HISTOMORPHOLOGICAL PHENOTYPE LEARNING (HPL)
MORPHOLOGICAL CLUSTERS LINKED TO PATIENT OUTCOME AND MOLECULAR PHENOTYPES

Adalberto Claudio Quiros, John Le Quesne, Ke Yuan
Nicolas Coudray, Anna Yeaton, Harvey Pass,
Andre Moreira, Luis Chiriboga, Navneet Narula, Afreen Karimkhan, Aristotelis Tsirigos

University of Glasgow, School of Computing Science/Beatson Institute
New York University, School of Medicine

July 22nd, 2022

OUTLINE

1. Methodology:
    A. Histomorphological phenotype clusters (HPC).
    B. HPC correlations and enrichment.
2. Results:
    C. LUAD vs LUSC classification.
    D. LUAD overall survival.
    E. LUAD recurrence free survival.
    F. HPC correlations:
        1. Immune landscape
        2. Tissue tile cell annotations
        3. Histological subtype tile annotations.
    G. LUAD recurrence free survival: Framework vs other annotations.
    H. LUAD Overall survival.

METHODOLOGY

METHODOLOGY - HPL

Methodology

A. Whole Slide Image (WSI) tiling and stain color normalization[1].

B. Self-supervised learning - Barlow Twins[2]:

- Training of CNN backbone $f_\theta(\,.\,)$ in order to create z tissue tile representations.

C. Tile vector representation clustering:

- Nearest neighbor graph definition from z tissue representations and Leiden community detection.

D. WSI vector representations:

- WSIs are defined by a vector with dimensionality equal to the total number of clusters.
- Each dimension corresponds to the percentage of the phenotype of the total tissue tiles.

[1] Reinhard, M. et al. Color transfer between images, 2001.
[2] Zbontar, L. et al. Barlow twins: Self-supervised learning via redundancy reduction, 2021.

METHODOLOGY - IMMUNE FEAT. & HISTO. ANNO.

Immune landscape features:

- Immune features such as TIL, leukocyte fraction, wound healing, proliferation, and others derived from RNA sequence data[3]
- Spearman correlation to account for relationships between clusters and immune features.
- Threshold: $pvalue < 0.01$

Histological subtype annotations:

- NYU manual tile annotations on tissue tiles.
- Measuring enrichment on histological subtype per cluster:
  1. Hypergeometric test:
     - Probability of k successes in n draws from a population N with K objects (No replacement).
  2. Use of fold to account for over/under rep:
     - $fold = k/expectation = kn(K/N)$
  3. Threshold: $pvalue < 0.01$
     - Over representation: $pvalue = 1 - CDF(k, n, K, N)$
     - Under representation: $pvalue = CDF(k, n, K, N)$

[3] Thorsson, V. et al. The Immune Landscape of Cancer, 2018.

METHODOLOGY - CELL ANNOTATIONS

Cell Annotations:

- Use of HoverNet[4] to annotate cells in each of the tissue tiles: neoplastic, connective, inflammatory, and dead cells.
- E.g.: Number of neoplastic cells per tile for the entire population vs cluster 15

Measuring enrichment on cell types per cluster:

1. Two Sample Kolmogorov–Smirnov test:
   - Critical value: largest difference between CDFs at any point of the support.

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|$$

2. Use critical value to measure the distributions shift:
   - Over representation $= +D_{n,m}$
   - Under representation $= -D_{n,m}$
3. Threshold: $pvalue < 0.01$

[4] Graham, S. et al. HoVer-Net: Simultaneous Segmentation and Classification of Nuclei in Multi-Tissue Histology Images, 2019.

RESULTS
LUAD VS LUSC CLASSIFICATION

RESULTS - LUAD VS LUSC CLASSIFICATION

LUAD vs LUSC classification:

- Logistic regression over WSI representations.
- 5-fold cross-validation with non-overlapping institutions across folds:
  - TCGA train, validation, and test sets.
  - NYU as an additional test set.
- AUC performance above 0.93.

AUC Mean and 95% CI for increasing number of clusters.

RESULTS - LUAD VS LUSC CLASSIFICATION

Log Odds Ratio - LR Coefficients

HPC	Log Odds Ratio	95% CI	P-Value	Predominant Lung Type	Purity %	Patients %
45	-0.36	(-0.6, -0.12)	0.003	LUSC	97	15.8
5	-0.2	(-0.38, -0.02)	0.028	LUSC	87	56.7
46	-0.17	(-0.39, 0.05)	0.129	LUSC	77	16.3
25	-0.14	(-0.34, 0.06)	0.166	LUSC	78	51.3
16	-0.13	(-0.32, 0.05)	0.148	LUSC	63	44.2
0	-0.11	(-0.28, 0.06)	0.2	LUSC	79	77.5
37	-0.05	(-0.24, 0.13)	0.573	LUAD	73	50.0
1	-0.05	(-0.26, 0.17)	0.654	LUSC	73	90.8
34	-0.03	(-0.26, 0.19)	0.773	LUSC	56	82.4
6	-0.03	(-0.23, 0.16)	0.745	LUAD	50	84.4
21	-0.03	(-0.22, 0.16)	0.758	LUSC	58	72.5
32	-0.03	(-0.2, 0.14)	0.751	LUSC	63	62.5
15	-0.01	(-0.18, 0.16)	0.908	LUSC	77	50.8
41	0.0	(-0.17, 0.18)	0.975	LUAD	71	35.4
39	0.0	(-0.22, 0.23)	0.971	LUAD	81	38.2
29	0.01	(-0.17, 0.18)	0.942	LUSC	68	54.4
23	0.01	(-0.2, 0.23)	0.898	LUSC	50	82.9
9	0.02	(-0.18, 0.21)	0.866	LUSC	62	78.8
4	0.09	(-0.1, 0.27)	0.378	LUAD	71	55.4
18	0.09	(-0.15, 0.33)	0.455	LUAD	97	31.7
26	0.11	(-0.11, 0.33)	0.324	LUAD	95	48.7
35	0.24	(0.09, 0.39)	0.002	LUAD	74	54.3
22	0.26	(0.05, 0.47)	0.014	LUAD	96	42.4
36	0.26	(0.06, 0.46)	0.009	LUAD	86	41.4
28	0.27	(0.06, 0.48)	0.013	LUAD	89	60.7
31	0.34	(0.1, 0.57)	0.006	LUSC	53	94.6
11	0.36	(0.16, 0.56)	0.0	LUAD	97	44.1

RESULTS - LUAD VS LUSC CLASSIFICATION

Log Odds Ratio - LR Coefficients

HPC	Log Odds Ratio	95% CI	P-Value
45	-0.36	(-0.6, -0.12)	0.003
5	-0.2	(-0.38, -0.02)	0.028
46	-0.17	(-0.39, 0.05)	0.129
25	-0.14	(-0.34, 0.06)	0.166
16	-0.13	(-0.32, 0.05)	0.148
0	-0.11	(-0.28, 0.06)	0.2
37	-0.05	(-0.24, 0.13)	0.573
1	-0.05	(-0.26, 0.17)	0.654
34	-0.03	(-0.26, 0.19)	0.773
6	-0.03	(-0.23, 0.16)	0.745
21	-0.03	(-0.22, 0.16)	0.758
32	-0.03	(-0.2, 0.14)	0.751
15	-0.01	(-0.18, 0.16)	0.908
41	0.0	(-0.17, 0.18)	0.975
39	0.0	(-0.22, 0.23)	0.971
29	0.01	(-0.17, 0.18)	0.942
23	0.01	(-0.2, 0.23)	0.898
9	0.02	(-0.18, 0.21)	0.866
4	0.09	(-0.1, 0.27)	0.378
18	0.09	(-0.15, 0.33)	0.455
26	0.11	(-0.11, 0.33)	0.324
35	0.24	(0.09, 0.39)	0.002
22	0.26	(0.05, 0.47)	0.014
36	0.26	(0.06, 0.46)	0.009
28	0.27	(0.06, 0.48)	0.013
31	0.34	(0.1, 0.57)	0.006
11	0.36	(0.16, 0.56)	0.0

LUSC Clusters

RESULTS - LUAD VS LUSC CLASSIFICATION

Log Odds Ratio - LR Coefficients

HPC	Log Odds Ratio	95% CI	P-Value
45	-0.36	(-0.6, -0.12)	0.003
5	-0.2	(-0.38, -0.02)	0.028
46	-0.17	(-0.39, 0.05)	0.129
25	-0.14	(-0.34, 0.06)	0.166
16	-0.13	(-0.32, 0.05)	0.148
0	-0.11	(-0.28, 0.06)	0.2
37	-0.05	(-0.24, 0.13)	0.573
1	-0.05	(-0.26, 0.17)	0.654
34	-0.03	(-0.26, 0.19)	0.773
6	-0.03	(-0.23, 0.16)	0.745
21	-0.03	(-0.22, 0.16)	0.758
32	-0.03	(-0.2, 0.14)	0.751
15	-0.01	(-0.18, 0.16)	0.908
41	0.0	(-0.17, 0.18)	0.975
39	0.0	(-0.22, 0.23)	0.971
29	0.01	(-0.17, 0.18)	0.942
23	0.01	(-0.2, 0.23)	0.898
9	0.02	(-0.18, 0.21)	0.866
4	0.09	(-0.1, 0.27)	0.378
18	0.09	(-0.15, 0.33)	0.455
26	0.11	(-0.11, 0.33)	0.324
35	0.24	(0.09, 0.39)	0.002
22	0.26	(0.05, 0.47)	0.014
36	0.26	(0.06, 0.46)	0.009
28	0.27	(0.06, 0.48)	0.013
31	0.34	(0.1, 0.57)	0.006
11	0.36	(0.16, 0.56)	0.0

LUAD Clusters

RESULTS - LUAD VS LUSC CLASSIFICATION

Log Odds Ratio - LR Coefficients

HPC	Log Odds Ratio	95% CI	P-Value
45	-0.36	(-0.6, -0.12)	0.003
5	-0.2	(-0.38, -0.02)	0.028
46	-0.17	(-0.39, 0.05)	0.129
25	-0.14	(-0.34, 0.06)	0.166
16	-0.13	(-0.32, 0.05)	0.148
0	-0.11	(-0.28, 0.06)	0.2
37	-0.05	(-0.24, 0.13)	0.573
1	-0.05	(-0.26, 0.17)	0.654
34	-0.03	(-0.26, 0.19)	0.773
6	-0.03	(-0.23, 0.16)	0.745
21	-0.03	(-0.22, 0.16)	0.758
32	-0.03	(-0.2, 0.14)	0.751
15	-0.01	(-0.18, 0.16)	0.908
41	0.0	(-0.17, 0.18)	0.975
39	0.0	(-0.22, 0.23)	0.971
29	0.01	(-0.17, 0.18)	0.942
23	0.01	(-0.2, 0.23)	0.898
9	0.02	(-0.18, 0.21)	0.866
4	0.09	(-0.1, 0.27)	0.378
18	0.09	(-0.15, 0.33)	0.455
26	0.11	(-0.11, 0.33)	0.324
35	0.24	(0.09, 0.39)	0.002
22	0.26	(0.05, 0.47)	0.014
36	0.26	(0.06, 0.46)	0.009
28	0.27	(0.06, 0.48)	0.013
31	0.34	(0.1, 0.57)	0.006
11	0.36	(0.16, 0.56)	0.0

LUAD Clusters

RESULTS - LUAD VS LUSC CLASSIFICATION

TCGA-33-4532-01Z-00-DX3 - LUSC WSI

RESULTS - LUAD VS LUSC CLASSIFICATION

TCGA-33-4532-01Z-00-DX3 - LUSC WSI

RESULTS - LUAD VS LUSC CLASSIFICATION

TCGA-69-7761-01Z-00-DX1 - LUAD WSI

RESULTS - LUAD VS LUSC CLASSIFICATION

TCGA-69-7761-01Z-00-DX1 - LUAD WSI

RESULTS - LUAD VS LUSC CLASSIFICATION

UMAP tile representations and Leiden cluster network based on PAGA[5].

[5] Wolf, F. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells, 2019.

RESULTS - LUAD VS LUSC CLASSIFICATION

Bi-Hierarchical clustering of LUAD and LUSC WSI representations based on clusters percent contribution to the slide

UMAP of WSI representations labeled by subtype

RESULTS
LUAD SURVIVAL

RESULTS - LUAD OVERALL SURVIVAL

Overall survival:

- Cox proportional hazard regression over WSI representations.
- 5-fold cross-validation:
  - TCGA train and test sets.
  - NYU as an additional test set.
- Combining Cox coefficients and significance across folds:
  - Coefficients are averaged.
  - Fisher's method to combine p-values:
    - Statistic: $S = -2\sum ln(p_k)$
    - Significance: $1 - \chi^2(S; v = 2k)$

Performance	C-Index Mean	C-Index CI
TCGA	0.60	0.56-0.63
NYU	0.65	0.63-0.67

Overall Survival Performance
TCGA & NYU

RESULTS - LUAD OVERALL SURVIVAL

5-Fold Cross-Validation
Log Hazard Ratio

HPC	Log Hazard Ratio	95% CI	P-Value	Patients %
31	-0.028	(-0.061, 0.005)	0.008	88.5
1	-0.028	(-0.052, -0.004)	0.0	87.3
37	-0.023	(-0.04, -0.006)	0.0	51.8
0	-0.021	(-0.046, 0.003)	0.006	84.8
16	-0.018	(-0.042, 0.005)	0.012	80.5
8	-0.016	(-0.033, 0.001)	0.001	50.7
9	-0.015	(-0.039, 0.009)	0.082	87.1
32	-0.013	(-0.035, 0.008)	0.109	68.8
44	-0.012	(-0.042, 0.017)	0.509	48.2
21	-0.012	(-0.032, 0.007)	0.083	67.0
43	-0.011	(-0.037, 0.014)	0.381	43.4
36	-0.011	(-0.035, 0.014)	0.407	69.5
5	-0.011	(-0.026, 0.005)	0.049	61.1
23	-0.009	(-0.032, 0.015)	0.429	83.0
45	-0.008	(-0.049, 0.032)	0.85	11.3
26	-0.008	(-0.027, 0.011)	0.387	58.1
13	-0.006	(-0.03, 0.017)	0.595	83.7
18	-0.006	(-0.027, 0.015)	0.772	46.2
33	-0.005	(-0.032, 0.021)	0.822	67.2
17	-0.004	(-0.03, 0.023)	0.954	95.5
28	-0.004	(-0.025, 0.017)	0.952	31.0
19	-0.002	(-0.02, 0.015)	0.982	47.5
6	-0.002	(-0.019, 0.016)	0.985	76.9
40	-0.001	(-0.028, 0.026)	0.994	61.8
12	-0.001	(-0.024, 0.022)	0.912	45.2
24	0.0	(-0.02, 0.021)	0.99	49.3
29	0.001	(-0.02, 0.021)	0.951	49.3
42	0.001	(-0.03, 0.032)	0.909	42.3
35	0.002	(-0.026, 0.031)	0.993	31.9
2	0.002	(-0.016, 0.021)	0.895	77.2
38	0.003	(-0.02, 0.026)	0.651	33.2
30	0.007	(-0.019, 0.033)	0.614	60.6
34	0.007	(-0.013, 0.027)	0.65	64.0
4	0.01	(-0.012, 0.032)	0.411	53.6
25	0.01	(-0.007, 0.027)	0.122	63.1
11	0.01	(-0.009, 0.03)	0.224	66.8
7	0.011	(-0.005, 0.026)	0.035	71.5
27	0.011	(-0.013, 0.036)	0.378	39.4
10	0.012	(-0.012, 0.036)	0.296	67.6
3	0.015	(-0.002, 0.032)	0.006	73.8
14	0.016	(-0.004, 0.035)	0.02	60.2
20	0.018	(-0.005, 0.041)	0.021	68.1
22	0.019	(-0.009, 0.047)	0.051	90.3
41	0.021	(-0.006, 0.047)	0.015	59.0
39	0.026	(-0.002, 0.055)	0.002	37.3
15	0.03	(0.014, 0.046)	0.0	43.0

Kaplan-Meier survival curve
High/Low-risk groups

RESULTS - LUAD OVERALL SURVIVAL

5-Fold Cross-Validation
Log Hazard Ratio

HPC	Log Hazard Ratio	95% CI	P-Value
31	-0.028	(-0.061, 0.005)	0.008
1	-0.028	(-0.052, -0.004)	0.0
37	-0.023	(-0.04, -0.006)	0.0
0	-0.021	(-0.046, 0.003)	0.006
16	-0.018	(-0.042, 0.005)	0.012
8	-0.016	(-0.033, 0.001)	0.001
9	-0.015	(-0.039, 0.009)	0.082
32	-0.013	(-0.035, 0.008)	0.109
44	-0.012	(-0.042, 0.017)	0.509
21	-0.012	(-0.032, 0.007)	0.083
43	-0.011	(-0.037, 0.014)	0.381
36	-0.011	(-0.035, 0.014)	0.407
5	-0.011	(-0.026, 0.005)	0.049
23	-0.009	(-0.032, 0.015)	0.429
45	-0.008	(-0.049, 0.032)	0.85
26	-0.008	(-0.027, 0.011)	0.387
13	-0.006	(-0.03, 0.017)	0.595
18	-0.006	(-0.027, 0.015)	0.772
33	-0.005	(-0.032, 0.021)	0.822
17	-0.004	(-0.03, 0.023)	0.954
28	-0.004	(-0.025, 0.017)	0.952
19	-0.002	(-0.02, 0.015)	0.982
6	-0.002	(-0.019, 0.016)	0.985
40	-0.001	(-0.028, 0.026)	0.994
12	-0.001	(-0.024, 0.022)	0.912
24	0.0	(-0.02, 0.021)	0.99
29	0.001	(-0.02, 0.021)	0.951
42	0.001	(-0.03, 0.032)	0.909
35	0.002	(-0.026, 0.031)	0.993
2	0.002	(-0.016, 0.021)	0.895
38	0.003	(-0.02, 0.026)	0.651
30	0.007	(-0.019, 0.033)	0.614
34	0.007	(-0.013, 0.027)	0.65
4	0.01	(-0.012, 0.032)	0.411
25	0.01	(-0.007, 0.027)	0.122
11	0.01	(-0.009, 0.03)	0.224
7	0.011	(-0.005, 0.026)	0.035
27	0.011	(-0.013, 0.036)	0.378
10	0.012	(-0.012, 0.036)	0.296
3	0.015	(-0.002, 0.032)	0.006
14	0.016	(-0.004, 0.035)	0.02
20	0.018	(-0.005, 0.041)	0.021
22	0.019	(-0.009, 0.047)	0.051
41	0.021	(-0.006, 0.047)	0.015
39	0.026	(-0.002, 0.055)	0.002
15	0.03	(0.014, 0.046)	0.0

RESULTS - LUAD OVERALL SURVIVAL

5-Fold Cross-Validation
Log Hazard Ratio

HPC	Log Hazard Ratio	95% CI	P-Value
31	-0.028	(-0.061, 0.005)	0.008
1	-0.028	(-0.052, -0.004)	0.0
37	-0.023	(-0.04, -0.006)	0.0
0	-0.021	(-0.046, 0.003)	0.006
16	-0.018	(-0.042, 0.005)	0.012
8	-0.016	(-0.033, 0.001)	0.001
9	-0.015	(-0.039, 0.009)	0.082
32	-0.013	(-0.035, 0.008)	0.109
44	-0.012	(-0.042, 0.017)	0.509
21	-0.012	(-0.032, 0.007)	0.083
43	-0.011	(-0.037, 0.014)	0.381
36	-0.011	(-0.035, 0.014)	0.407
5	-0.011	(-0.026, 0.005)	0.049
23	-0.009	(-0.032, 0.015)	0.429
45	-0.008	(-0.049, 0.032)	0.85
26	-0.008	(-0.027, 0.011)	0.387
13	-0.006	(-0.03, 0.017)	0.595
18	-0.006	(-0.027, 0.015)	0.772
33	-0.005	(-0.032, 0.021)	0.822
17	-0.004	(-0.03, 0.023)	0.954
28	-0.004	(-0.025, 0.017)	0.952
19	-0.002	(-0.02, 0.015)	0.982
6	-0.002	(-0.019, 0.016)	0.985
40	-0.001	(-0.028, 0.026)	0.994
12	-0.001	(-0.024, 0.022)	0.912
24	0.0	(-0.02, 0.021)	0.99
29	0.001	(-0.02, 0.021)	0.951
42	0.001	(-0.03, 0.032)	0.909
35	0.002	(-0.026, 0.031)	0.993
2	0.002	(-0.016, 0.021)	0.895
38	0.003	(-0.02, 0.026)	0.651
30	0.007	(-0.019, 0.033)	0.614
34	0.007	(-0.013, 0.027)	0.65
4	0.01	(-0.012, 0.032)	0.411
25	0.01	(-0.007, 0.027)	0.122
11	0.01	(-0.009, 0.03)	0.224
7	0.011	(-0.005, 0.026)	0.035
27	0.011	(-0.013, 0.036)	0.378
10	0.012	(-0.012, 0.036)	0.296
3	0.015	(-0.002, 0.032)	0.006
14	0.016	(-0.004, 0.035)	0.02
20	0.018	(-0.005, 0.041)	0.021
22	0.019	(-0.009, 0.047)	0.051
41	0.021	(-0.006, 0.047)	0.015
39	0.026	(-0.002, 0.055)	0.002
15	0.03	(0.014, 0.046)	0.0

RESULTS - LUAD RECURRENCE FREE SURVIVAL

Recurrence free survival:

- Cox proportional hazard regression over WSI representations.
- 5-fold-cross validation:
  - NYU train and test sets.
- Recurrence runs: Prediction of systemic and locoregional
- Combining Cox coefficients and significance across folds:
  - Coefficients are averaged.
  - Fisher's method to combine p-values:
    - Statistic: $S = -2\sum ln(p_k)$
    - Significance: $1 - \chi^2(S; v = 2k)$

Recurrence Free Survival
systemic and locoregional

Performance	C-Index Mean	C-Index CI
NYU	0.74	0.70-0.80

RESULTS - LUAD RECURRENCE FREE SURVIVAL

5 Fold Cross Validation
Log Hazard Ratio

HPC	Log Hazard Ratio	95% CI	P-Value	Patients %
39	-0.153	(-0.305, -0.001)	0.0	6.9
45	-0.148	(-0.294, -0.001)	0.0	2.2
29	-0.14	(-0.259, -0.021)	0.0	10.1
27	-0.117	(-0.212, -0.023)	0.0	19.2
22	-0.071	(-0.152, 0.01)	0.006	98.2
36	-0.069	(-0.129, -0.009)	0.0	74.3
32	-0.068	(-0.134, -0.001)	0.0	96.0
34	-0.055	(-0.155, 0.045)	0.211	27.5
0	-0.049	(-0.109, 0.011)	0.009	79.0
18	-0.044	(-0.139, 0.051)	0.393	27.9
31	-0.041	(-0.151, 0.07)	0.624	96.0
37	-0.038	(-0.074, -0.003)	0.0	79.4
21	-0.038	(-0.089, 0.014)	0.039	72.5
23	-0.037	(-0.121, 0.047)	0.417	94.6
13	-0.035	(-0.096, 0.026)	0.146	94.9
33	-0.027	(-0.136, 0.082)	0.596	48.9
2	-0.024	(-0.083, 0.035)	0.551	98.2
42	-0.02	(-0.103, 0.063)	0.876	54.0
26	-0.02	(-0.067, 0.027)	0.46	87.0
40	-0.02	(-0.086, 0.047)	0.757	72.5
17	-0.018	(-0.099, 0.064)	0.844	84.4
16	-0.016	(-0.098, 0.066)	0.936	75.0
41	-0.016	(-0.116, 0.085)	0.409	43.5
30	-0.012	(-0.109, 0.085)	0.93	38.8
10	-0.011	(-0.068, 0.045)	0.706	86.6
12	-0.01	(-0.112, 0.091)	0.936	28.3
19	-0.01	(-0.073, 0.053)	0.946	44.2
1	-0.009	(-0.072, 0.054)	0.877	89.9
4	-0.006	(-0.068, 0.056)	0.996	54.0
3	0.005	(-0.053, 0.064)	0.994	97.1
9	0.009	(-0.051, 0.07)	0.951	94.9
43	0.014	(-0.073, 0.101)	0.84	47.8
14	0.015	(-0.041, 0.071)	0.516	66.3
20	0.019	(-0.045, 0.083)	0.767	75.4
35	0.021	(-0.116, 0.159)	0.966	12.3
7	0.024	(-0.04, 0.087)	0.522	93.8
25	0.025	(-0.028, 0.079)	0.308	54.3
8	0.026	(-0.013, 0.066)	0.064	69.2
28	0.041	(-0.045, 0.128)	0.325	23.6
38	0.05	(-0.069, 0.17)	0.484	16.7
24	0.053	(-0.011, 0.116)	0.009	52.9
5	0.057	(-0.017, 0.131)	0.027	35.9
44	0.078	(-0.03, 0.187)	0.045	40.6
6	0.08	(0.019, 0.142)	0.0	63.0
15	0.112	(0.056, 0.169)	0.0	43.8
11	0.14	(0.069, 0.21)	0.0	59.1

Recurrence free survival:

- Prediction of systemic and locoregional.

High-Low risk groups
NYU

UMAP of patient vector representations labeled by risk group

RESULTS - LUAD RECURRENCE FREE SURVIVAL

5 Fold Cross Validation
Log Hazard Ratio

HPC	Log Hazard Ratio	95% CI	P-Value
39	-0.153	(-0.305, -0.001)	0.0
45	-0.148	(-0.294, -0.001)	0.0
29	-0.14	(-0.259, -0.021)	0.0
27	-0.117	(-0.212, -0.023)	0.0
22	-0.071	(-0.152, 0.01)	0.006
36	-0.069	(-0.129, -0.009)	0.0
32	-0.068	(-0.134, -0.001)	0.0
34	-0.055	(-0.155, 0.045)	0.211
0	-0.049	(-0.109, 0.011)	0.009
18	-0.044	(-0.139, 0.051)	0.393
31	-0.041	(-0.151, 0.07)	0.624
37	-0.038	(-0.074, -0.003)	0.0
21	-0.038	(-0.089, 0.014)	0.039
23	-0.037	(-0.121, 0.047)	0.417
13	-0.035	(-0.096, 0.026)	0.146
33	-0.027	(-0.136, 0.082)	0.596
2	-0.024	(-0.083, 0.035)	0.551
42	-0.02	(-0.103, 0.063)	0.876
26	-0.02	(-0.067, 0.027)	0.46
40	-0.02	(-0.086, 0.047)	0.757
17	-0.018	(-0.099, 0.064)	0.844
16	-0.016	(-0.098, 0.066)	0.936
41	-0.016	(-0.116, 0.085)	0.409
30	-0.012	(-0.109, 0.085)	0.93
10	-0.011	(-0.068, 0.045)	0.706
12	-0.01	(-0.112, 0.091)	0.936
19	-0.01	(-0.073, 0.053)	0.946
1	-0.009	(-0.072, 0.054)	0.877
4	-0.006	(-0.068, 0.056)	0.996
3	0.005	(-0.053, 0.064)	0.994
9	0.009	(-0.051, 0.07)	0.951
43	0.014	(-0.073, 0.101)	0.84
14	0.015	(-0.041, 0.071)	0.516
20	0.019	(-0.045, 0.083)	0.767
35	0.021	(-0.116, 0.159)	0.966
7	0.024	(-0.04, 0.087)	0.522
25	0.025	(-0.028, 0.079)	0.308
8	0.026	(-0.013, 0.066)	0.064
28	0.041	(-0.045, 0.128)	0.325
38	0.05	(-0.069, 0.17)	0.484
24	0.053	(-0.011, 0.116)	0.009
5	0.057	(-0.017, 0.131)	0.027
44	0.078	(-0.03, 0.187)	0.045
6	0.08	(0.019, 0.142)	0.0
15	0.112	(0.056, 0.169)	0.0
11	0.14	(0.069, 0.21)	0.0

RESULTS - LUAD RECURRENCE FREE SURVIVAL

5 Fold Cross Validation
Log Hazard Ratio

HPC	Log Hazard Ratio	95% CI	P-Value
39	-0.153	(-0.305, -0.001)	0.0
45	-0.148	(-0.294, -0.001)	0.0
29	-0.14	(-0.259, -0.021)	0.0
27	-0.117	(-0.212, -0.023)	0.0
22	-0.071	(-0.152, 0.01)	0.006
36	-0.069	(-0.129, -0.009)	0.0
32	-0.068	(-0.134, -0.001)	0.0
34	-0.055	(-0.155, 0.045)	0.211
0	-0.049	(-0.109, 0.011)	0.009
18	-0.044	(-0.139, 0.051)	0.393
31	-0.041	(-0.151, 0.07)	0.624
37	-0.038	(-0.074, -0.003)	0.0
21	-0.038	(-0.089, 0.014)	0.039
23	-0.037	(-0.121, 0.047)	0.417
13	-0.035	(-0.096, 0.026)	0.146
33	-0.027	(-0.136, 0.082)	0.596
2	-0.024	(-0.083, 0.035)	0.551
42	-0.02	(-0.103, 0.063)	0.876
26	-0.02	(-0.067, 0.027)	0.46
40	-0.02	(-0.086, 0.047)	0.757
17	-0.018	(-0.099, 0.064)	0.844
16	-0.016	(-0.098, 0.066)	0.936
41	-0.016	(-0.116, 0.085)	0.409
30	-0.012	(-0.109, 0.085)	0.93
10	-0.011	(-0.068, 0.045)	0.706
12	-0.01	(-0.112, 0.091)	0.936
19	-0.01	(-0.073, 0.053)	0.946
1	-0.009	(-0.072, 0.054)	0.877
4	-0.006	(-0.068, 0.056)	0.996
3	0.005	(-0.053, 0.064)	0.994
9	0.009	(-0.051, 0.07)	0.951
43	0.014	(-0.073, 0.101)	0.84
14	0.015	(-0.041, 0.071)	0.516
20	0.019	(-0.045, 0.083)	0.767
35	0.021	(-0.116, 0.159)	0.966
7	0.024	(-0.04, 0.087)	0.522
25	0.025	(-0.028, 0.079)	0.308
8	0.026	(-0.013, 0.066)	0.064
28	0.041	(-0.045, 0.128)	0.325
38	0.05	(-0.069, 0.17)	0.484
24	0.053	(-0.011, 0.116)	0.009
5	0.057	(-0.017, 0.131)	0.027
44	0.078	(-0.03, 0.187)	0.045
6	0.08	(0.019, 0.142)	0.0
15	0.112	(0.056, 0.169)	0.0
11	0.14	(0.069, 0.21)	0.0

RESULTS - IMMUNE LANDSCAPE

Methodology:

- Spearman correlation between HPC % and immune feature value across patients.

Cox Coefficient colors:

- Purple - favors dead/recurrence.
- Green - against dead/recurrence.
- Color gradient based on cox coefficient.

Conclusion:

- Good outcomes are associated with HPCs that correlate with TIL, leukocyte, and storma fraction.
- Poor outcomes are associated with HPCs that correlate with proliferation and wound healing.

RESULTS - CELL ANNOTATIONS

Cell annotations:

- Column dendrogram according to immune features and HPC correlations.

Motivation:

- Relation between immune features (RNA seq), cell annotations (tissue), and HPCs.

Conclusions:

- Good outcomes are associated with HPCs that correlate with TIL, leukocyte, storma fraction, and over-representation of inflammatory and dead cells.
- Poor outcomes are associated with HPCs that correlate with proliferation, wound healing, and under-representations of inflammatory and dead cells.

RESULTS - HISTOLOGICAL ANNOTATIONS

Histological subtype annotations:

- Column dendrogram according to immune features and HPC correlations.

Motivation:

- Relation between immune features (RNA seq), histological subtype pattern (tissue), and HPCs.

Conclusions:

- Good outcomes are associated with HPCs that correlate with TIL, leukocyte, storma fraction, and under-representation for solid, over-representation lepidic.
- Poor outcomes are associated with HPCs that correlate with proliferation, wound healing, and solid patterns.

LUAD SURVIVAL - CLUSTERS

RESULTS
LUAD RECURRENCE FREE SURVIVAL
FRAMEWORK VS HISTOLOGICAL SUBTYPE ANNOTATIONS/HOVERNET

RESULTS - LUAD RECURRENCE FREE SURVIVAL

Motivation:

- Compare framework performance to other annotations like pathologist histological subtype percentage or HoverNet cell annotations.

LUAD Recurrence free survival:

- Prediction of systemic and locoregional.
- Cox regressions with covariates: clusters, histological % annotations, or HoverNet cell type counts.

Conclusion:

- Framework clusters showing best mean C-Index performance.
- Histological subtype percentage annotations showing smallest variance across 5 fold cross validation.

Cox - Clusters

Cox - Hist. Subtype annotations %

Cox - HoverNet Raw cell type counts

Cox - HoverNet Ratios cell type

RESULTS - LUAD RECURRENCE FREE SURVIVAL

LUAD Recurrence free survival:

- Prediction of systemic and locoregional.
- Cox regressions with covariates: clusters, histological % annotations, or HoverNet cell type counts.
- Log Hazard Ratio - Coefficient relevance for Histological Subtype percentage and HoverNet cell type counts.

Cox - Hist. Subtype annotations %

Cox - HoverNet Raw cell type counts

RESULTS - LUAD RECURRENCE FREE SURVIVAL

LUAD Recurrence free survival:

- Prediction of systemic and locoregional.
- Log Hazard Ratio - Coefficient relevance for Histological Subtype percentage and framework clusters.
- Coefficient values in the same range for both.
- Frame clusters show higher C-Index 0.74 vs 0.69

Cox - Hist. Subtype annotations %

Cox - Clusters

RESULTS - LUAD RECURRENCE FREE SURVIVAL

Cox Log Hazard Ratio: $ln\frac{h(t)}{h_0(t)} = \sum_{\#covarites} a_i X_i$

- Increase in hazard per unit increase in X_i while X_{-i} constant, how sharp the survival function declines with a covariate (acc hazard tru time, $S(t) = e^{\int_0^t h(u)du}$).
- Caveat: Need to come back from Centered Log Ratio Transformation:

$$clr(x) = [log(\frac{x_1}{g(x)}), \ldots, log(\frac{x_D}{g(x)})]$$

Cox - Hist. Subtype annotations %

Cox - Clusters

RESULTS
LUSC SURVIVAL

RESULTS - LUAD OVERALL SURVIVAL

Overall survival:

- Cox proportional hazard regression over WSI representations.
- 5 fold cross validation:
  - TCGA train and test sets.
- Combining Cox coefficients and significance across folds:
  - Coefficients are averaged.
  - Fisher's method to combine p-values:
    - Statistic: $S = -2\sum ln(p_k)$
    - Significance: $1 - \chi^2(S; v = 2k)$

Performance	C-Index Mean	C-Index CI
TCGA	0.58	0.54-0.63

Overall Survival Performance
TCGA

RESULTS - LUAD OVERALL SURVIVAL

5 Fold Cross Validation Log Hazard Ratio

Cluster	Log Hazard Ratio	95% CI	P-Value	Tile Counts	Mean Tile Pat.	Max Tile Pat. %	Patients %
8	-0.04	(-0.073, -0.006)	0.0	11936.0	45	12.9	70.4
14	-0.036	(-0.091, 0.018)	0.074	10792.0	30	8.1	92.1
33	-0.036	(-0.077, 0.004)	0.004	3068.0	35	7.5	23.0
28	-0.032	(-0.068, 0.005)	0.006	4721.6	28	18.0	44.4
24	-0.024	(-0.059, 0.011)	0.046	8753.6	30	9.1	76.2
10	-0.022	(-0.056, 0.012)	0.074	12724.8	46	14.8	73.2
17	-0.021	(-0.053, 0.01)	0.036	10014.4	43	20.2	61.2
4	-0.017	(-0.052, 0.017)	0.239	14477.6	46	4.7	82.8
5	-0.017	(-0.053, 0.02)	0.41	14721.6	43	6.8	89.3
26	-0.017	(-0.06, 0.027)	0.593	5658.4	22	6.0	66.7
23	-0.014	(-0.048, 0.02)	0.46	8061.6	32	5.5	66.3
29	-0.014	(-0.053, 0.026)	0.622	4349.6	21	7.9	53.4
20	-0.011	(-0.048, 0.026)	0.783	8517.6	31	8.3	71.2
22	-0.008	(-0.048, 0.031)	0.906	8372.0	27	13.2	80.9
31	-0.008	(-0.043, 0.028)	0.937	3373.6	19	6.1	44.8
11	-0.003	(-0.025, 0.019)	0.982	12728.8	54	3.5	61.8
0	-0.002	(-0.053, 0.049)	0.978	19318.4	51	7.1	99.4
27	0.003	(-0.04, 0.046)	1.0	4409.6	37	34.2	31.1
1	0.006	(-0.027, 0.04)	0.964	17430.4	53	6.3	87.1
7	0.007	(-0.024, 0.038)	0.836	14690.4	45	3.5	87.1
18	0.008	(-0.022, 0.037)	0.825	9637.6	37	4.5	68.7
30	0.009	(-0.037, 0.056)	0.887	3036.0	14	19.2	54.5
15	0.012	(-0.022, 0.045)	0.495	10636.0	33	8.7	83.3
3	0.012	(-0.016, 0.04)	0.477	16990.4	59	6.3	76.2
9	0.013	(-0.024, 0.05)	0.646	12341.6	38	10.6	84.8
13	0.014	(-0.02, 0.048)	0.497	11059.2	37	4.7	79.2
32	0.024	(-0.027, 0.074)	0.392	2900.0	12	9.6	57.9
12	0.025	(-0.011, 0.061)	0.05	12595.2	41	4.2	81.1
21	0.029	(-0.012, 0.07)	0.04	8284.0	27	12.2	80.3
19	0.029	(-0.005, 0.063)	0.006	9212.8	31	7.5	76.4
2	0.03	(-0.007, 0.066)	0.008	16785.6	49	3.0	89.7
16	0.031	(-0.004, 0.065)	0.005	9254.4	30	5.6	80.5
6	0.033	(-0.003, 0.069)	0.003	14336.0	46	8.2	82.6
25	0.04	(-0.016, 0.096)	0.044	6456.0	18	5.9	91.4

Kaplan-Meier survival curve High/Low risk groups TCGA

Overall Survival Performance TCGA

RESULTS - LUAD OVERALL SURVIVAL

5 Fold Cross Validation
Log Hazard Ratio

Cluster	Log Hazard Ratio	95% CI	P-Value
8	-0.04	(-0.073, -0.006)	0.0
14	-0.036	(-0.091, 0.018)	0.074
33	-0.036	(-0.077, 0.004)	0.004
28	0.032	(-0.068, 0.005)	0.006
24	0.024	(-0.059, 0.011)	0.046
10	-0.022	(-0.056, 0.012)	0.074
17	-0.021	(-0.053, 0.01)	0.036
4	-0.017	(-0.052, 0.017)	0.239
5	-0.017	(-0.053, 0.02)	0.41
26	-0.017	(-0.06, 0.027)	0.593
23	-0.014	(-0.048, 0.02)	0.46
29	-0.014	(-0.053, 0.026)	0.622
20	-0.011	(-0.048, 0.026)	0.783
22	-0.008	(-0.048, 0.031)	0.906
31	-0.008	(-0.043, 0.028)	0.937
11	-0.003	(-0.025, 0.019)	0.982
0	-0.002	(-0.053, 0.049)	0.978
27	0.003	(-0.04, 0.046)	1.0
1	0.006	(-0.027, 0.04)	0.964
7	0.007	(-0.024, 0.038)	0.836
18	0.008	(-0.022, 0.037)	0.825
30	0.009	(-0.037, 0.056)	0.887
15	0.012	(-0.022, 0.045)	0.495
3	0.012	(-0.016, 0.04)	0.477
9	0.013	(-0.024, 0.05)	0.646
13	0.014	(-0.02, 0.048)	0.497
32	0.024	(-0.027, 0.074)	0.392
12	0.025	(-0.011, 0.061)	0.05
21	0.029	(-0.012, 0.07)	0.04
19	0.029	(-0.005, 0.063)	0.006
2	0.03	(-0.007, 0.066)	0.008
16	0.031	(-0.004, 0.065)	0.005
6	0.033	(-0.003, 0.069)	0.003
25	0.04	(-0.016, 0.096)	0.044

RESULTS - LUAD OVERALL SURVIVAL

5 Fold Cross Validation
Log Hazard Ratio

Cluster	Log Hazard Ratio	95% CI	P-Value
8	-0.04	(-0.073, -0.006)	0.0
14	-0.036	(-0.091, 0.018)	0.074
33	-0.036	(-0.077, 0.004)	0.004
28	0.032	(-0.068, 0.005)	0.006
24	0.024	(-0.059, 0.011)	0.046
10	-0.022	(-0.056, 0.012)	0.074
17	-0.021	(-0.053, 0.01)	0.036
4	-0.017	(-0.052, 0.017)	0.239
5	-0.017	(-0.053, 0.02)	0.41
26	-0.017	(-0.06, 0.027)	0.593
23	-0.014	(-0.048, 0.02)	0.46
29	-0.014	(-0.053, 0.026)	0.622
20	-0.011	(-0.048, 0.026)	0.783
22	-0.008	(-0.048, 0.031)	0.906
31	-0.008	(-0.043, 0.028)	0.937
11	-0.003	(-0.025, 0.019)	0.982
0	-0.002	(-0.053, 0.049)	0.978
27	0.003	(-0.04, 0.046)	1.0
1	0.006	(-0.027, 0.04)	0.964
7	0.007	(-0.024, 0.038)	0.836
18	0.008	(-0.022, 0.037)	0.825
30	0.009	(-0.037, 0.056)	0.887
15	0.012	(-0.022, 0.045)	0.495
3	0.012	(-0.016, 0.04)	0.477
9	0.013	(-0.024, 0.05)	0.646
13	0.014	(-0.02, 0.048)	0.497
32	0.024	(-0.027, 0.074)	0.392
12	0.025	(-0.011, 0.061)	0.05
21	0.029	(-0.012, 0.07)	0.04
19	0.029	(-0.005, 0.063)	0.006
2	0.03	(-0.007, 0.066)	0.008
16	0.031	(-0.004, 0.065)	0.005
6	0.033	(-0.003, 0.069)	0.003
25	0.04	(-0.016, 0.096)	0.044

RESULTS - LUAD OVERALL SURVIVAL

5 Fold Cross Validation
Log Hazard Ratio

Cluster	Log Hazard Ratio	95% CI	P-Value
8	-0.04	(-0.073, -0.006)	0.0
14	-0.036	(-0.091, 0.018)	0.074
33	-0.036	(-0.077, 0.004)	0.004
28	0.032	(-0.068, 0.005)	0.006
24	0.024	(-0.059, 0.011)	0.046
10	-0.022	(-0.056, 0.012)	0.074
17	-0.021	(-0.053, 0.01)	0.036
4	-0.017	(-0.052, 0.017)	0.239
5	-0.017	(-0.053, 0.02)	0.41
26	-0.017	(-0.06, 0.027)	0.593
23	-0.014	(-0.048, 0.02)	0.46
29	-0.014	(-0.053, 0.026)	0.622
20	-0.011	(-0.048, 0.026)	0.783
22	-0.008	(-0.048, 0.031)	0.906
31	-0.008	(-0.043, 0.028)	0.937
11	-0.003	(-0.025, 0.019)	0.982
0	-0.002	(-0.053, 0.049)	0.978
27	0.003	(-0.04, 0.046)	1.0
1	0.006	(-0.027, 0.04)	0.964
7	0.007	(-0.024, 0.038)	0.836
18	0.008	(-0.022, 0.037)	0.825
30	0.009	(-0.037, 0.056)	0.887
15	0.012	(-0.022, 0.045)	0.495
3	0.012	(-0.016, 0.04)	0.477
9	0.013	(-0.024, 0.05)	0.646
13	0.014	(-0.02, 0.048)	0.497
32	0.024	(-0.027, 0.074)	0.392
12	0.025	(-0.011, 0.061)	0.05
21	0.029	(-0.012, 0.07)	0.04
19	0.029	(-0.005, 0.063)	0.006
2	0.03	(-0.007, 0.066)	0.008
16	0.031	(-0.004, 0.065)	0.005
6	0.033	(-0.003, 0.069)	0.003
25	0.04	(-0.016, 0.096)	0.044

RESULTS - LUAD OVERALL SURVIVAL

5 Fold Cross Validation
Log Hazard Ratio

Cluster	Log Hazard Ratio	95% CI	P-Value
8	-0.04	(-0.073, -0.006)	0.0
14	-0.036	(-0.091, 0.018)	0.074
33	-0.036	(-0.077, 0.004)	0.004
28	0.032	(-0.068, 0.005)	0.006
24	0.024	(-0.059, 0.011)	0.046
10	-0.022	(-0.056, 0.012)	0.074
17	-0.021	(-0.053, 0.01)	0.036
4	-0.017	(-0.052, 0.017)	0.239
5	-0.017	(-0.053, 0.02)	0.41
26	-0.017	(-0.06, 0.027)	0.593
23	-0.014	(-0.048, 0.02)	0.46
29	-0.014	(-0.053, 0.026)	0.622
20	-0.011	(-0.048, 0.026)	0.783
22	-0.008	(-0.048, 0.031)	0.906
31	-0.008	(-0.043, 0.028)	0.937
11	-0.003	(-0.025, 0.019)	0.982
0	-0.002	(-0.053, 0.049)	0.978
27	0.003	(-0.04, 0.046)	1.0
1	0.006	(-0.027, 0.04)	0.964
7	0.007	(-0.024, 0.038)	0.836
18	0.008	(-0.022, 0.037)	0.825
30	0.009	(-0.037, 0.056)	0.887
15	0.012	(-0.022, 0.045)	0.495
3	0.012	(-0.016, 0.04)	0.477
9	0.013	(-0.024, 0.05)	0.646
13	0.014	(-0.02, 0.048)	0.497
32	0.024	(-0.027, 0.074)	0.392
12	0.025	(-0.011, 0.061)	0.05
21	0.029	(-0.012, 0.07)	0.04
19	0.029	(-0.005, 0.063)	0.006
2	0.03	(-0.007, 0.066)	0.008
16	0.031	(-0.004, 0.065)	0.005
6	0.033	(-0.003, 0.069)	0.003
25	0.04	(-0.016, 0.096)	0.044

RESULTS - IMMUNE LANDSCAPE

Methodology:

- Spearman correlation between cluster % and immune feature value across patient.

Cox Coefficient colors:

- Purple - favors dead/recurrence.
- Green - against dead/recurrence.
- Color gradient based on cox coefficient.

Conclusion:

- Better outcomes are not as clearly associated with clusters that correlate wit TIL, leukocyte, and storm fraction, as compared to the LUAD OS case.

RESULTS - CELL ANNOTATIONS

Cell annotations:

- Column dendrogram according to immune features and cluster correlations.

Motivation:

- Relation between immune features (RNA seq), cell annotations (tissue) and clusters.

Conclusions:

- Clusters that correlate with TIL, leukocyte, and storm fraction and over-representations for inflammatory and dead cells. Not with over-representations of neoplastic as in the LUAD OS case.