A Methodology for Designing and Validating Computational Pathology Scores for Immune Cell Clustering in Tumor Biopsy Samples

Logan Cerkovnik, Karen Ryall, G. David Young, Kristen Wilson, Joseph Krueger
Flagship Biosciences, Inc, Westminster, CO.

Flagship BIOSCIENCES

POSTER # 5684

Abstract

BACKGROUND

Immune cell clustering is commonly observed in histopathology images. As the frequency and nature of immune cell clustering may represent biological phenomena critical to an immunotherapy response, they are import features to measure to differentiate immune phenotypes, which may predict patient responses to immunomodulating therapies. Computational pathology datasets and unsupervised machine-learning approaches are capable of measuring immune cell clustering using a variety of methods; however, it has not been clear how such measurements might be applied to generate a validated computational pathology score that truly captures the immune phenotype. This work explores a methodology for the development and analytical validation of digital pathology scores for immune cell clustering derived from the application of unsupervised learning to digital pathology datasets.

MATERIALS AND METHODS

Computational pathology data derived using Flagship's Computational Tissue Analysis (cTA®) platform from 12 non–small cell lung cancer (NSCLC) biopsy samples stained with a validated CD8–Ki-67 duplex immunohistochemistry (IHC) assay were used to create a virtual library of scores that described the clustering of CD8-positive (CD8+) stained cells. The library was created using different combinations of clustering methods, parameters, and scoring schemes. Scores with high distinguishability measured via a 2-way intraclass correlation coefficient, interrun precision measured via the coefficient of variation, and dynamic range were considered analytically validated. Principal component analysis and hierarchical clustering of the analytically validated scores were used to further optimize selection of the most informative subset of clustering scores from the library.

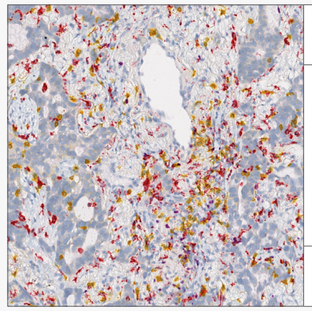
RESULTS

A total of 17 clustering scores passed the validation criteria. Of these 17 scores, 3 appear to be relatively uncorrelated with each other and capture unique information about the spatial relationships of CD8+ cells. A further analysis of these scores demonstrates the ability to distinguish different immune cell clustering profiles in samples that contain similar biomarker expression levels using the common scoring methods of both overall percentage of positive cells and percentage of positive cells per tissue area.

CONCLUSIONS

This process for screening and analytically validating a virtual library of computational biomarker scores appears to hold much promise for bringing cluster-derived computational pathology scores into future clinical applications in a way that is analogous to the analytical validation of traditional IHC assays in support of oncology drug development.

Immune Cell Clustering:
Spatial Heterogeneity Reflecting Immune Cell Biology



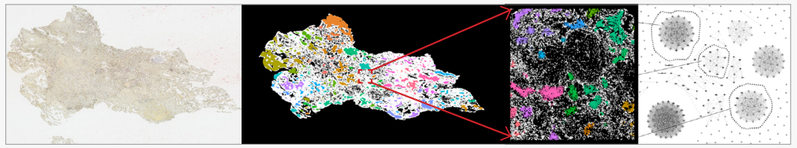
Random Immune Infiltrate	Competent Immune Infiltrate	Suppressed Immune Infiltrate	Excluded Immune Infiltrate
Tumor Epithelium	Tumor Epithelium	Tumor Epithelium	Tumor Epithelium
Immunological Incompetence	Immunological Induction	Immunological Tolerance	Adaptive Resistance

CD68 (red) **CD8** (yellow) **FoxP3** (purple)

Example: CD8+ Immune Clustering Analysis Using cTA

COMPUTATIONAL TISSUE SCORE

This is a continuous numeric variable describing an intrinsic property or feature from image analysis of a histopathology slide.



Flagship's cTA platform captures all cells across a whole tissue section and defines their spatial relationships. This spatial data is computed to create meaningful biological end points such as immune cell clustering.

Creating a Computational Pathology Score for CD8+ Cell Clustering

IMMUNE CLUSTERING SCORE

This is a computational tissue score that reports the degree of immune cell clustering for a whole slide as a continuous variable (example: CD8+ cells in an NSCLC tissue sample).

SCORING CHALLENGE

Many mathematical models exist for creating clusters from spatial data arrays and can be applied to cTA data derived from tissue; however, it is unclear what approach to defining and scoring clusters is most appropriate for capturing the biological relevance of immune cells in tissue biopsy samples.

APPROACH

A library of cluster scoring methods was evaluated to identify approaches that can generate unique, distinguishable, and reproducible scoring paradigms for immune cell clustering.

Immune Clustering Score Validation Approach

- Create library of candidate cTA scores derived from Biofeatures™
- Screen library using repeatability statistics, biological hypothesis, and visual confirmation
- Select candidate scores for further evaluation
- Evaluate and validate candidate scores using comparative datasets
- Utilize scores for improved outcome, patient stratification

Validation Criteria for Immune Clustering Score

Qualification Mark	Formula
Repeatability: A sample will give similar scores across sections	**Coefficient of Variation** $CV = \frac{\sigma}{\mu}$
Distinguishability: Repeated measurements of a sample do not change its relative ranking	**Interclass Correlation Coefficient** $\rho = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_e}$
Dynamic Range (DR): How much greater the score's maximum value is than its standard deviation at the low end	**Dynamic Range** $DR = log\left(\frac{max\ (score)}{\sigma}\right)$
Signal to Noise Ratio (SNR): How much greater a score's average value is than its standard deviation	**Signal to Noise Ratio** $SNR = \frac{\mu}{\sigma}$

Principal component analysis and hierarchical clustering of the analytically validated scores were used to further optimize selection of the most informative subset of clustering scores from the library.

Score Library Validation Methodology

Examining Degree of Clustering Scores for Whole-Slide CD8+ Cells in NSCLC

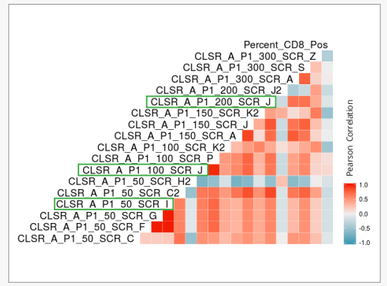
240+ Different Scores

17 with High Interrun Precision
2-way intraclass correlation coefficient lower confidence limit > 0.3

13 with Low Correlation with % Positive
|Pearson Correlation| < 0.4

3 Groups Noncorrelated
|Pearson correlation| < ~ 0.5

Three statistical methods were defined from a library of 240+ scores, which created a reproducible scoring method that distinguishes CD8+ cell clustering between samples.

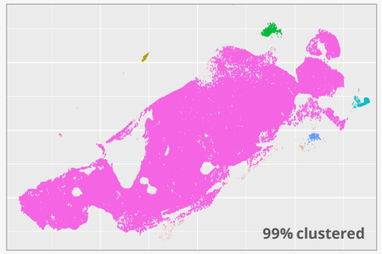


Pearson Correlation Coefficients of Distinguishable Library Scores

Beyond Percent Positive:
CD8+ Cell Abundance vs CD8+ Cell Clustering

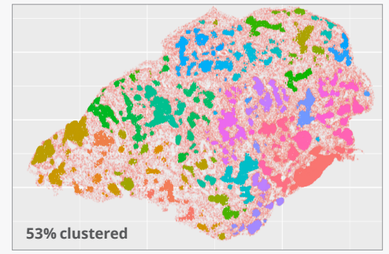
Two samples may have the same CD8 positivity score, but different immune phenotype!

21% CD8+ cells



The percentage of CD8+ cells demonstrates immune infiltration, and a high prevalence of immune cell clustering suggests stronger immune competence than less prevalent immune cell clustering.

23% CD8+ cells



The percentage of CD8+ cells demonstrates immune infiltration, but a lack of immune cell clustering suggests reduced immune competence compared with a tissue sample with more prevalent immune cell clustering.

Conclusion

- Multiple different scoring methodologies for CD8+ immune cell clustering in NSCLC samples are ready for further evaluation in clinical cohorts.
- At least 3 distinct types of cluster scores appear to be present in the NSCLC validation cohort.
- Cluster scoring methods can generate unique, distinguishable, and reproducible features for computational biomarkers and digital pathology scoring paradigms.