Machine learning-based identification of H&E-derived morphologic features associated with CD8+ T cell immune exclusion

Yanchao Wang[1], Fredrick D. Gootkind[1], Florent Peyraud[2], Jean-Philippe Guégan[3], Antoine Italiano[2], G. Travis Clifton[1], Laura A. Dillon[1], Xinwei Sher[1]

[1]Incendia Therapeutics, Boston, MA, USA; [2]Early Phase Trials and Sarcoma Unit, Institut Bergonié, Bordeaux, France; University of Bordeaux, Bordeaux, France; [3]Explicyte Immuno-Oncology, Bordeaux, France

Background
Immune exclusion is characterized by a predominance of CD8+ T cells in the stroma of the tumor microenvironment (TME) which are not in contact with tumor cells. Immune exclusion is a significant obstacle to effective cancer immunotherapy and reversing it has emerged as a new therapeutic opportunity. However, the objective assessment of immune exclusion remains a challenge due to the complexity of the TME and the need for immunostaining to quantify CD8+ cells in tumor and stroma. While artificial intelligence (AI) models applied to H&E images can quantify spatial tissue and cell features at scale, they are unable to identify the molecular subtypes of cells. Here we describe the development of a machine learning model that classifies tumors by their CD8-defined immune phenotypes using features extracted from H&E whole slide images alone.

Methods
AI-powered TME models (PathAI PathExplore™) were deployed on H&E images from 39 non-small lung cancer (NSCLC) and triple-negative breast cancer (TNBC) samples. CD8-based assessment of immune phenotypes was done using an adjacent section. 115 H&E-derived features from each of the 25 tumor samples (12 NSCLC, 13 TNBC) classified as immune excluded and 7 tumor samples (3 NSCLC, 4 TNBC) classified as immune infiltrated were used as input into a feature selection process to identify the subset of H&E features predictive of immune phenotype. The feature selection process was driven by an ensemble L1 norm support vector machine (Ens-L1-SVM) model containing 500 L1-SVM base models each trained by 80% of the data. The degree of immune exclusion was determined in the feature space by the distance to the separation hyperplane defined by the model. Out-of-bag error estimation was employed on the remaining 20% of data from each base model to assess model performance.

Results
The Ens-L1-SVM model identified 6 H&E features predictive of CD8-defined immune phenotypes, including the density ratio of lymphocytes to macrophages in the cancer epithelium, the density of fibroblasts, and the spatial distributions of cancer cells, macrophages, and lymphocytes. Immune phenotypes were classified by the Ens-L1-SVM model with an estimated accuracy of 97%. Additionally, all selected features and

their direction of association aligned with prior knowledge of the mechanisms or manifestations of immune exclusion or infiltration in the TME.

Conclusion

This work illustrated that morphologic features derived from H&E images using AI-powered models can be effective predictors of CD8-defined immune exclusion, providing an option for patient stratification by immune phenotype using widely-available H&E images. While the Ens-L1-SVM model showed robust performance on a small number of samples and is expected to have good generalizability due to the linear nature of the model, additional validation across diverse datasets and various tumor indications is needed.