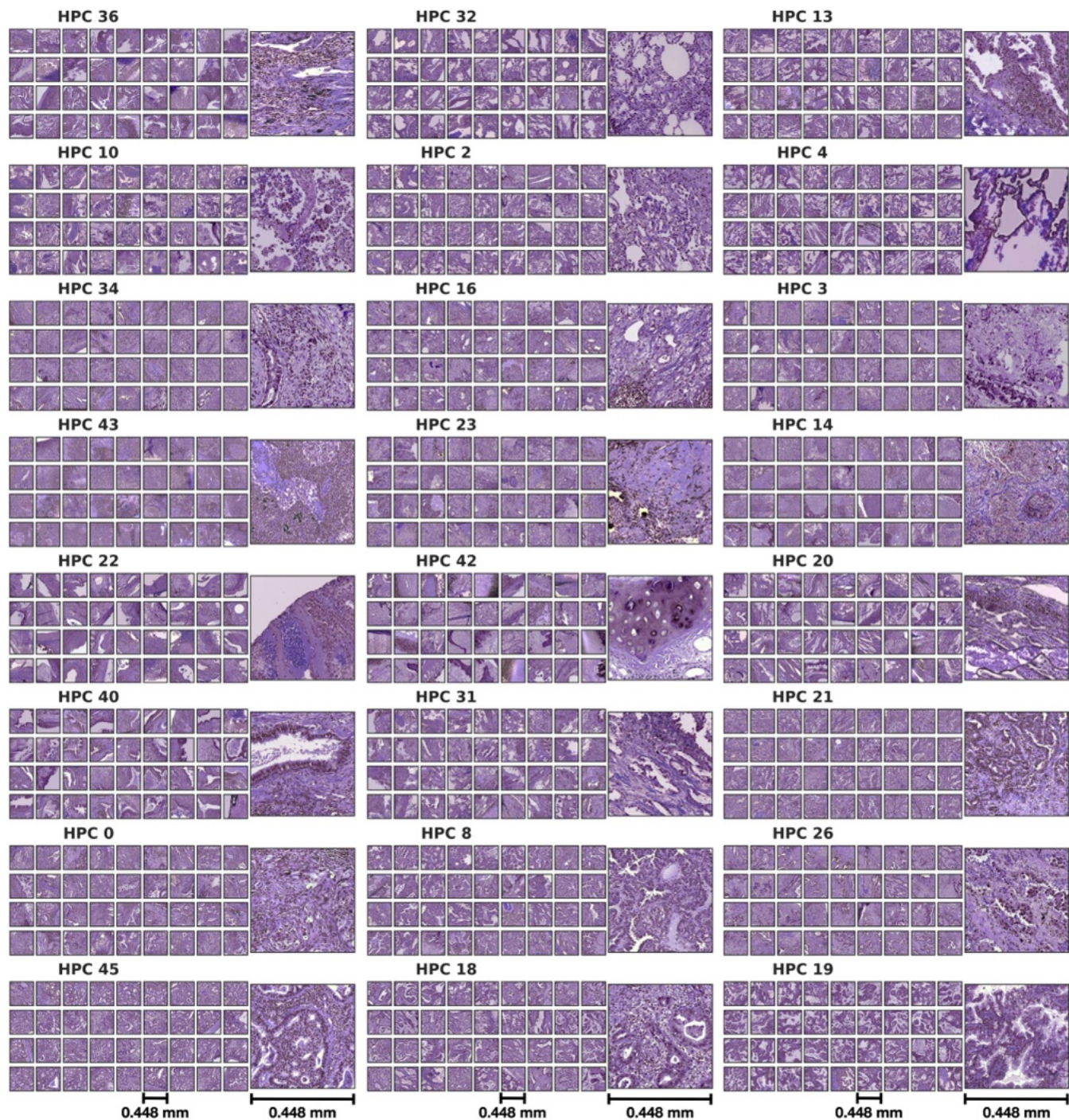


Variant adenocarcinoma appearances

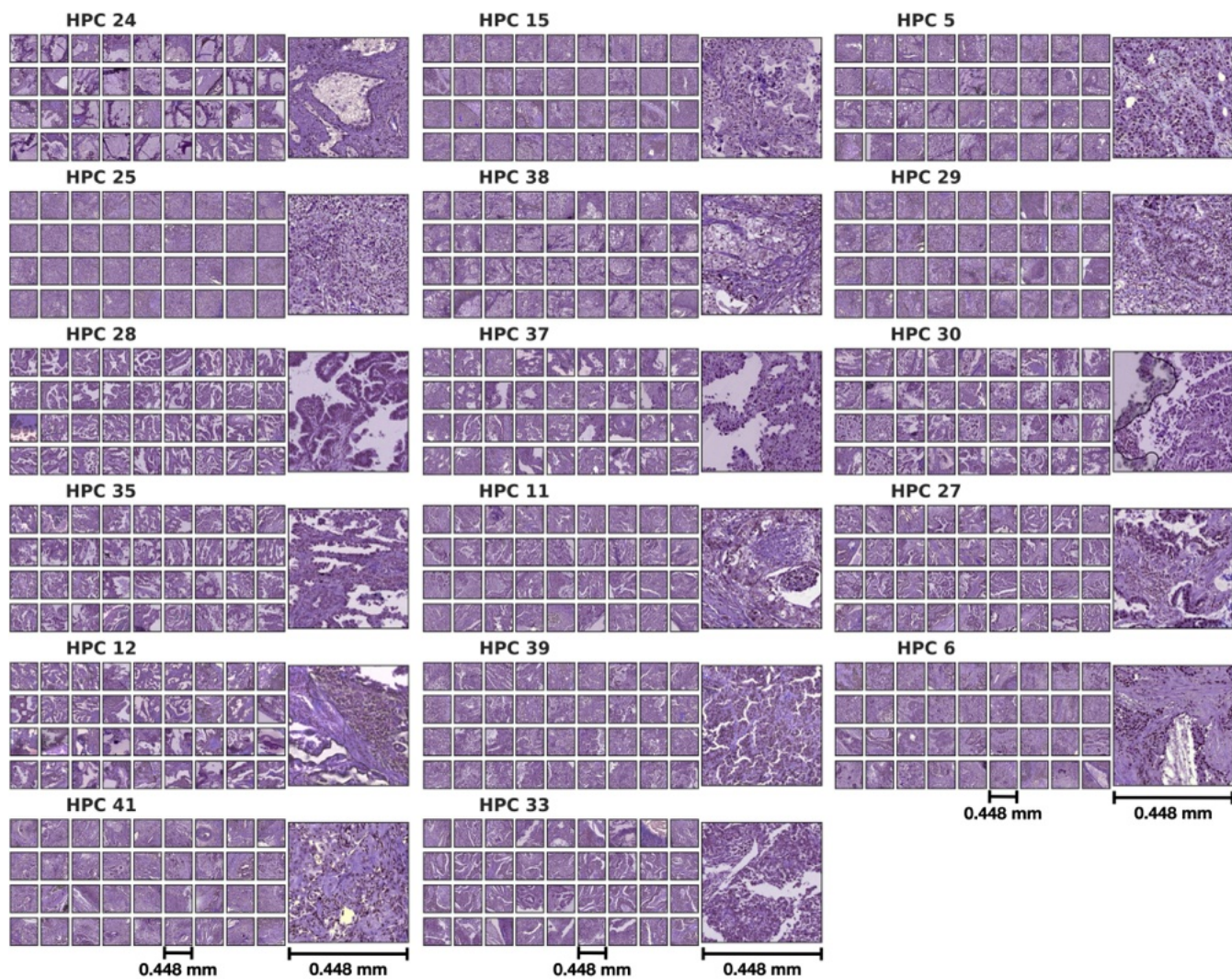
Stroma-predominant tumor and artefacts



Supplementary Figure 2. Consensus description of HPCs with their representative tiles. Source data are provided as a Source Data file.

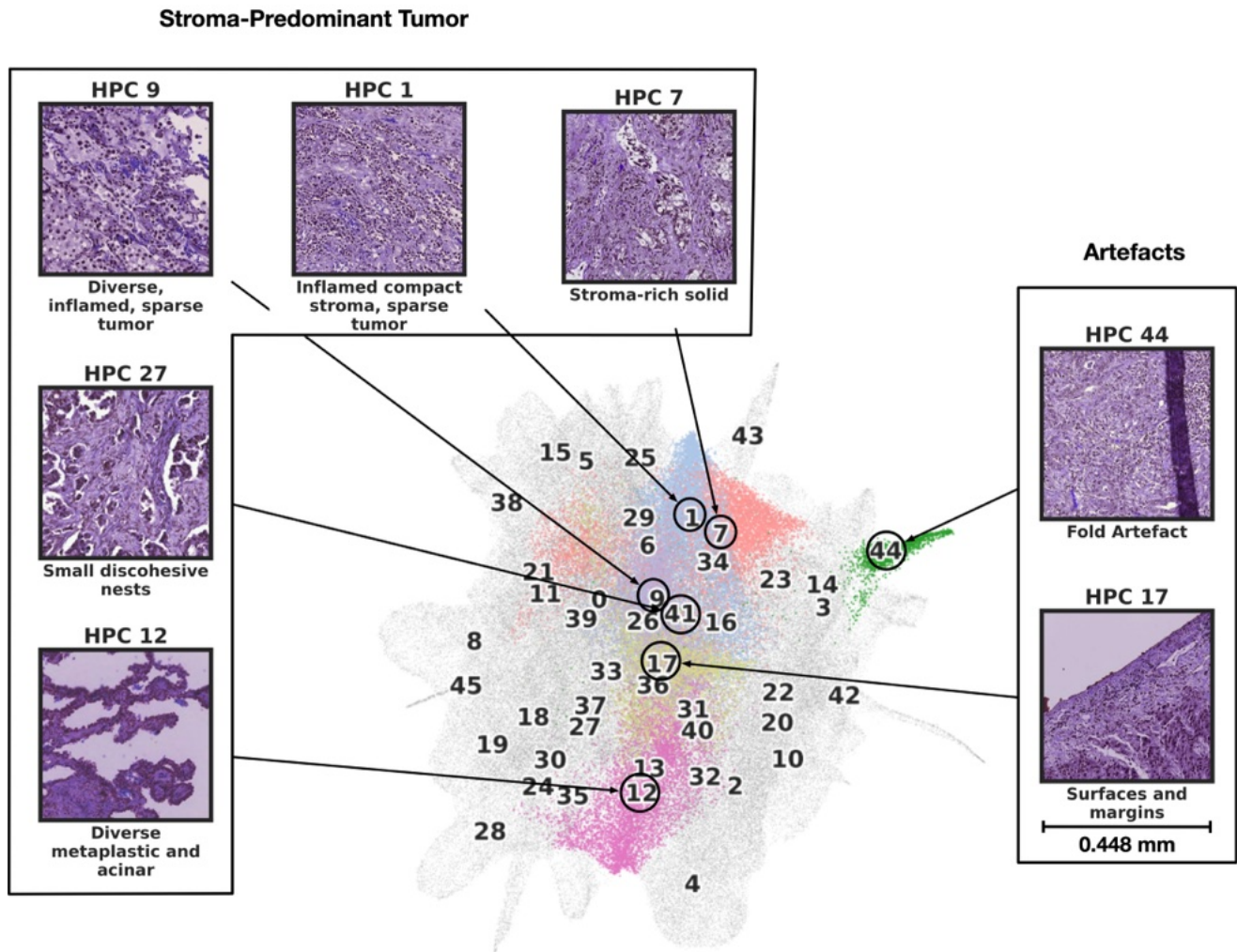


Supplementary Figure 3. More examples of tiles from HPC discussed in Figures 3 and 4. Clusters are ordered as they appear in those two main figures. Source data are provided as a Source Data file.

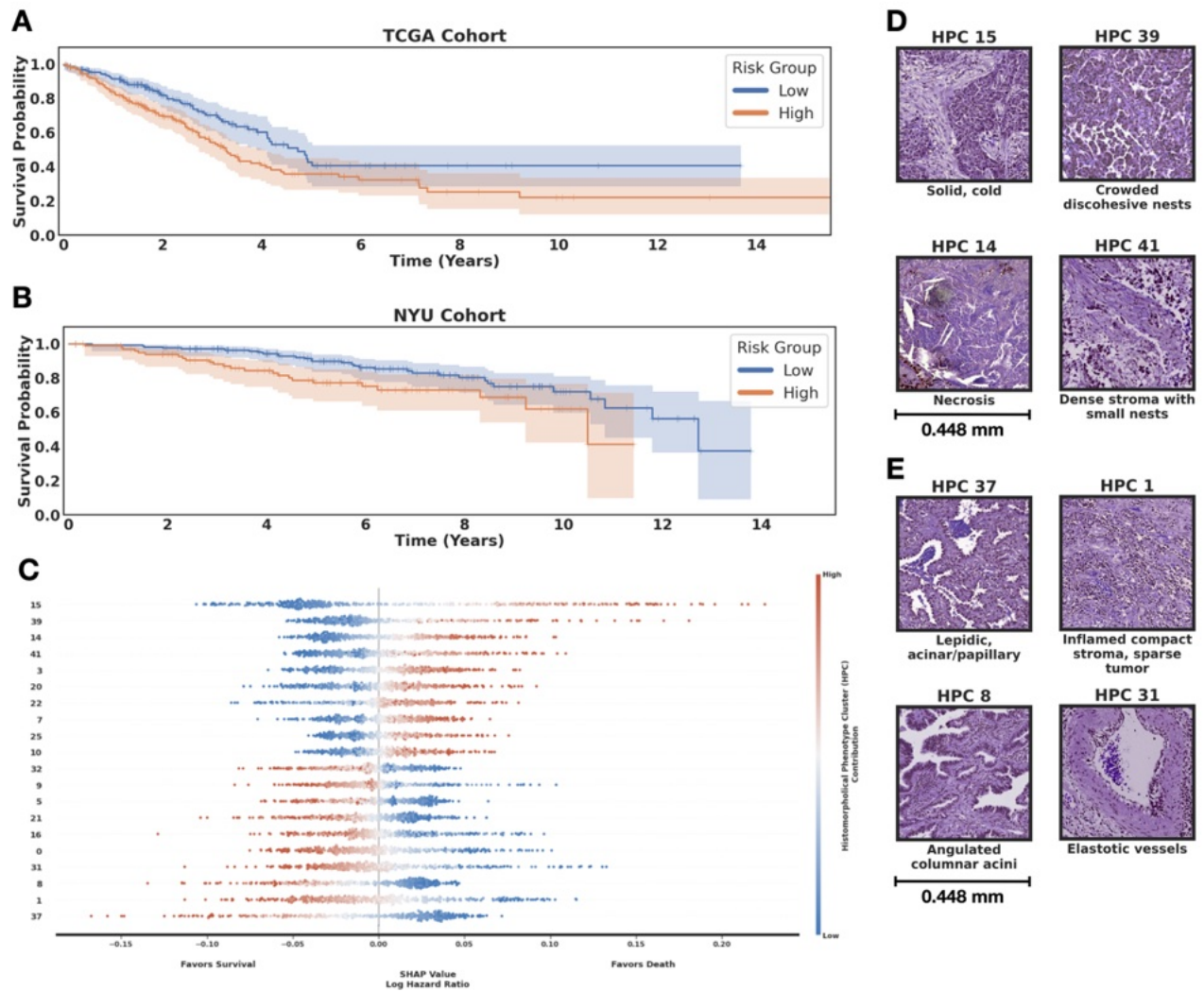


Supplementary Figure 4. More examples of tiles from HPC discussed in Figures 3 and 4. Clusters are ordered as they appear in those two main figures. Source data are provided as a Source Data file.

Stroma-predominant tumor and artefacts

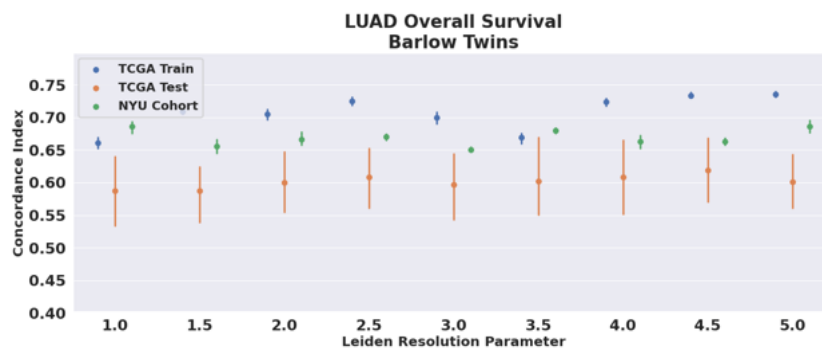


Supplementary Figure 5. Consensus description of HPCs enriched in stroma-predominant tumor and artefacts with their representative tiles). Source data are provided as a Source Data file.

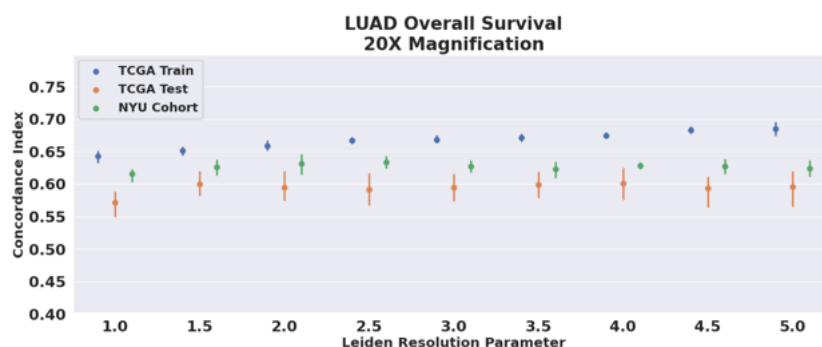


Supplementary Figure 6. Lung adenocarcinoma overall survival. **A.** SHAP (SHapley Additive exPlanations) plot showing how HPC (y-axis) weight on survival outcome. **B.** High and low risk groups on TCGA cohort showing statistical significance ($p\text{-value } 3.7 \times 10^{-3} < 0.05$) and c-index of 0.60. **C.** High and low risk groups on NYU_1 cohort showing statistical significance ($p\text{-value } 1.6 \times 10^{-2} < 0.05$) and c-index of 0.65. **D.** Representative tiles for HPCs associated with poor survival in panel C. **E.** Representative tiles for HPCs associated with good survival in panel C. Source data are provided as a Source Data file.

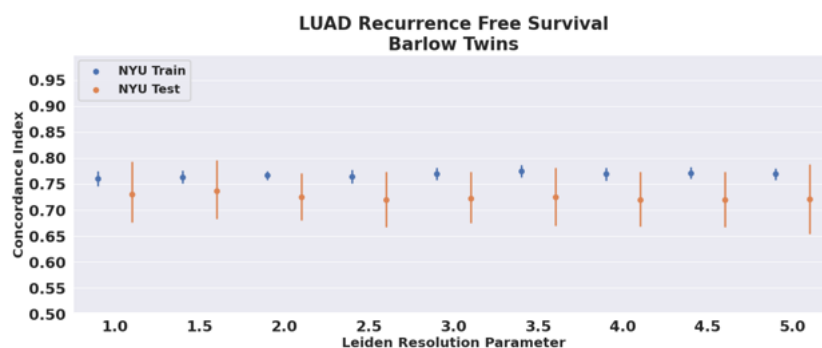
A



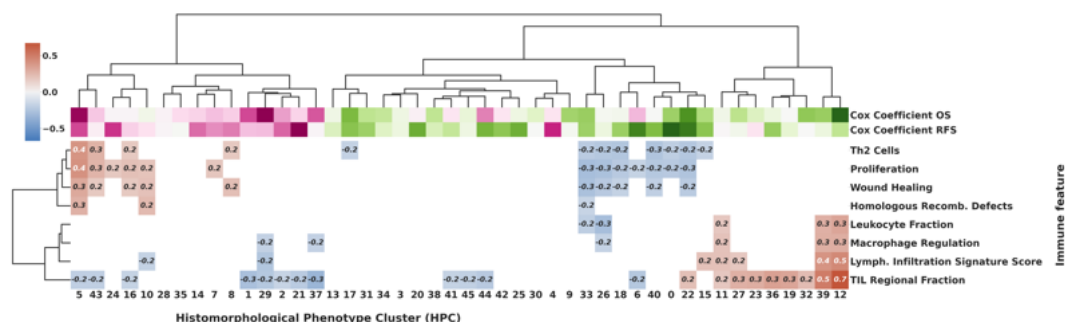
B



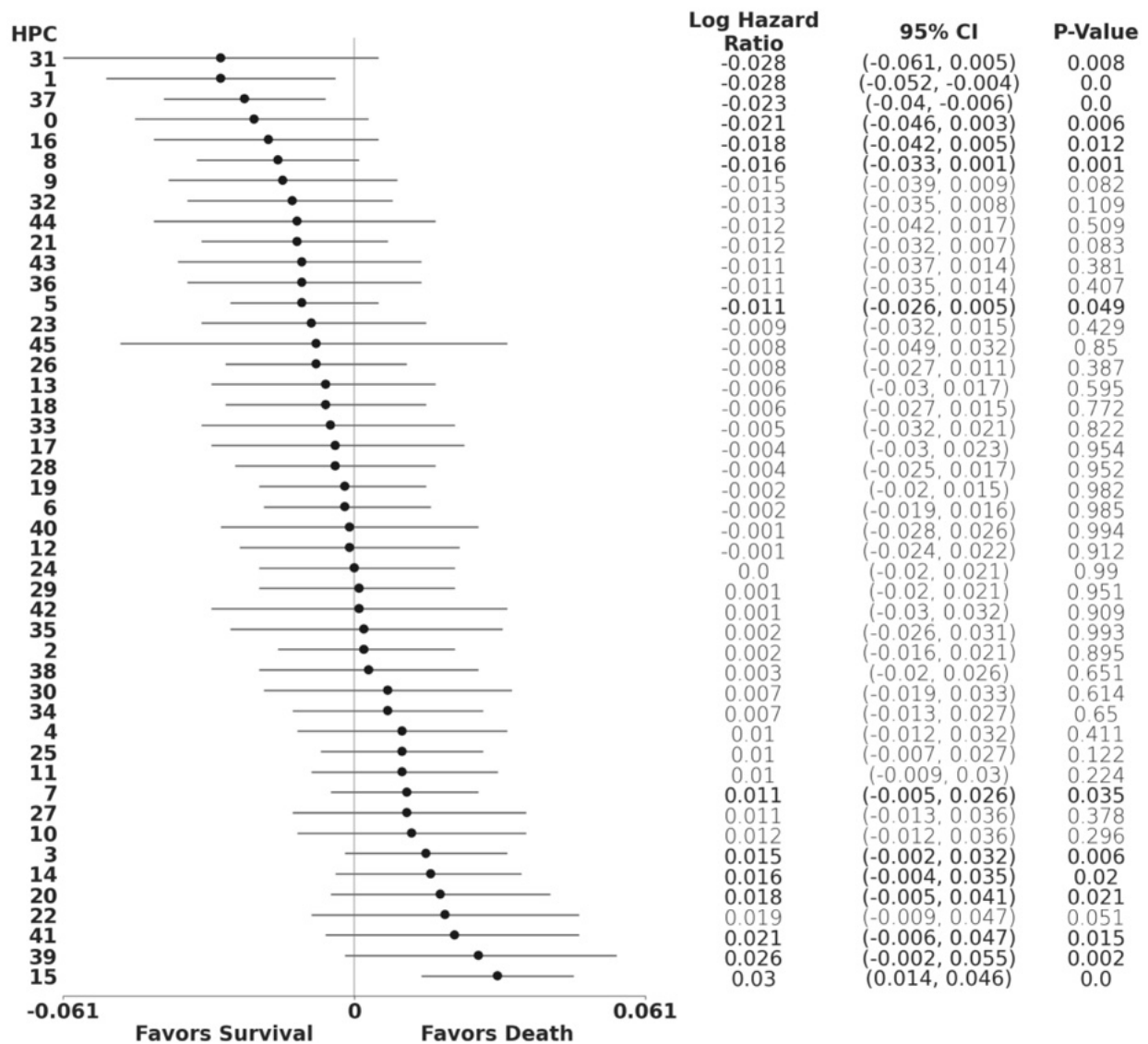
C



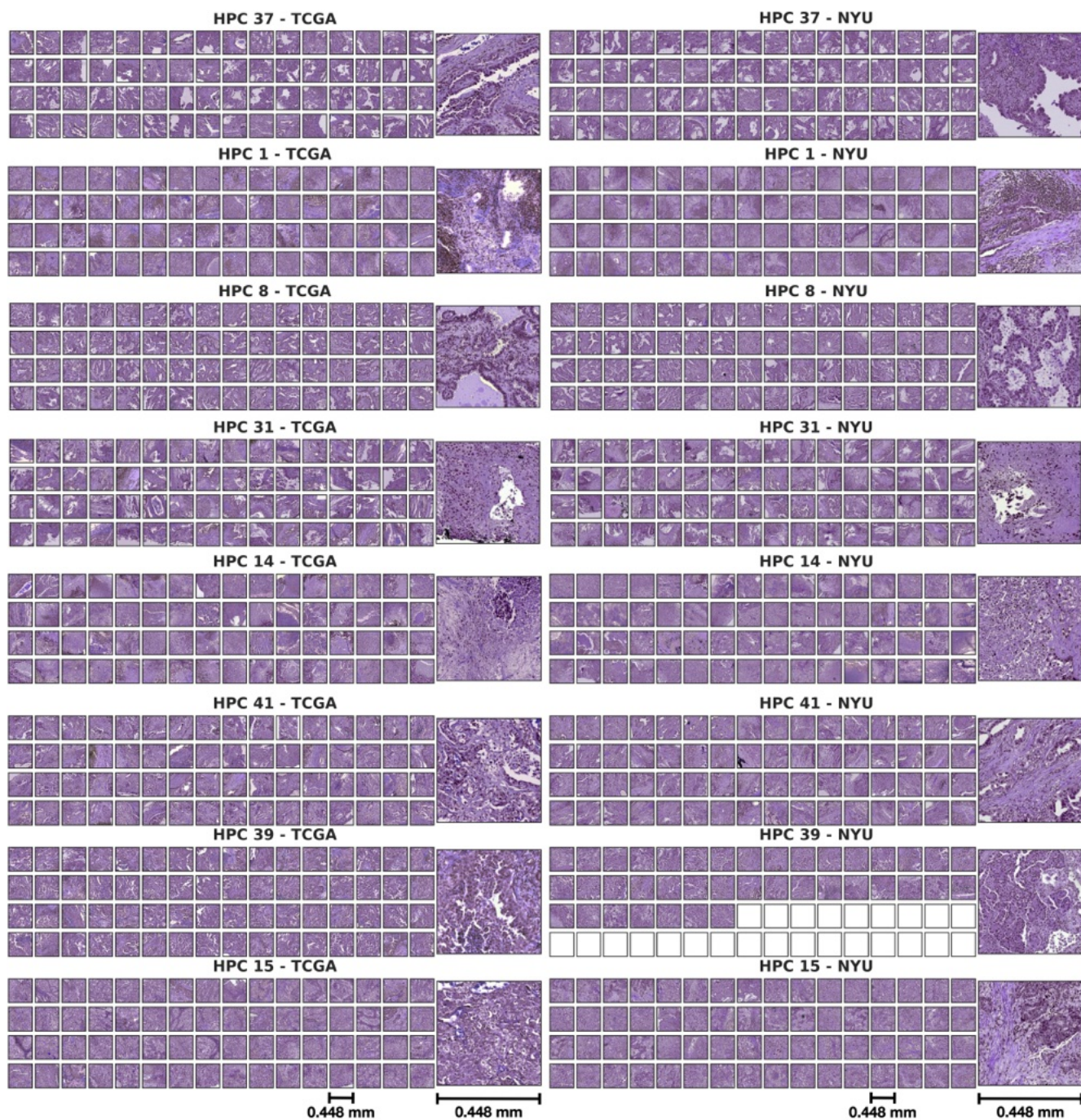
D



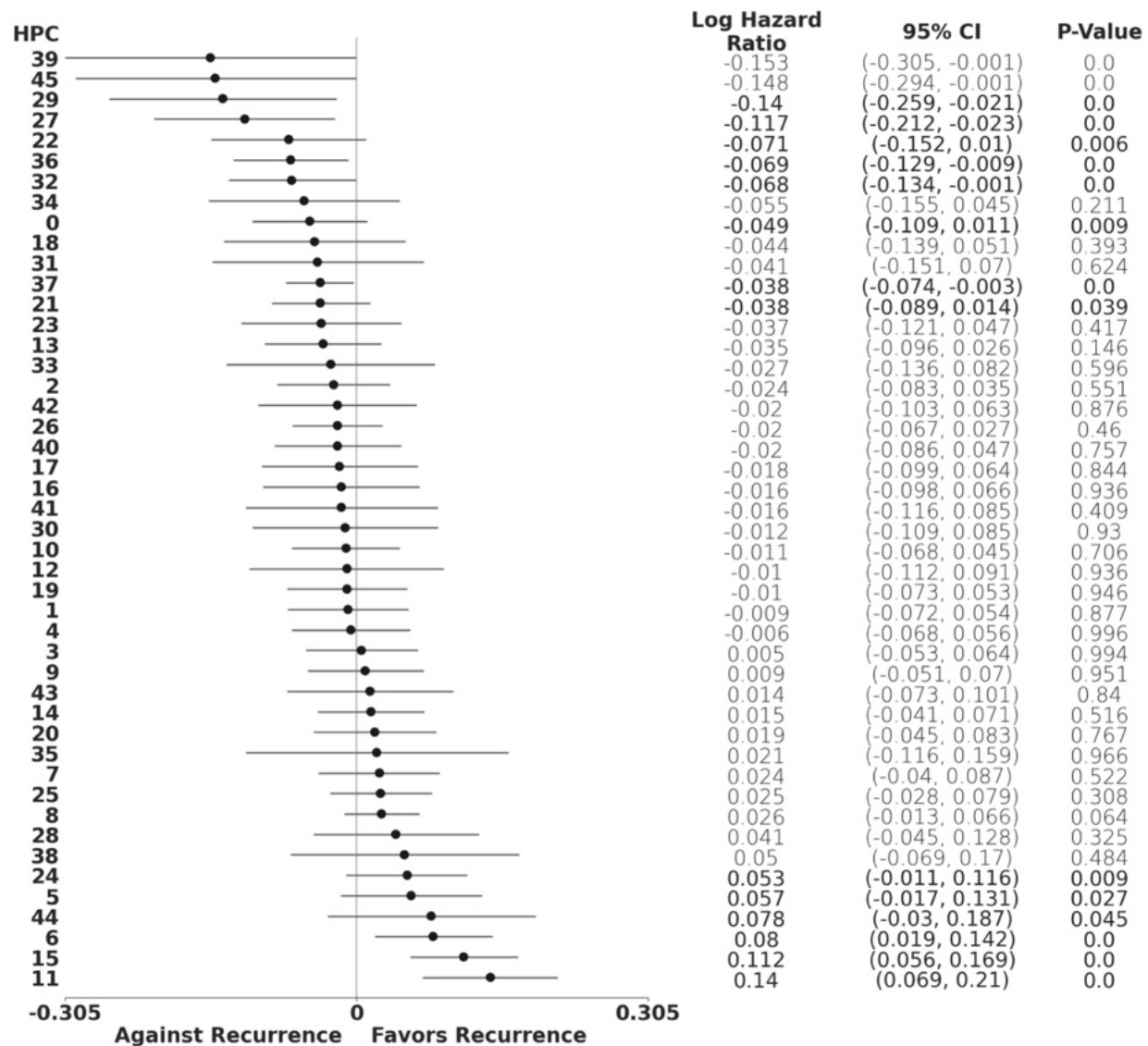
Supplementary Figure 7. Survival analysis at different resolutions and after training Barlow-Twins without sub-sampling the training dataset. **A** OS measured from HPL run at 5x. **B** OS measured from HPL run at 20x. **C** RFS measured from HPL run at 5x. **D** Bi-hierarchical clustering of HPCs and immune signature correlations where positive correlations are shown in red and negative as blue, corresponding to run at 20x and Leiden resolution of 2.0. Cox coefficients for overall and recurrence free survival are colored with a gradient scheme, from favoring death or recurrence as purple to favoring survival or no recurrence as green. All runs were done following a 5-fold cross-validation split similar to the one used in previous studies^{1,2}. Source data are provided as a Source Data file.



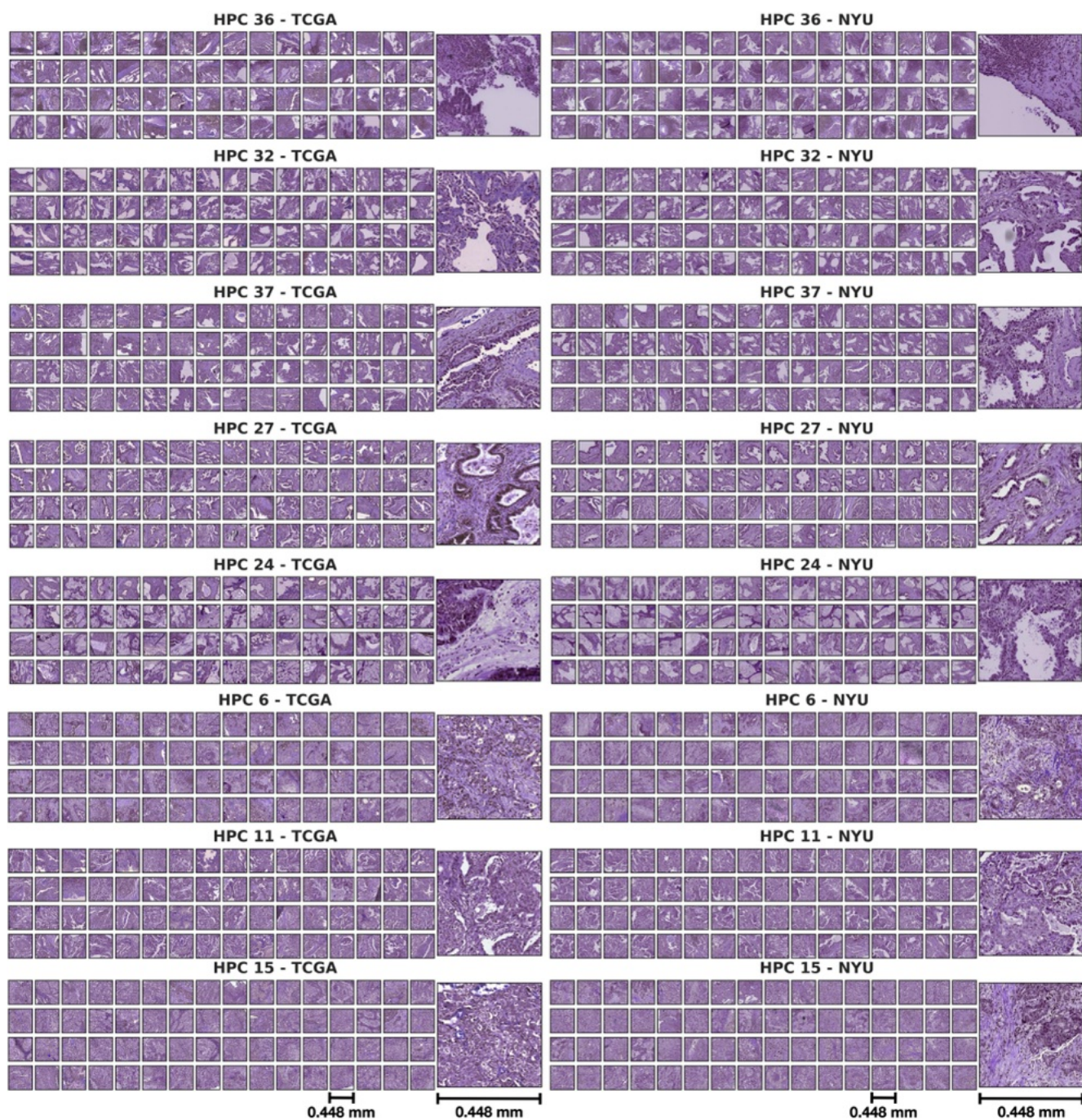
Supplementary Figure 8. Lung adenocarcinoma overall survival Forest plot. Log hazard ratio of Cox proportional hazards model over the train sets of a 5-fold cross-validation over TCGA data and NYU as independent cohort. We averaged coefficients across fold and combined p-values with Fisher's combined probability test. Statistically significant HPCs (p-value < 0.05) and HPCs that contain at least 10% of total patients are in bold, which is motivated by finding tissue patterns that can generalize across the cohort. This plot shows the relationship between HPCs and their relevance in predicting overall survival, negative values show association with survival while positive values are associated with a death event. Source data are provided as a Source Data file.



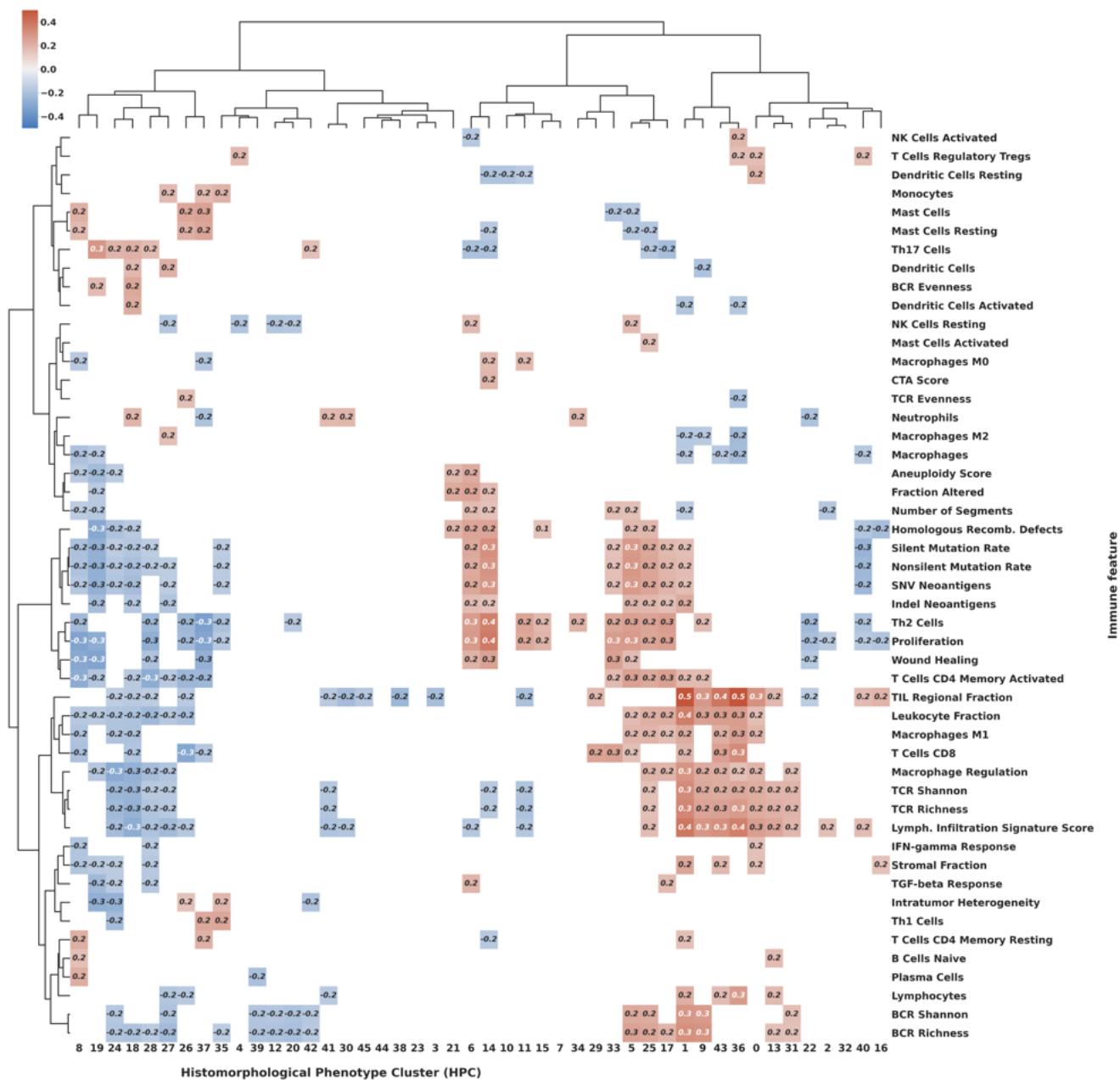
Supplementary Figure 9. The Cancer Genome Atlas (TCGA) and New York University NYU_1 tile samples for statistically significant HPCs on lung adenocarcinoma (LUAD) overall survival. Source data are provided as a Source Data file.



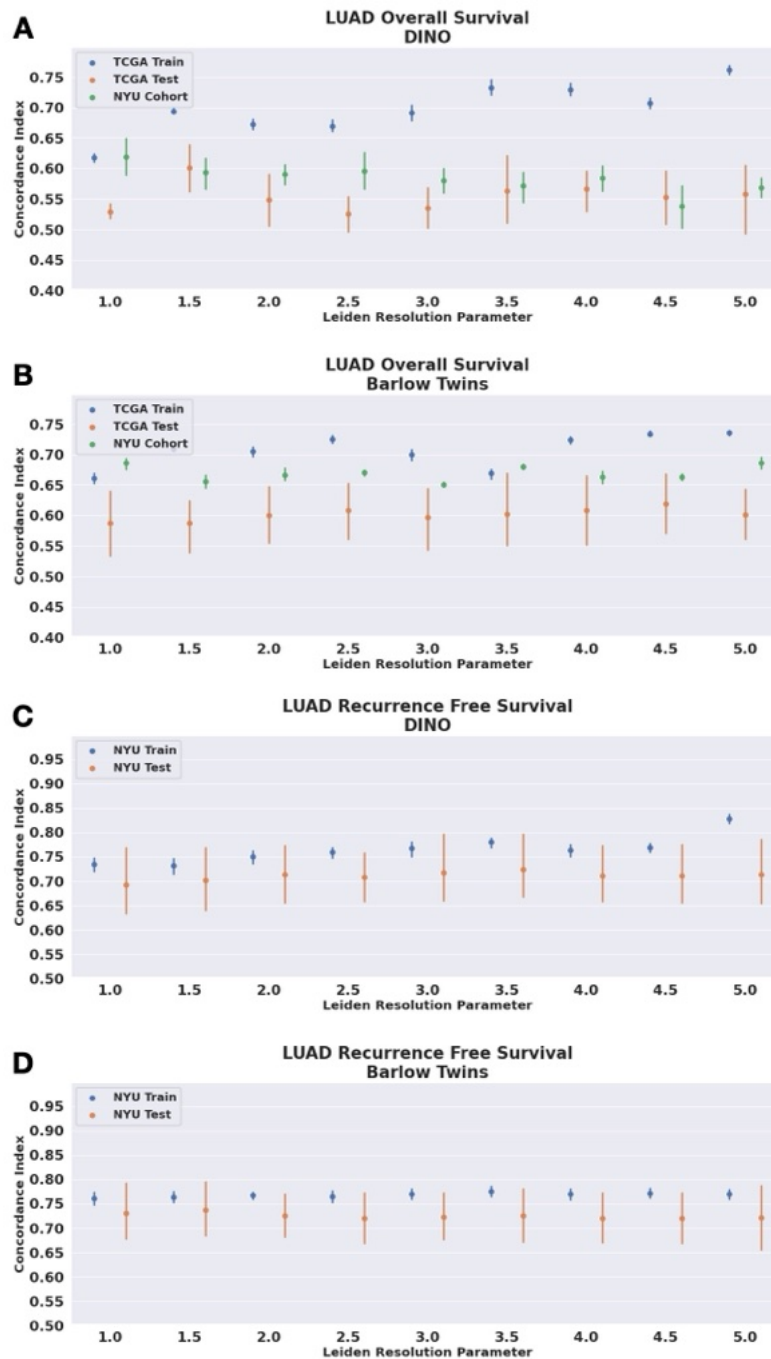
Supplementary Figure 10. Lung adenocarcinoma recurrence free survival Forest plot. Log hazard ratio of Cox proportional hazards model over the train sets of a 5-fold cross-validation on the NYU cohort. We averaged coefficients across fold and combined p-values with Fisher's combined probability test. This plot shows the relationship between HPCs and their positive or negative relevance in predicting recurrence. In order to find tissue patterns that can generalize across the cohort, we focused on HPCs that contain at least 10% of the total NYU number of patients and show statistical significance (p-value < 0.05); these are displayed in bold. Source data are provided as a Source Data file.



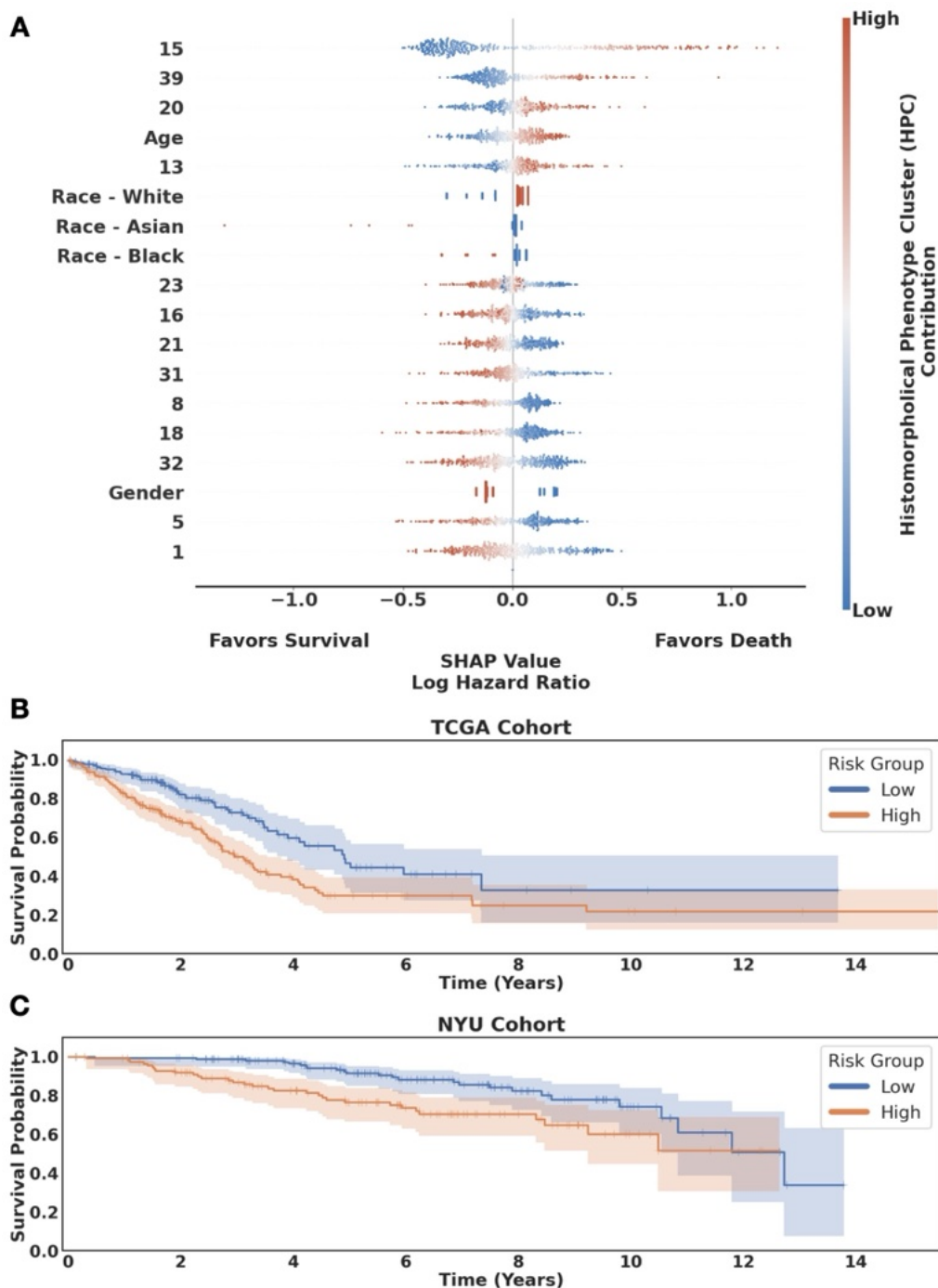
Supplementary Figure 11. The Cancer Genome Atlas (TCGA) and New York University *NYU*₁ tile samples for statistically significant HPCs on lung adenocarcinoma (LUAD) recurrence free survival. Source data are provided as a Source Data file.



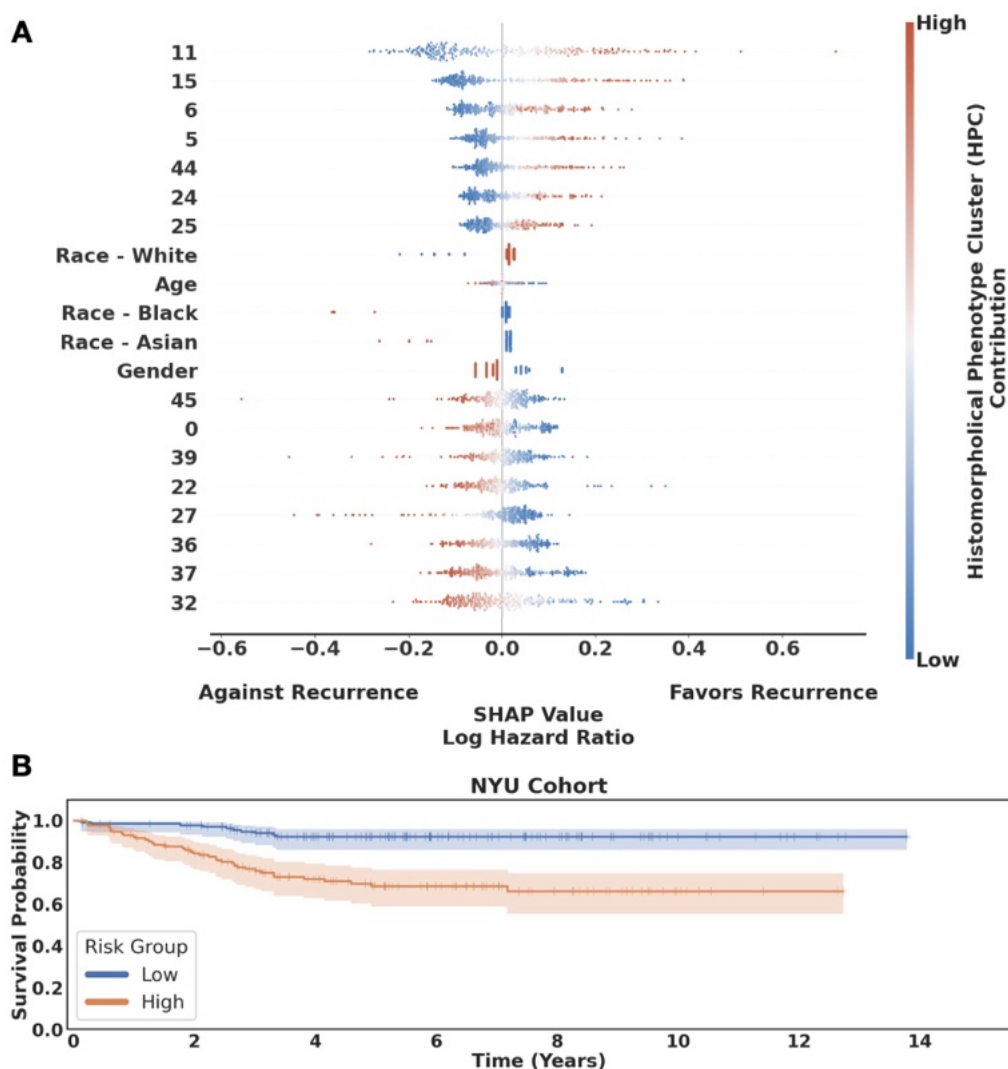
Supplementary Figure 12. Lung adenocarcinoma survival analysis and Histomorphological Phenotype Cluster (HPC) correlations. Bi-hierarchical clustering of HPCs and immune signature³ spearman correlations. Source data are provided as a Source Data file.



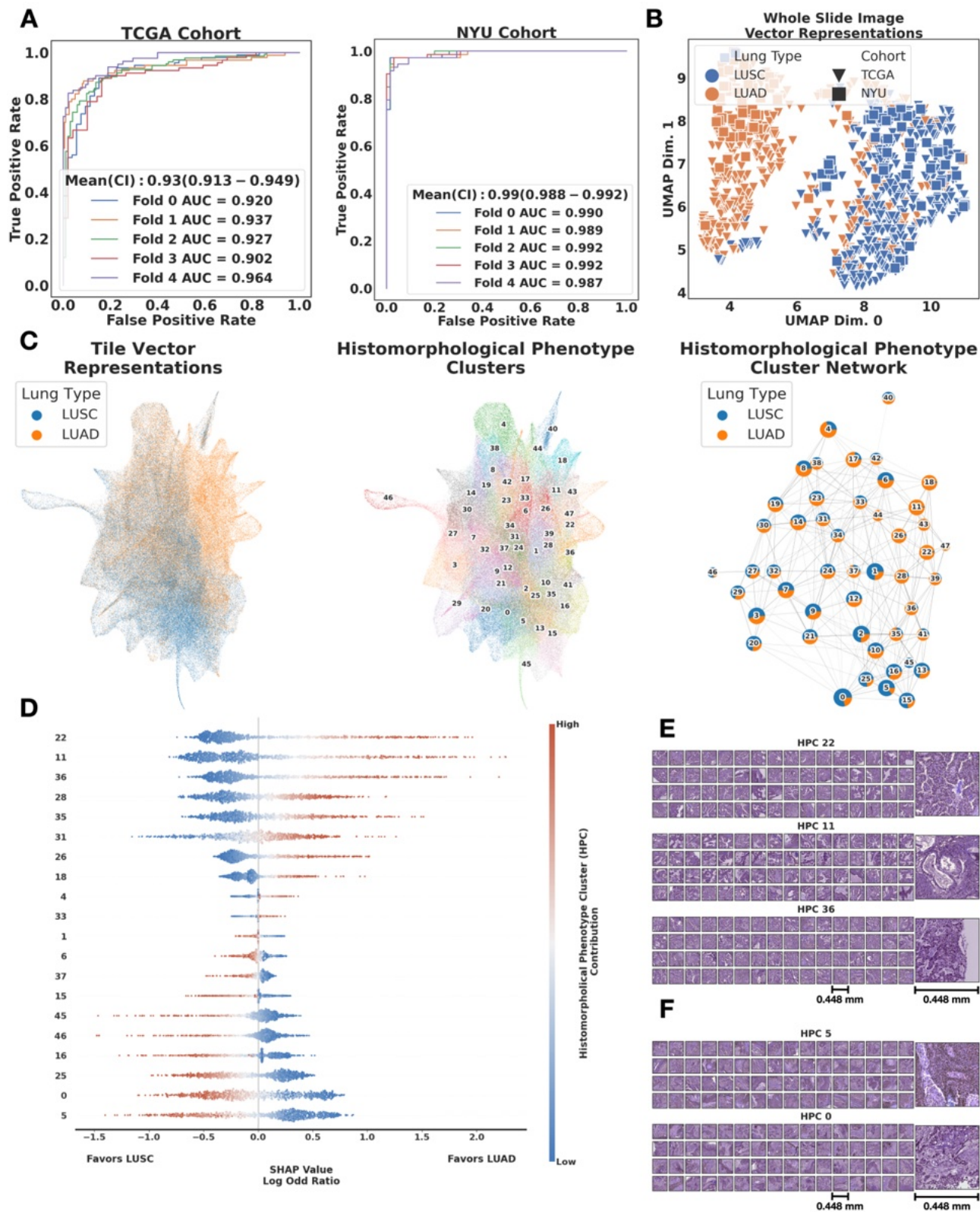
Supplementary Figure 13. LUAD survival analysis comparing Barlow Twins and DINO at different resolutions. **A** OS measured from HPL using DINO as self-supervised method. **B** OS measured from HPL using Barlow Twins as self-supervised method. **C** RFS measured from HPL using DINO as self-supervised method. **D** RFS measured from HPL using Barlow Twins as self-supervised method. DINO LUAD OS (Resolution 1.5) achieves a c-index of 0.612 with 95% confidence intervals (CI) of 0.569-0.643 in the TCGA Test set and a c-index of 0.597 with 95% CI of 0.565-0.625 in the NYU cohort. Barlow Twins (Resolution 2.0) achieves a TCGA Test set c-index of 0.600 with 95% CI of 0.551-0.649, and a NYU c-index of 0.666 with 95% CI of 0.656-0.676. DINO LUAD RFS (Resolution 1.5) achieves a comparable result to Barlow Twins with a c-index of 0.701 with 95% CI of 0.642-0.768. In comparison, Barlow Twins (Resolution 2.0) reaches a c-index of 0.725 with 95% CI of 0.681-0.770. The performance of both self-supervised methods is similar. However, DINO c-index fluctuates more with respect to the Leiden resolution parameter (resulting in more/less clusters) while Barlow Twins remains rather stable at different resolutions. All runs were done following a 5-fold cross-validation split similar to the one used in previous studies^{1,2}. Source data are provided as a Source Data file.



Supplementary Figure 14. LUAD Overall Survival analysis including patient metadata (age, sex, and race). Patient metadata is included as covariates in the Cox model along with HPC values. We evaluated the model using a 5-fold cross-validation over the TCGA data with a cohort from NYU as an additional independent set. **A.** SHAP (SHapley Additive exPlanations) plot showing how HPC (y-axis) weight on survival outcome (all race and ethnicities for TCGA data were submitted by the Tissue Source Sites that collected the sample and are presumed to be self-described/reported by the patient). **B.** High and low risk groups on TCGA cohort showing statistical significance ($p\text{-value } 1.86 \times 10^{-4} < 0.05$) and c-index of 0.636 with 95% confidence intervals (CI) of 0.596 – 0.680. **C.** High and low risk groups on NYU₁ cohort showing statistical significance ($p\text{-value } 3.57 \times 10^{-3} < 0.05$) and c-index of 0.650 with 95% CI of 0.636 – 0.664. Source data are provided as a Source Data file.

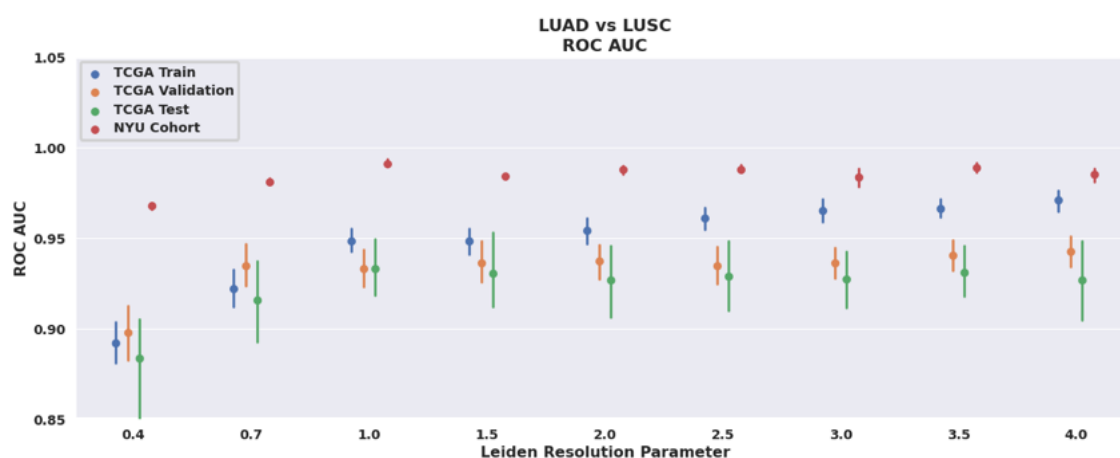


Supplementary Figure 15. LUAD Recurrence Free Survival analysis including patient metadata (age, sex, and race). Patient metadata is included as covariates in the Cox model along with HPC values. We evaluated the model using a 5-fold cross-validation over the NYU cohort (race was self-described by patients). **A.** SHAP (SHapley Additive exPlanations) plot showing how HPC (y-axis) weight on survival outcome. **B.** High and low risk groups on NYU cohort showing statistical significance ($p\text{-value } 8.82 \times 10^{-7} < 0.05$) and c-index of 0.748 with 95% confidence intervals (CI) of 0.698 – 0.808. Source data are provided as a Source Data file.

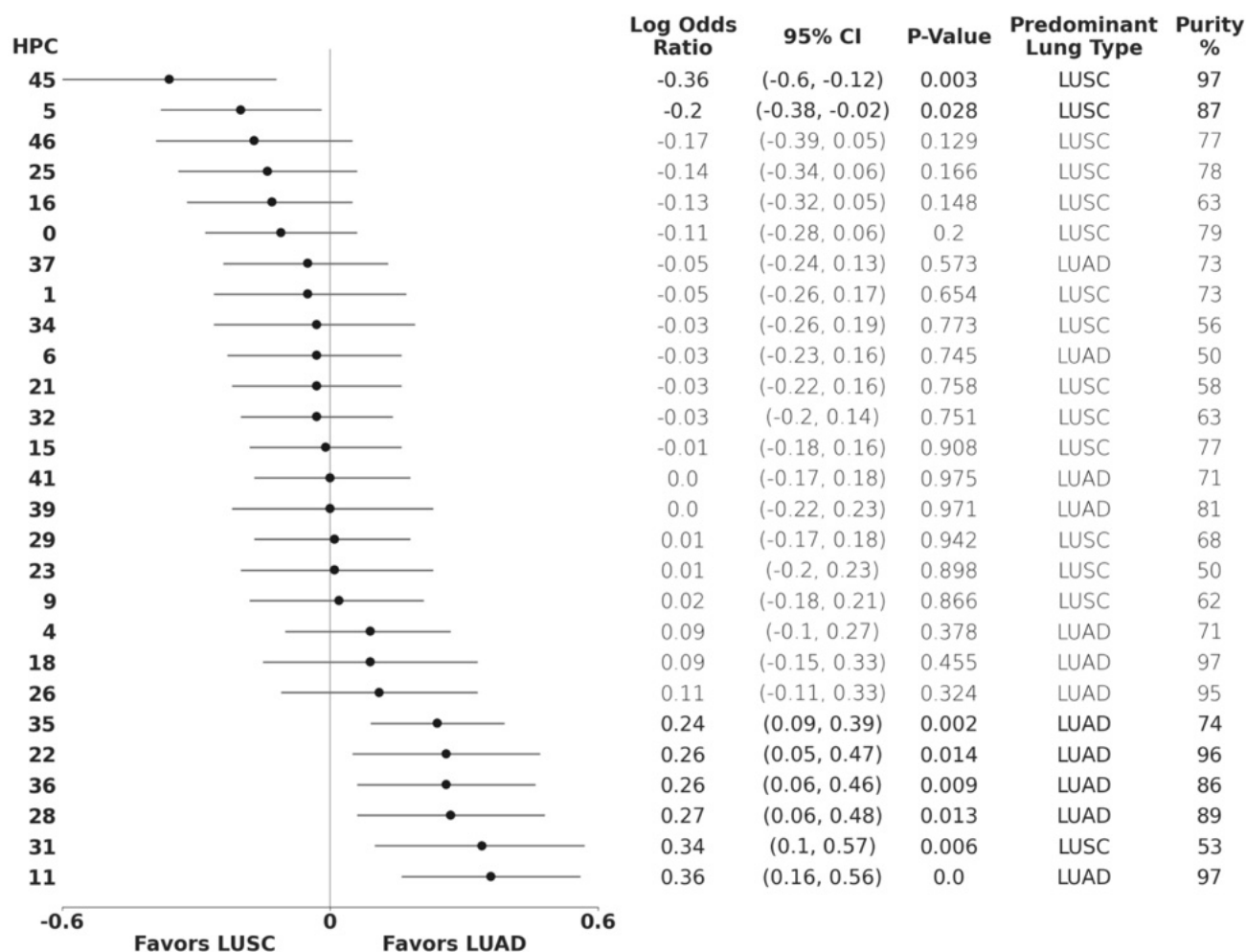


Supplementary Figure 16. Lung adenocarcinoma (LUAD) versus squamous cell carcinoma (LUSC)

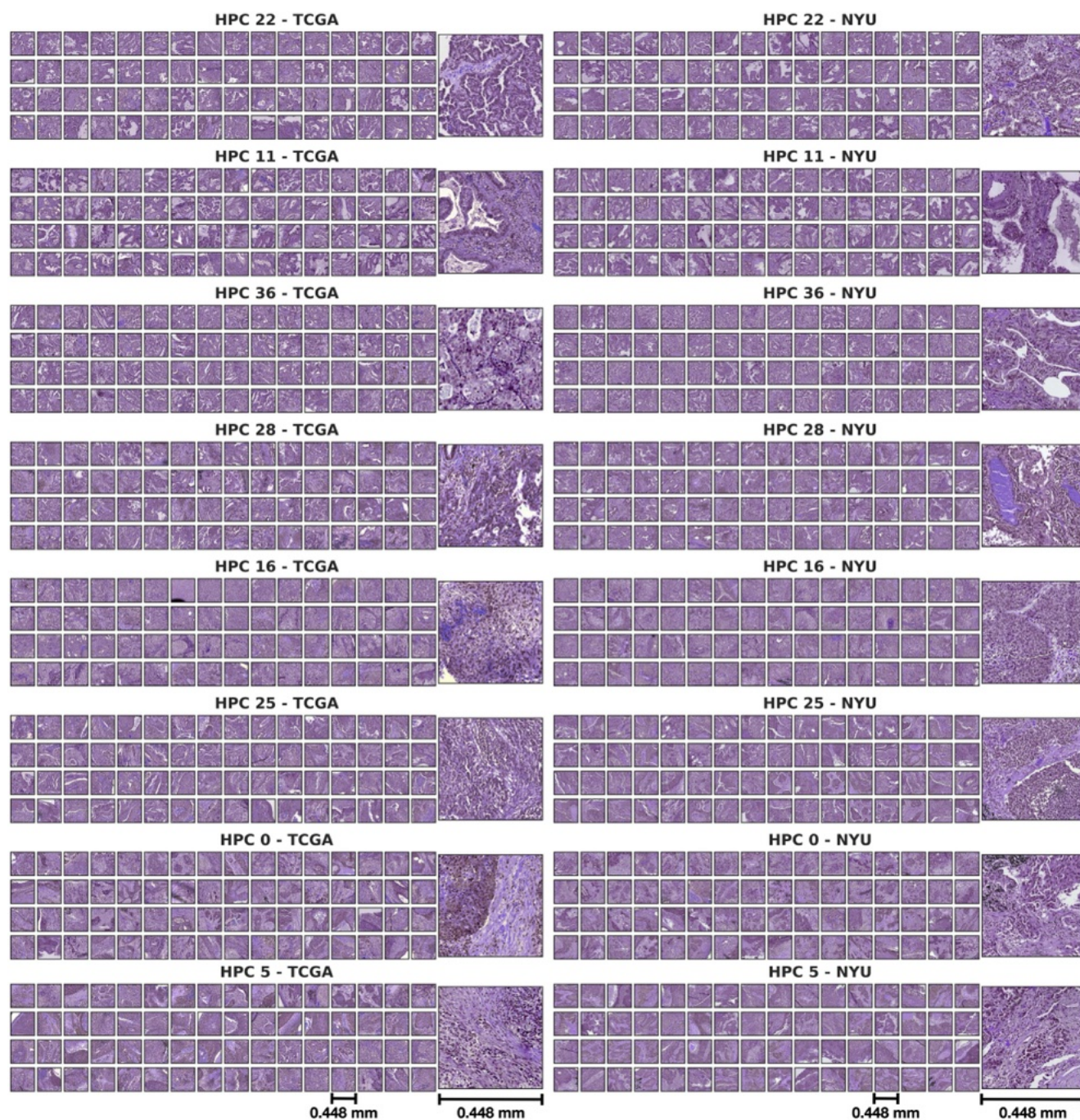
classification performance and representation analysis. **A.** Receiver operating characteristic (ROC) curve over a 5-fold cross-validation on TCGA dataset and NYU_2 dataset as an independent cohort, with mean AUCs of 0.93 and 0.99 respectively. **B.** Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction of whole slide image (WSI) compositional vector representations for TCGA and NYU_2 cohorts. Each representation is labeled with the corresponding lung type, LUSC (blue) and LUAD (orange). **C.** UMAP dimensionality reduction of tile vector representations, each tile label corresponds to the associated WSI lung type (left) and HPC membership (middle), and partition-based graph abstraction (PAGA) of HPC and their connections (right). **D.** SHAP (SHapley Additive exPlanations) plot. **E.** Examples of tiles randomly selected from the 3 top HPCs enriched in LUAD-specific phenotypes according to SHAP analysis from panel D. **F.** Examples of tiles randomly selected from the 3 top HPCs enriched in LUSC-specific phenotypes according to SHAP analysis from panel D. Source data are provided as a Source Data file.



Supplementary Figure 17. Effect of the Leiden resolution on the LUAD/LUSC classification task. This run was done on a 5 fold cross-validation. Source data are provided as a Source Data file.

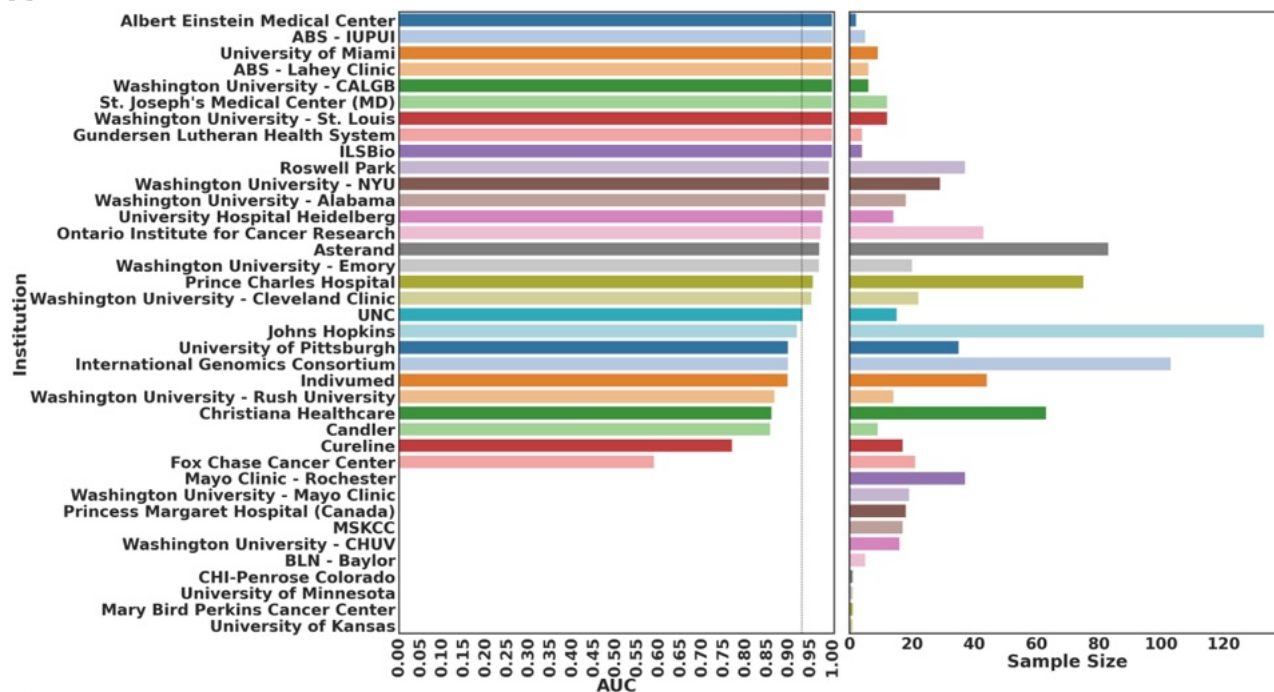


Supplementary Figure 18. Lung adenocarcinoma (LUAD) versus squamous cell carcinoma (LUSC) Forest plot. Forest plot of the logistic regression's coefficients over the train sets of a 5-fold cross-validation. We averaged coefficients across fold and combined p-values with Fisher's combined probability test. We highlighted statistically significant HPCs (p-value < 0.05) and included each HPC's dominant type and purity. We calculated purity as the percentage of the dominant cancer type tiles with respect to the total HPC tiles. Source data are provided as a Source Data file.

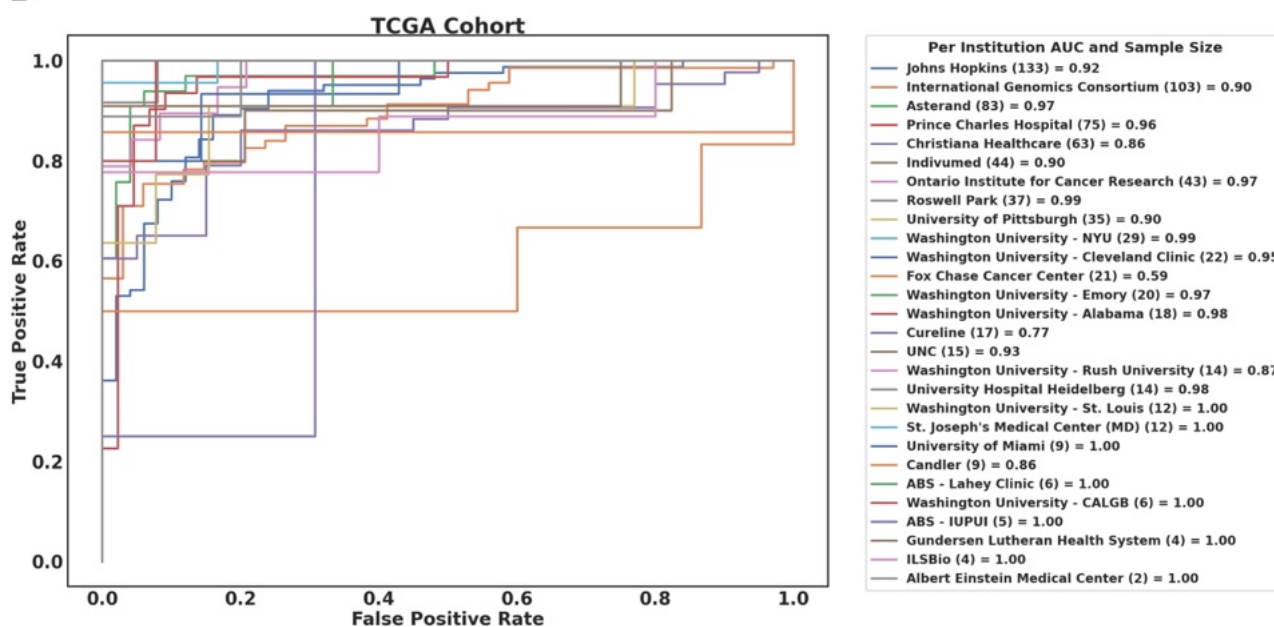


Supplementary Figure 19. The Cancer Genome Atlas (TCGA) and New York University NYU_2 tile samples for statistically significant HPCs on lung adenocarcinoma (LUAD) versus cell squamous cell carcinoma (LUSC) classification. Source data are provided as a Source Data file.

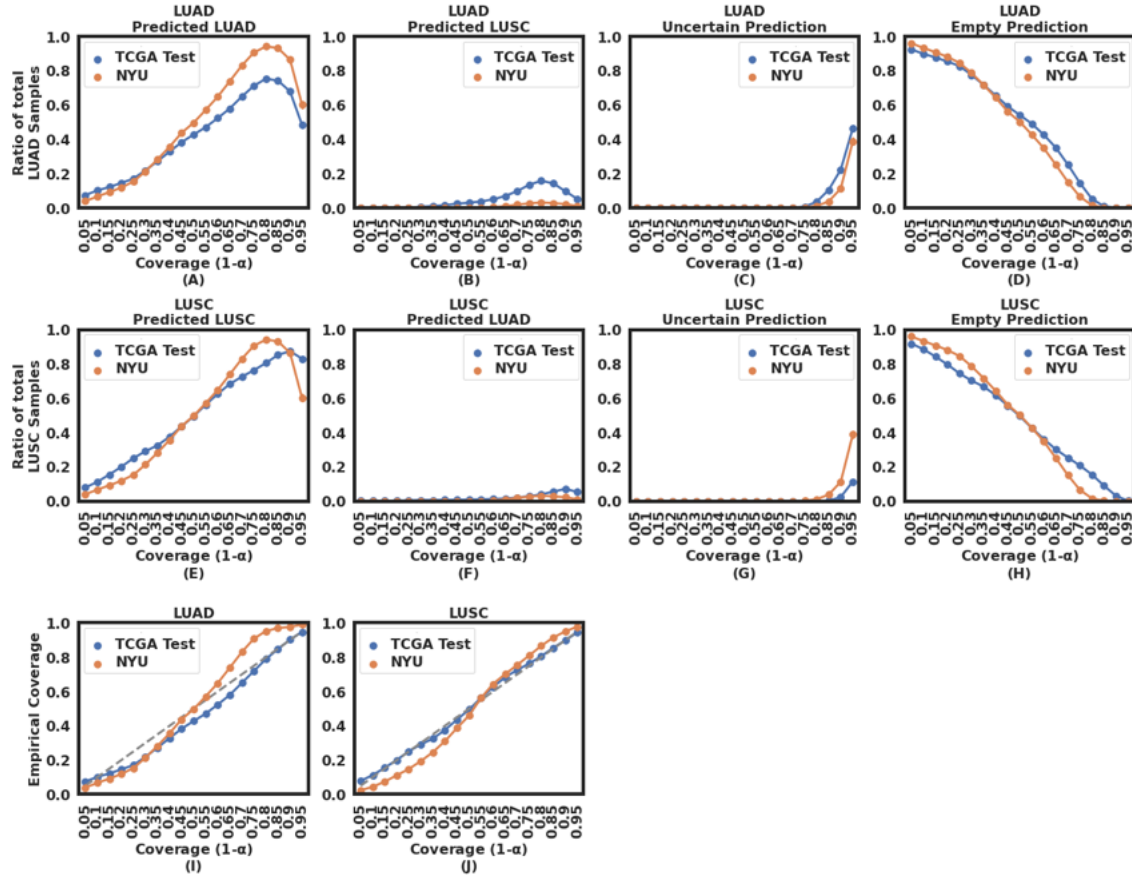
A



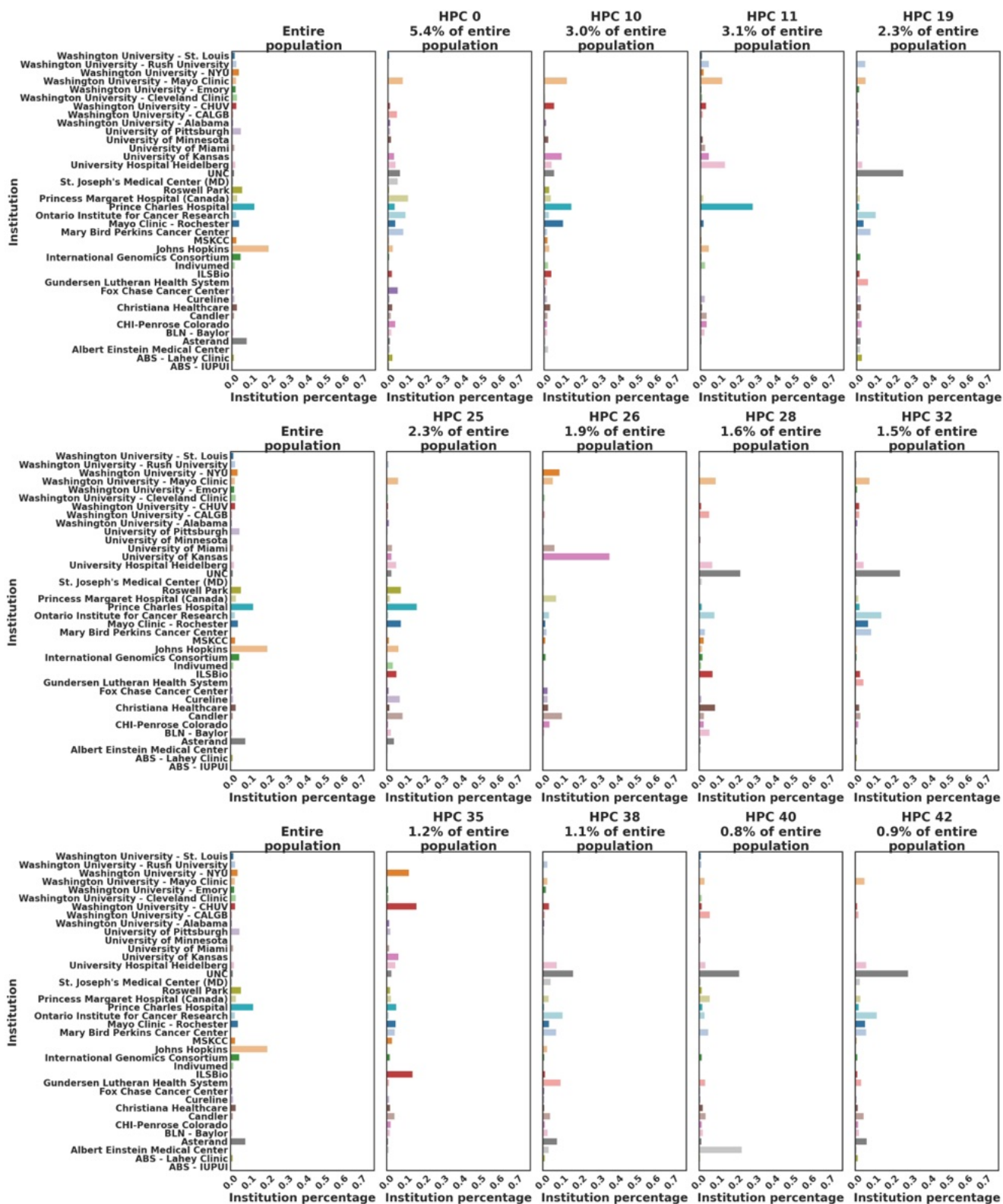
B



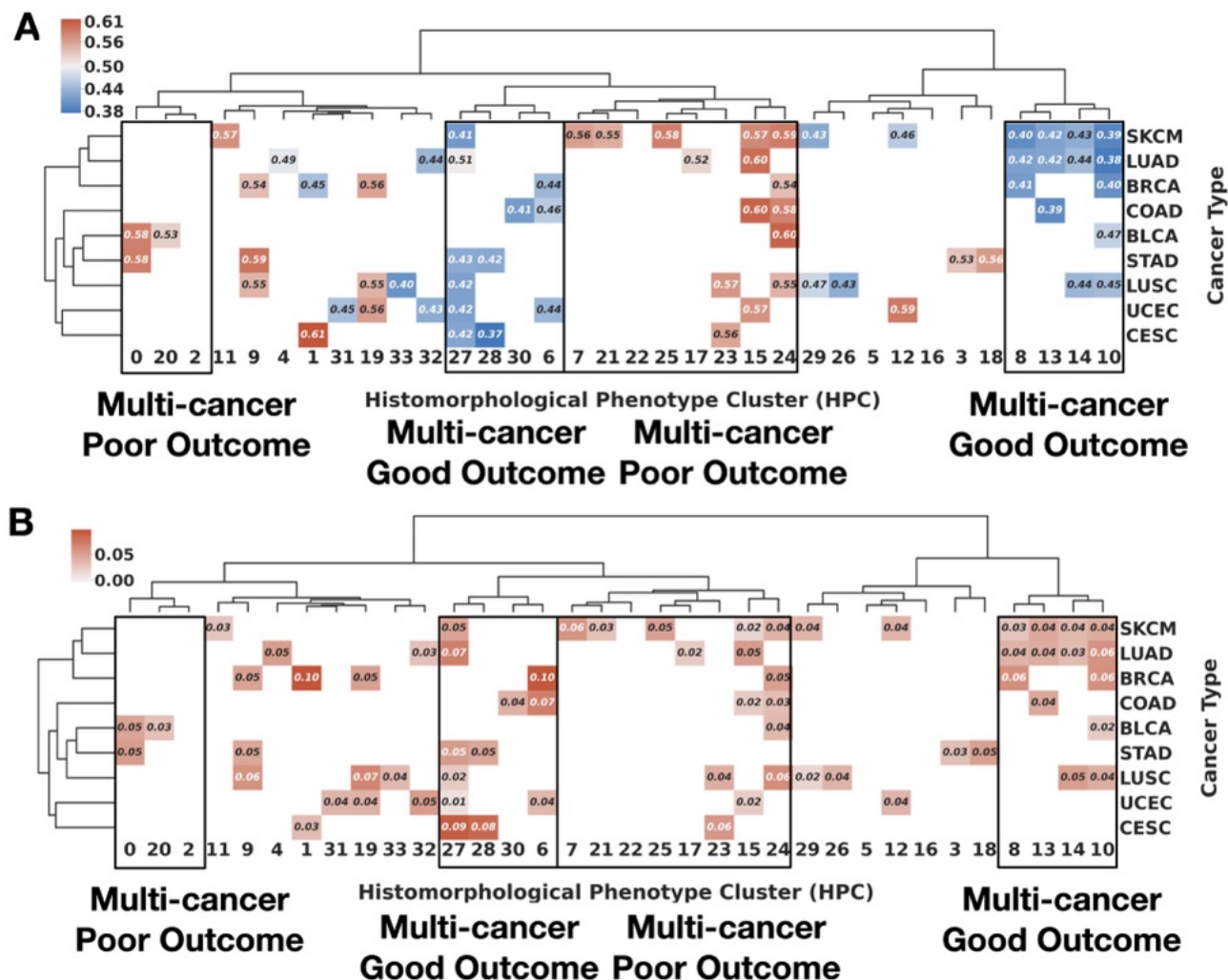
Supplementary Figure 20. Lung adenocarcinoma (LUAD) versus squamous cell carcinoma (LUSC) classification performance per institution. A Receiver operating characteristic (ROC) area under the curve (AUC) per TCGA institution and WSI sample size. Institutions without an AUC value only have one kind of lung type, not allowing to compute the ROC AUC. **B** ROC curve per TCGA institution. These results correspond to the 5-fold cross-validation on TCGA dataset and NYU dataset as independent cohort as included in **Figure 13 A**. Source data are provided as a Source Data file.



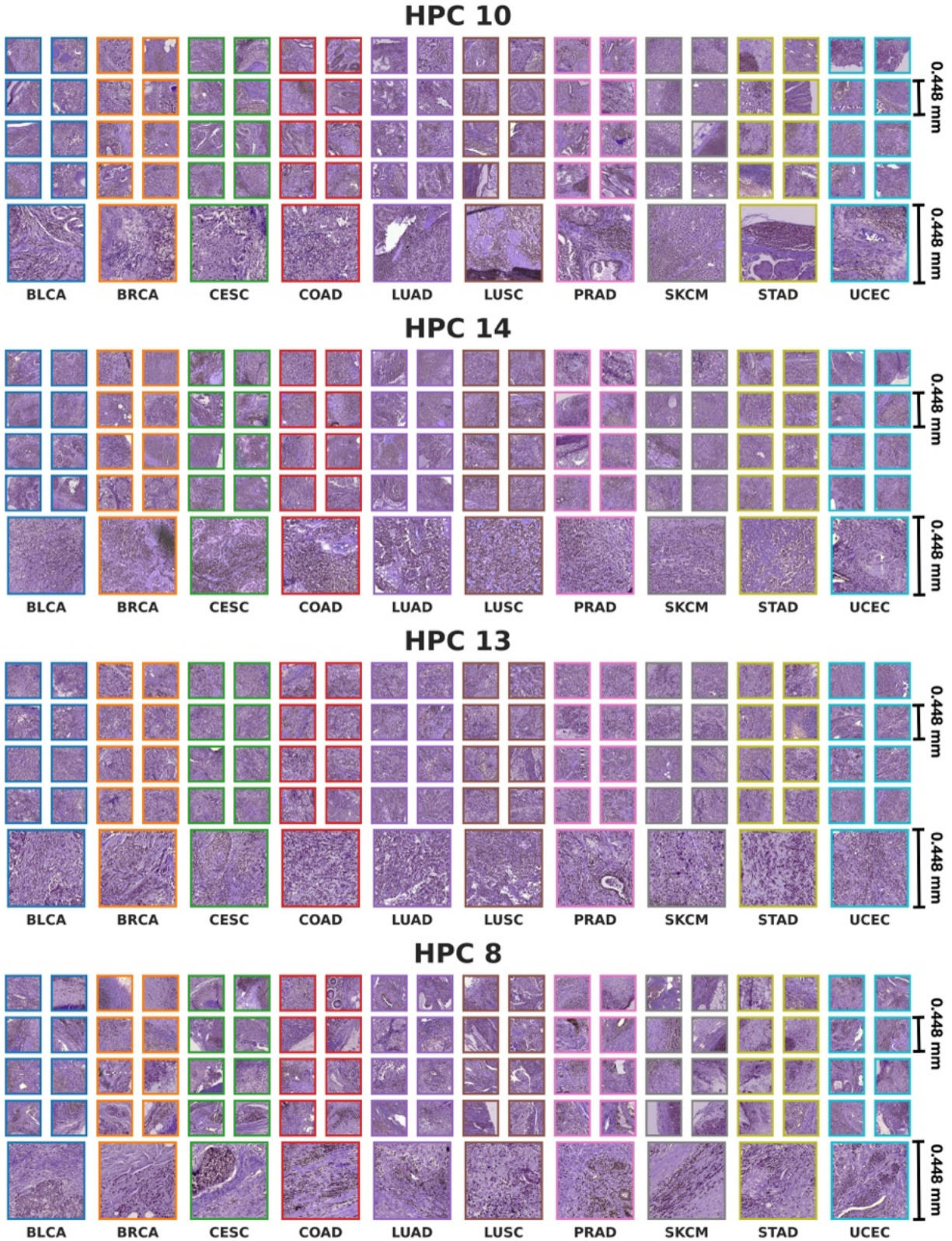
Supplementary Figure 21. Label-Conditional Conformal Prediction on the LUAD vs LUSC prediction task. We performed MICP⁴ for a range of coverage $(1 - \alpha)$ from 0.05 to 0.95. The figure above shows for LUAD and LUSC the ratios of correct single predictions (A,F), incorrect single predictions (B,G), uncertain predictions (C,H), and empty predictions (D,I). Uncertain predictions are defined as prediction sets that contain both labels and empty predictions are as prediction sets that contain no label. Finally, we include the empirical coverage for LUAD (I) and LUSC (J), highlighting the calibration coverage in the gray dashed line. We define non-conformity measure as one minus the probability of the correct class. For each coverage measure $(1 - \alpha)$, we obtained 1000 samples by running 200 trials per fold, out of the initial folds used in the 5-fold cross-validation LUAD vs LUSC task. For each trial, we randomly subsampled 50% of the total samples on the TCGA validation set (calibration set), TCGA test set (quantification set), and NYU cohort (additional quantification set), resulting into approximately 100 samples for the TCGA validation and test set, and 70 samples for the NYU cohort. We can see higher uncertainty in TCGA LUAD samples compared to TCGA LUSC samples. The model is also more sensitive to misclassifications in LUAD, barely making any in LUSC. Indeed, this result may be explained by LUAD solid samples, which are harder to differentiate from LUSC than other LUAD subtypes such as acinar, papillary, micropapillary, or lepidic. HPL displays similar confidence on LUAD and LUSC samples of the NYU cohort, with higher uncertainty in LUSC compared to TCGA. Nonetheless, HPL makes larger correct single predictions on both subtypes in the NYU cohort than in TCGA. Possibly due to NYU WSIs being cleaner and less noisy samples than the TCGA counterparts, in particular on the LUAD cases. Source data are provided as a Source Data file.



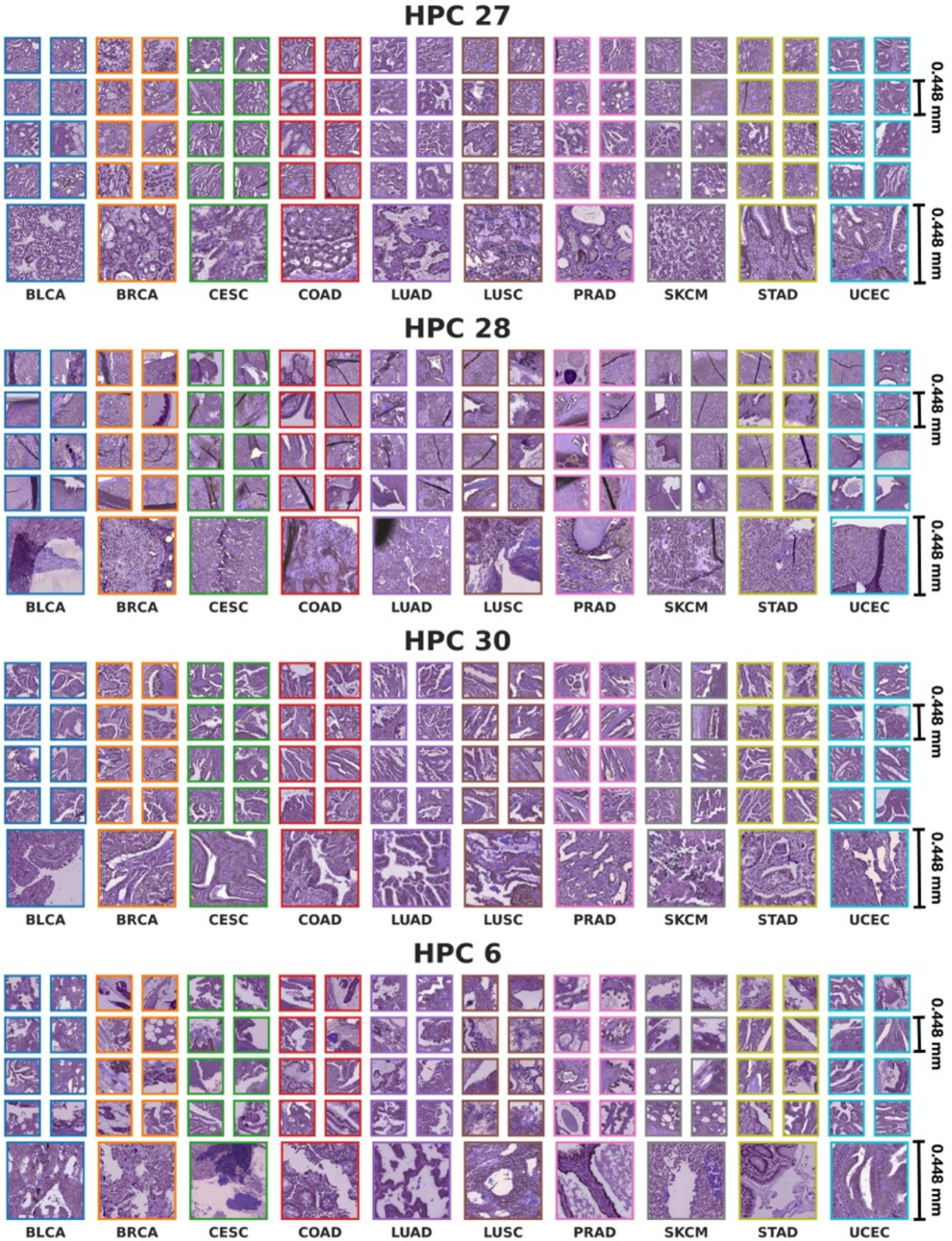
Supplementary Figure 22. Lung adenocarcinoma (LUAD) versus squamous cell carcinoma (LUSC) HPCs: institution distribution per HPC and for the entire population of tissue tiles. we show 12 randomly sampled HPCs. Source data are provided as a Source Data file.



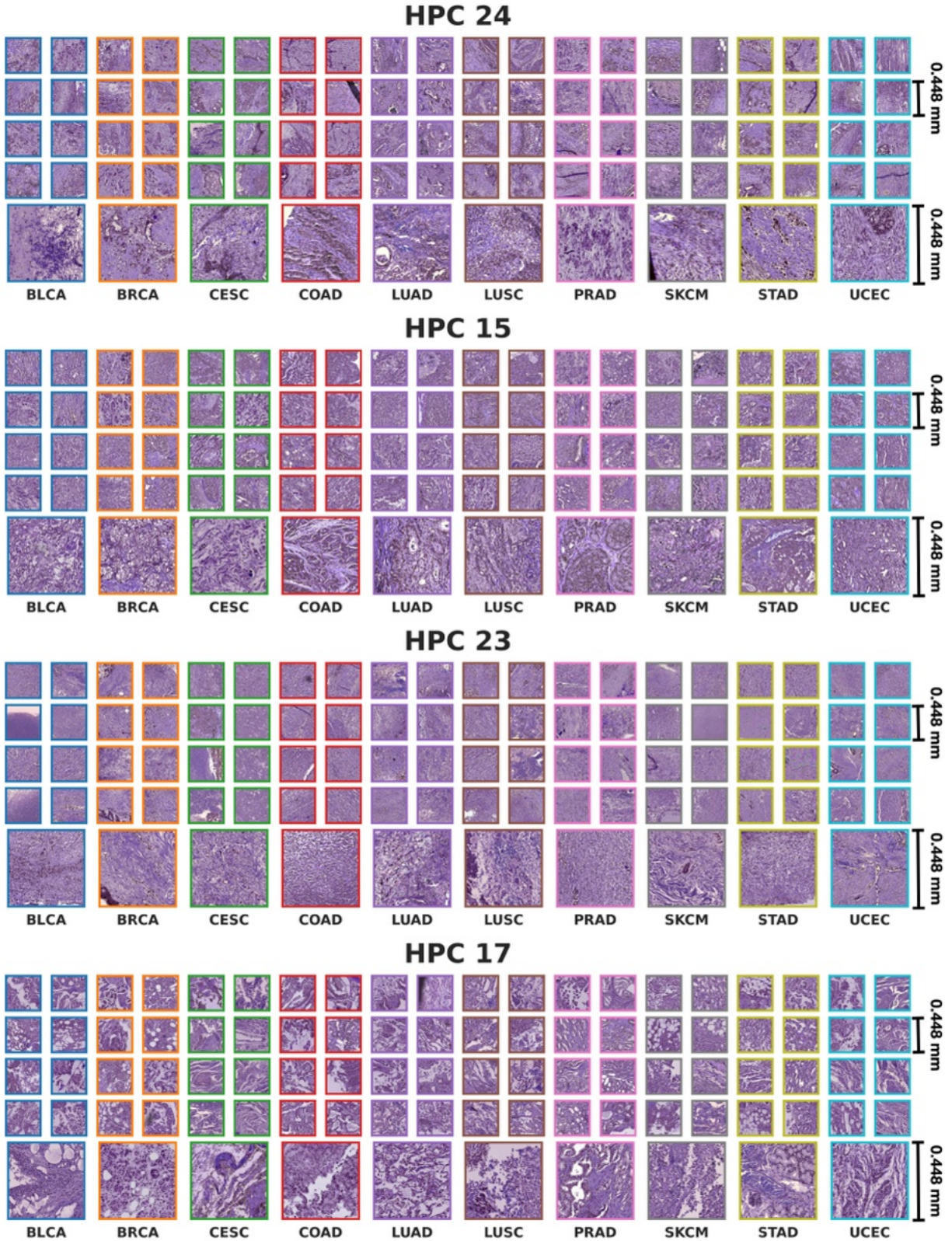
Supplementary Figure 23. Concordance index between multi-cancer HPCs and overall survival. **A** Mean C-index values across the 5-fold cross validation. Values below 0.5 (blue) indicate that higher percentage of the HPC favors longer survival (good outcome), while those above 0.5 (red) indicate that higher percentage the HPC favors shorter survival (poor outcome). **B** Variation for the 95% confidence interval over the 5-fold cross validation. We only display statistically significant values of the log rank test for the high and low risk patient split (p -value < 0.05). Risk groups are defined by taking the median risk in the training set (in this casem the HPC value). Source data are provided as a Source Data file.



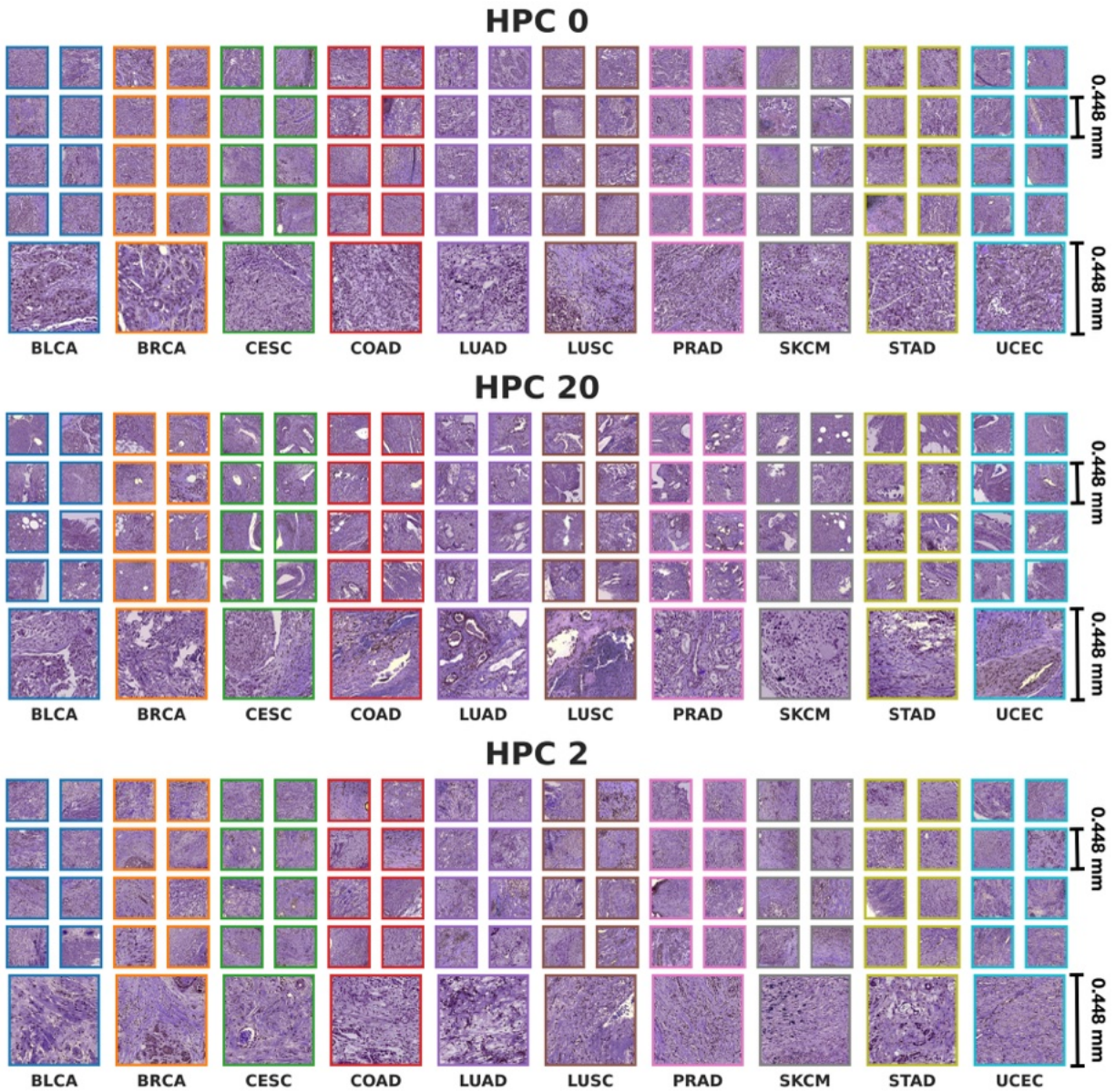
Supplementary Figure 24. Examples of tiles from The Cancer Genome Atlas (TCGA) assigned to HPCs associated with good outcomes, and corresponding to the right-hand side of the bi-herarchical clustering in Figure 9A. Source data are provided as a Source Data file.



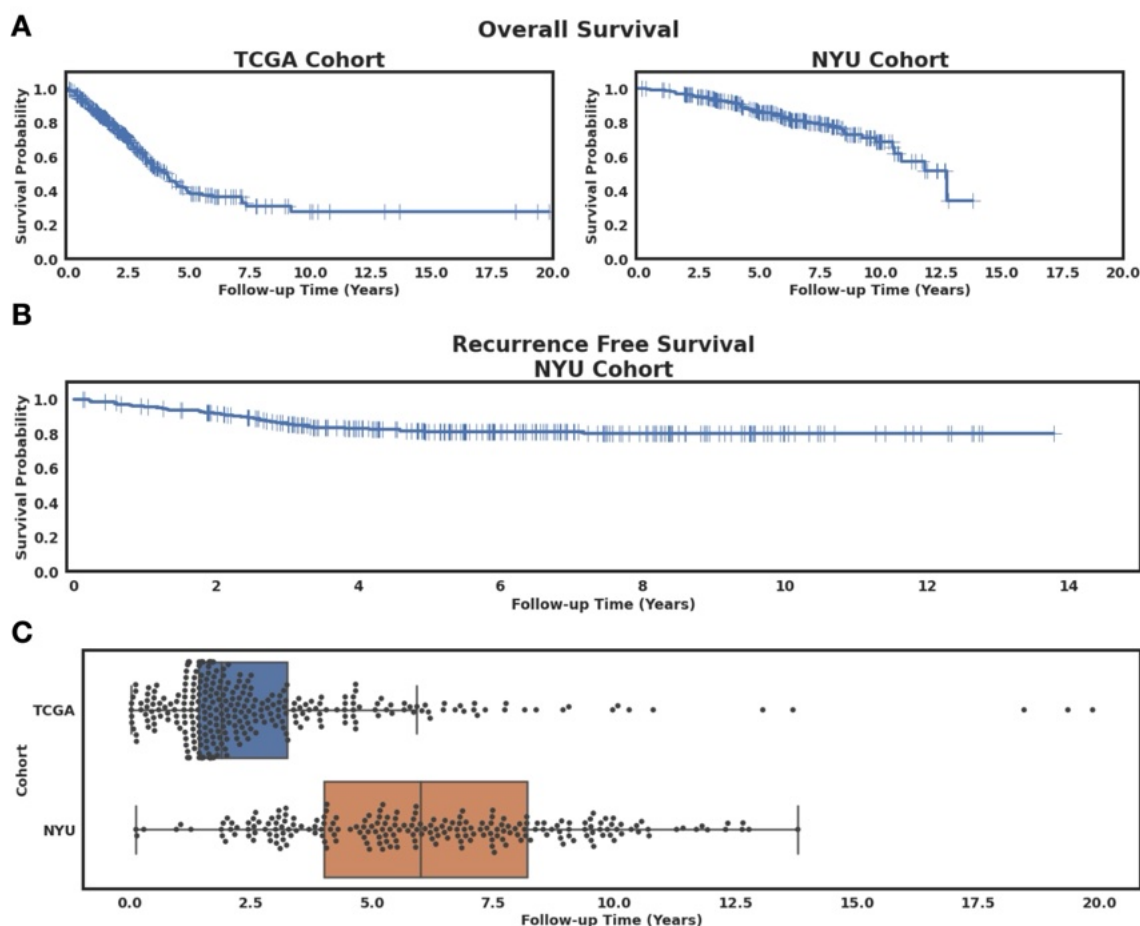
Supplementary Figure 25. Examples of tiles from The Cancer Genome Atlas (TCGA) assigned to HPCs associated with good outcomes, and corresponding to the left-hand side of the bi-herarchical clustering in Figure 9A. Source data are provided as a Source Data file.



Supplementary Figure 26. Examples of tiles from The Cancer Genome Atlas (TCGA) assigned to HPCs associated with poor outcomes, and corresponding to the right-hand side of the bi-herarchical clustering in Figure 9A. Source data are provided as a Source Data file.



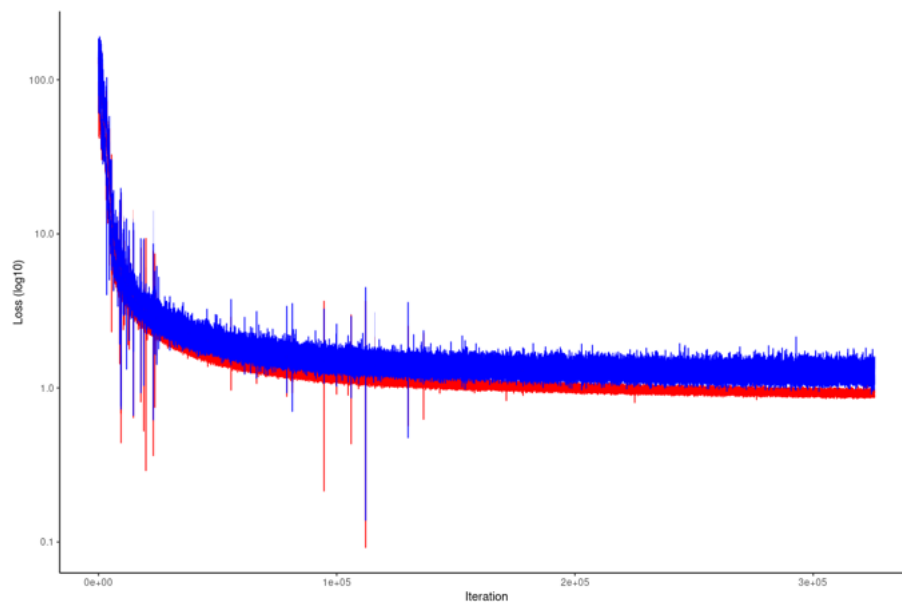
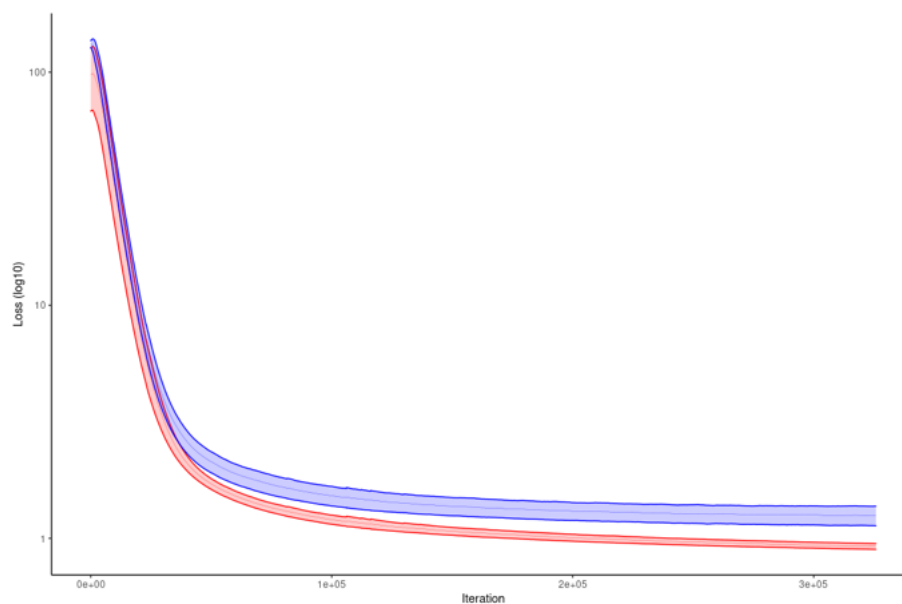
Supplementary Figure 27. Examples of tiles from The Cancer Genome Atlas (TCGA) assigned to HPCs associated with poor outcomes, and corresponding to the left-hand side of the bi-hierarchical clustering in Figure 9A. Source data are provided as a Source Data file.



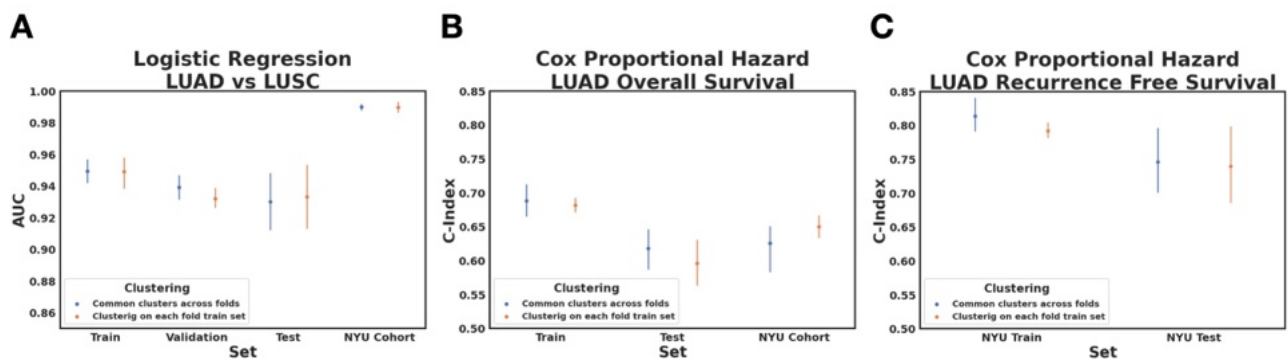
Supplementary Figure 28. Dataset description for lung adenocarcinoma overall and recurrence free survival. **A** Kaplan-Meier curves of overall survival events of lung adenocarcinoma (LUAD) of The Cancer Genome Atlas (TCGA) and New York University *NYU*₁ cohorts. Censored samples are represented by a cross. **B** Kaplan-Meier curves of recurrence free survival events of LUAD systemic and loco-regional recurrence for the *NYU*₁ cohort. The NYU cohort is composed by three types: systemic, loco-regional, and new primary. We censored new primary cases as we focus on systemic and loco-regional. Censored samples are represented by a cross. **C** Follow-up times from patients of TCGA and *NYU*₁ cohorts. NYU patients are considerably larger follow-up times allowing a more refined study of recurrence. Source data are provided as a Source Data file.

Encoder Network $E : x \rightarrow z$
$x \in \mathbb{R}^{224 \times 224 \times 3}$
Conv2D Layer, 7x7, stride 2, pad downscale, Batch Norm, and ReLU $224 \times 224 \times 3 \rightarrow 112 \times 112 \times 32$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $112 \times 112 \times 32 \rightarrow 112 \times 112 \times 32$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $112 \times 112 \times 32 \rightarrow 112 \times 112 \times 32$
Conv2D Layer, 4x4, stride 2, pad downscale, Batch Norm, and ReLU $112 \times 112 \times 32 \rightarrow 56 \times 56 \times 32$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $56 \times 56 \times 32 \rightarrow 56 \times 56 \times 32$
Attention Layer at $56 \times 56 \times 32$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $56 \times 56 \times 32 \rightarrow 28 \times 28 \times 64$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $28 \times 28 \times 64 \rightarrow 28 \times 28 \times 64$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $28 \times 28 \times 64 \rightarrow 28 \times 28 \times 64$
Conv2D Layer, 4x4, stride 2, pad downscale, Batch Norm, and ReLU $28 \times 28 \times 64 \rightarrow 14 \times 14 \times 128$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $14 \times 14 \times 128 \rightarrow 14 \times 14 \times 128$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $14 \times 14 \times 128 \rightarrow 14 \times 14 \times 128$
Conv2D Layer, 4x4, stride 2, pad downscale, Batch Norm, and ReLU $14 \times 14 \times 128 \rightarrow 7 \times 7 \times 256$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $7 \times 7 \times 256 \rightarrow 7 \times 7 \times 256$
ResNet Conv2D Layer, 5x5, stride 1, pad same, Batch Norm, and ReLU $7 \times 7 \times 256 \rightarrow 7 \times 7 \times 256$
Conv2D Layer, 4x4, stride 2, pad downscale, Batch Norm, and ReLU $7 \times 7 \times 256 \rightarrow 4 \times 4 \times 512$
Flatten $4 \times 4 \times 512 \rightarrow 1024$
Dense Layer, Batch Norm and ReLU, $1024 \rightarrow 1024$
Dense Layer, Batch Norm and ReLU, $1024 \rightarrow 1024$
Dense Layer, Batch Norm and ReLU, $1024 \rightarrow 128$

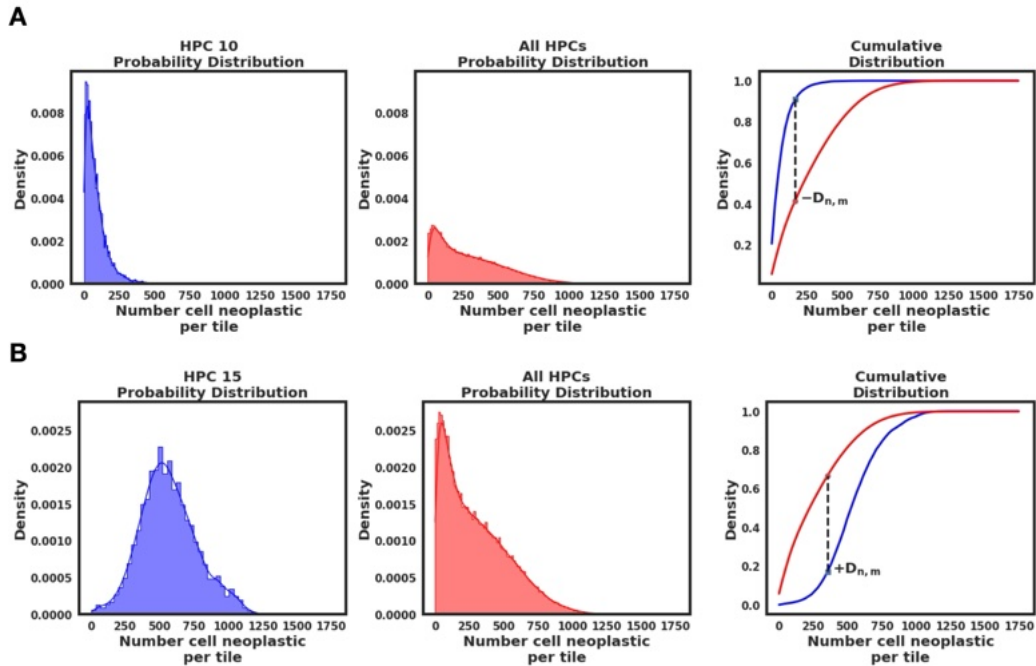
Supplementary Figure 29. Detailed description of the CNN layers. Source data are provided as a Source Data file.

A**B**

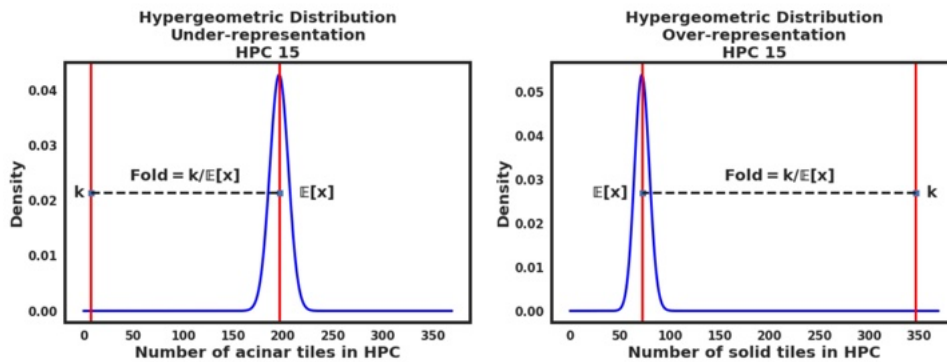
Supplementary Figure 30. Loss during training of Barlow-Twins-based network. **A** Raw losses for the train (red) and validation (blue) sets. **B** Same as panel A but after smoothing the curve using exponential moving average of the batch losses over 1 epoch for clearer display. Error shows the standard deviation from the 5-fold cross-validation runs. Source data are provided as a Source Data file.



Supplementary Figure 31. Clustering impact on performance. Performance variation when clustering is done on each fold's train set compared to when a particular cluster configuration is hold the same across folds for the classification or survival task. **A** Performance variation on logistic regression for LUSC versus LUAD classification. **B** Performance variation on Cox proportional hazard for LUAD overall survival. **C** Performance variation on Cox proportional hazard for LUAD recurrence free survival. Source data are provided as a Source Data file.



Supplementary Figure 32. Kolmogorov-Smirnov test for lung adenocarcinoma cell type over and under-representation. Representations of usage of Two-sample Kolmogorov-Smirnov (K-S) test to account for over and under-representation of cell types in an HPC. We used the K-S statistic $D_{n,m}$ to quantify for over-representation and under-representations, assigning $+D_{n,m}$ if there is over-representation (A) and $-D_{n,m}$ if there is under-representation (B). The Two-sample Kolmogorov-Smirnov test uses the statistic $D_{n,m}$ to quantify the distance between empirical distributions. The statistic is defined as $D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|$ where $F_{1,n}(x)$, $F_{2,m}(x)$ are the empirical distribution functions of the first and second samples, n is the sample size for the F_1 , m is the sample size for the F_2 , and x the support for both distributions. Source data are provided as a Source Data file.



Supplementary Figure 33. Hypergeometric test for lung adenocarcinoma histological subtype enrichment. The hypergeometric test uses the hypergeometric distribution in order to measure the statistical significance of seeing k successes in n draws from a population of size N with K successes, without replacement. In our case, we specify k as the number of tiles with a particular histological subtype given as the number of total tiles in the HPC n , K is the number of tiles with the same particular histological subtype in the entire population of annotated tiles of size N . We measure enrichment by using a *fold* metric defined as $k/\mathbb{E}[x]$ where $\mathbb{E}[x]$ is the expectation of the hypergeometric distribution given n, K, N . Enrichment for a particular histological subtype shows as *fold* > 1 and depletion as *fold* < 1 . This figure shows an example of the under-representation of acinar growth pattern and over-representation of the solid growth pattern in HPC 15. Source data are provided as a Source Data file.

Supplementary References

1. Chen, R. J. *et al.* Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* **40**, 865–878 (2022).
2. Chen, R. J. *et al.* Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16144–16155 (2022).
3. Thorsson, V. *et al.* The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).
4. Toccaceli, P. & Gammerman, A. Combination of inductive mondrian conformal predictors. *Mach. Learn.* **108**, 489–510 (2019).