A deep learning analysis pipeline for multiplex imaging identifies spatial features associated with clinical outcome in colorectal cancer

Etta Markovits1, Tal Dankovich1, Roman Guskii1, Ido Weiss1, Tommy Dicker2, Sun Dagan1, Ron Eitan1, Bexey Arbin1, Ami Bart1, Ori Zelichov1,2

Nucleai, Tel Aviv, Israel. Corresponding authors: etta.markovits@nucleai.ai, ori.zelichov@nucleai.ai

INTRODUCTION

- Multiplex immunofluorescence (mIF) can provide invaluable insights on spatial biology and the complexities of the immune tumor microenvironment (iTME), however, current analysis methods are laborious and user-dependent.
- We applied a novel end-to-end deep learning (DL) pipeline to mIF tumor-microarray images, to investigate associations between iTME composition and clinical outcome (Fig. 1A).

METHODS

- A publicly available CODEX dataset, consisting of 140 tissue cores from 35 colorectal cancer (CRC) patients stained with 56 protein markers and matching H&E slides was analyzed.
- We identified stromal abundance of a plasma cell-enriched neighborhood (“N_6”) as the top feature associated with worse prognosis (p=0.009), while enrichment of LAG3+ (>75% accuracy) by qualitative assessment in 97.7% of cores, thus markedly outperforming existing clustering-based cell typing approaches, which exhibited 65.9% accuracy (Fig. 2B).
- Phenotypic markers binary quantification demonstrated high accuracy and robustness, as reflected by >90% accuracy in almost all phenotypic markers, even though the model was not trained on these channels (Fig. 2C).
- Validation of the binary classifier performance in a melanoma cohort stained by Phenoimager for 5 lineage markers demonstrated a 91.7% balanced accuracy and a F1-score above 80% in all channels (Fig 2D). As the DL binary classifier was trained on the CRC CODEX cohort, we demonstrate robustness of the binary classifier for tumor type, different markers and imaging modalities.
- Upregulation of LAG3+CD4+ T-cells in a CD8+ T-cell enriched neighborhood (“N_2”) was also associated with better prognosis (p=0.006, Fig. 3B). These cells were shown to produce interferon-gamma in several previous works.
- Clustering of “local” cell neighborhoods resulted in an optimal solution of 12 cellular neighborhoods (Fig. 3A).
- We identified stromal abundance of a plasma cell-enriched neighborhood (“N_6”) as the top feature associated with worse prognosis (p=0.009), while enrichment of LAG3+ (>75% accuracy) by qualitative assessment in 97.7% of cores, thus markedly outperforming existing clustering-based cell typing approaches, which exhibited 65.9% accuracy (Fig. 2B).
- Phenotypic markers binary quantification demonstrated high accuracy and robustness, as reflected by >90% accuracy in almost all phenotypic markers, even though the model was not trained on these channels (Fig. 2C).
- Validation of the binary classifier performance in a melanoma cohort stained by Phenoimager for 5 lineage markers demonstrated a 91.7% balanced accuracy and a F1-score above 80% in all channels (Fig 2D). As the DL binary classifier was trained on the CRC CODEX cohort, we demonstrate robustness of the binary classifier for tumor type, different markers and imaging modalities.

RESULTS

- Model performance was evaluated quantitatively on 1,800 annotations from 14 test cores, and qualitatively on all cores by expert pathologists.
- For further validation of the binary classifier on a different imaging modality, we evaluated its performance on 5,000 annotations from a melanoma dataset stained with Phenoimager.
- Cell neighborhoods were identified by clustering the 10 nearest-neighbors for each cell utilizing the CLARA algorithm. Optimal number of clusters was identified using the “elbow rule” on within-cluster sum of squares (WSS) as measure of the variability of the observations within each cluster.
- Combining data from merged mIF & H&E cell typing, phenotypic marker positivity and neighborhood assignment, over 600 spatial features were calculated and correlated with good vs. bad prognosis defined by 2-year overall-survival (OS) cutoff.

CONCLUSIONS

- A novel DL pipeline for mIF analysis demonstrated high accuracy in classifying cell types and phenotypic markers, demonstrating marked improvement over current methods.
- Moreover, our DL binary classifier demonstrated robustness across cancer types, imaging modalities and different markers, thus paving the way for efficient and accurate analysis of multiplex imaging.
- Lastly, we demonstrate the importance of complex spatial features, which combine cell typing, neighborhood assignment and phenotypic marker expression, for capturing important information for outcome prediction.

A multiplex imaging analysis pipeline combining 5 DL models for cell segmentation, IF and H&E based cell typing, phenotypic markers binary quantification, and H&E areas segmentation. (B) Cell identification methodology based on marker expression and the immunomodulatory proteins quantified for each cell population.

Figure 1. (A) A multiplex imaging analysis pipeline combining 5 DL models for cell segmentation, IF and H&E based cell typing, phenotypic markers binary quantification, and H&E areas segmentation. (B) Cell identification methodology based on marker expression and the immunomodulatory proteins quantified for each cell population.

Figure 2. (A) Per-class model performance of the cell typing annotation-based DL model. (B) Comparison of the per-class F1 scores of cell typing by our DL model vs. clustering-based cell typing. (C) Accuracy of our DL binary marker classifier in the lineage markers, which it was trained upon, and phenotypic markers, which it was not trained upon. (D) Model performance of the binary classifier, which was trained on CRC codes images, in a melanoma cohort stained by Phenoimager.

Figure 2. (A) Per-class model performance of the cell typing annotation-based DL model. (B) Comparison of the per-class F1 scores of cell typing by our DL model vs. clustering-based cell typing. (C) Accuracy of our DL binary marker classifier in the lineage markers, which it was trained upon, and phenotypic markers, which it was not trained upon. (D) Model performance of the binary classifier, which was trained on CRC codes images, in a melanoma cohort stained by Phenoimager.

Figure 3. (A) Heatmap of cell type fraction in each neighborhood. (B) The total within-cluster sum of squares as a function of the number of clusters identified by the CLARA clustering method. (C) Representative images for cell typing and cell neighborhoods. (D) The top spatial features associated with clinical outcome of CRC patients.

A novel DL pipeline for mIF analysis demonstrated high accuracy in classifying cell types and phenotypic markers, demonstrating marked improvement over current methods.

Moreover, our DL binary classifier demonstrated robustness across cancer types, imaging modalities and different markers, thus paving the way for efficient and accurate analysis of multiplex imaging.

Lastly, we demonstrate the importance of complex spatial features, which combine cell typing, neighborhood assignment and phenotypic marker expression, for capturing important information for outcome prediction.