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

Ettai Markovits[1,2], Tal Dankovich[1], Roman Gluskin[1], Ido Weiss[1], Tomer Dicker[1], Sun Dagan[1], Ron Elran[1], Becky Arbiv[1], Amit Bart[1], Ori Zelichov[1,2]

[1]Nucleai, Tel Aviv, Israel, [2]Corresponding authors: ettai.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).

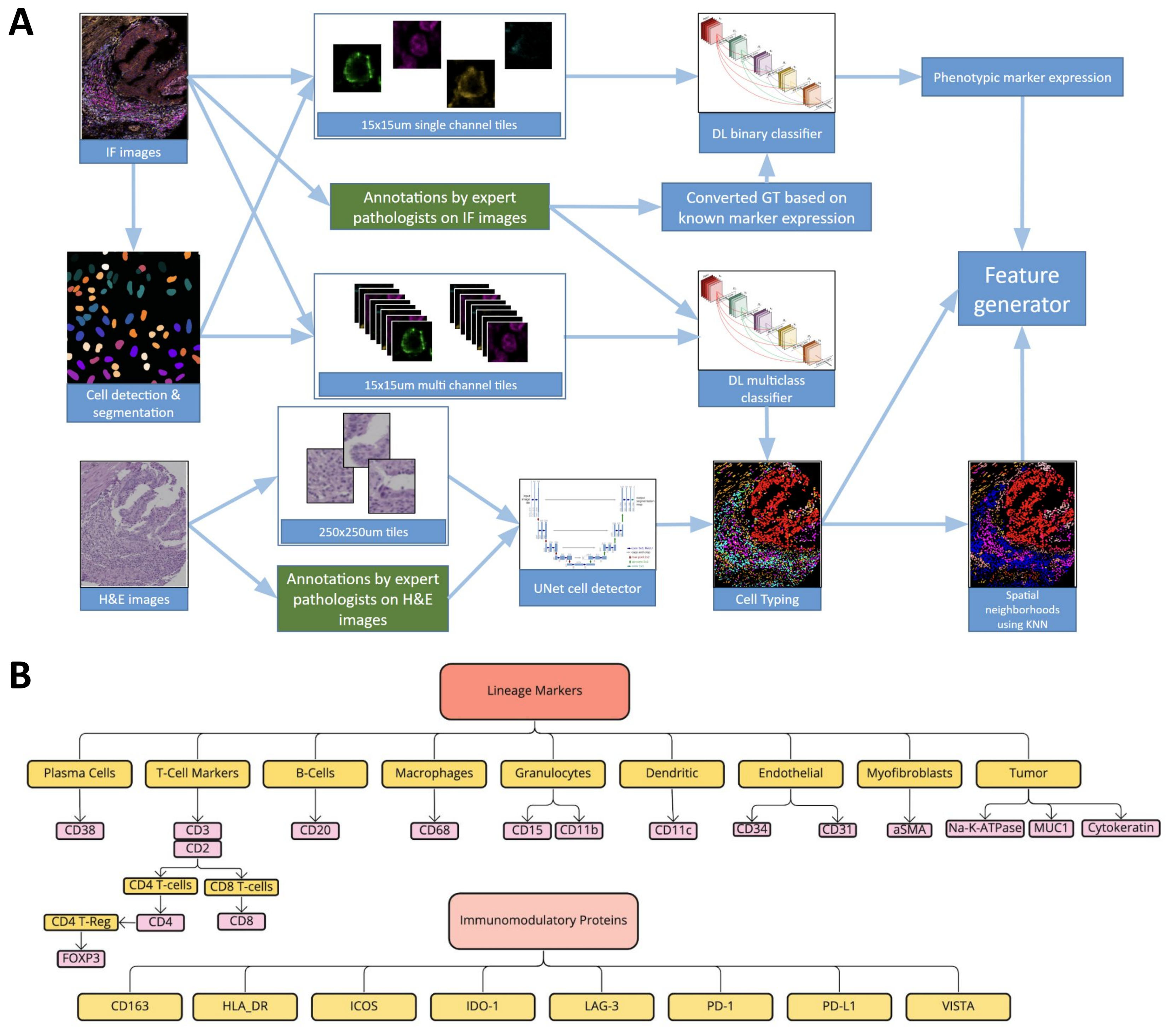

A
IF images
15x15um single channel tiles
DL binary classifier
Phenotypic marker expression
Annotations by expert pathologists on IF images
Converted GT based on known marker expression
Feature generator
Cell detection & segmentation
15x15um multi channel tiles
DL multiclass classifier
H&E images
250x250um tiles
Annotations by expert pathologists on H&E images
UNet cell detector
Cell Typing
Spatial neighborhoods using KNN
B
Lineage Markers
Plasma Cells
T-Cell Markers
B-Cells
Macrophages
Granulocytes
Dendritic
Endothelial
Myofibroblasts
Tumor
CD38
CD3
CD2
CD20
CD68
CD15
CD11b
CD11c
CD34
CD31
aSMA
Na-K-ATPase
MUC1
Cytokeratin
CD4 T-cells
CD8 T-cells
CD4 T-Reg
CD4
CD8
FOXP3
Immunomodulatory Proteins
CD163
HLA_DR
ICOS
IDO-1
LAG-3
PD-1
PD-L1
VISTA

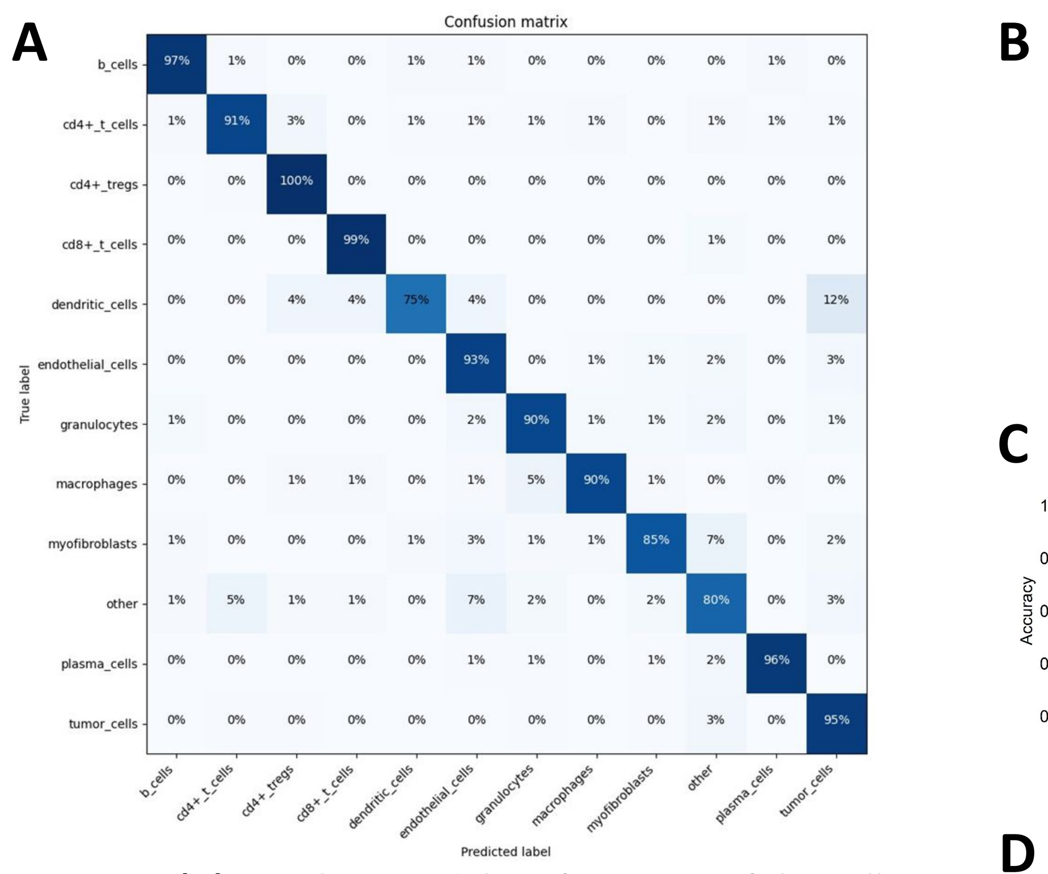
Figure 1. (A) A multiplex imaging analysis pipeline combining 5 DL models for cell segmentation, IF and H&E based cell typing, phenotypic markers binary quantification, and H&E areas segmentation. **(B)** Cell identification methodology based on marker expression and the immunomodulatory proteins quantified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

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.
- 7,000 cell annotations on 57 tissue cores were used to train DL models to segment cells and classify them into 13 subpopulations (Fig.1B). Positivity for 8 immunomodulatory markers was predicted by a DL binary classifier which was trained on 100,000 binary single channel annotations, extracted from cell type annotations. H&E slides were utilized to annotate tumor and stromal areas and identify fibroblasts and marker-negative tumor cells.
- 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 55,000 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 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.

RESULTS

- Cell typing DL model based on annotations reached a 91.9% accuracy (Fig. 2A) and good performance (>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.


A
Confusion matrix
True label
Predicted label
b_cells 97% 1% 0% 0% 1% 1% 0% 0% 0% 0% 1% 0%
cd4+_t_cells 1% 91% 3% 0% 1% 1% 1% 1% 0% 1% 1% 1%
cd4+_tregs 0% 0% 100% 0% 0% 0% 0% 0% 0% 0% 0% 0%
cd8+_t_cells 0% 0% 0% 99% 0% 0% 0% 0% 0% 1% 0% 0%
dendritic_cells 0% 0% 4% 4% 75% 4% 0% 0% 0% 0% 0% 12%
endothelial_cells 0% 0% 0% 0% 0% 93% 0% 1% 1% 2% 0% 3%
granulocytes 1% 0% 0% 0% 0% 2% 90% 1% 1% 2% 0% 1%
macrophages 0% 0% 1% 1% 0% 1% 5% 90% 1% 0% 0% 0%
myofibroblasts 1% 0% 0% 0% 1% 3% 1% 1% 85% 7% 0% 2%
other 1% 5% 1% 1% 0% 7% 2% 0% 2% 80% 0% 3%
plasma_cells 0% 0% 0% 0% 0% 1% 1% 0% 1% 2% 96% 0%
tumor_cells 0% 0% 0% 0% 0% 0% 0% 0% 0% 3% 0% 95%
b_cells cd4+_t_cells cd4+_tregs cd8+_t_cells dendritic_cells endothelial_cells granulocytes macrophages myofibroblasts other plasma_cells tumor_cells

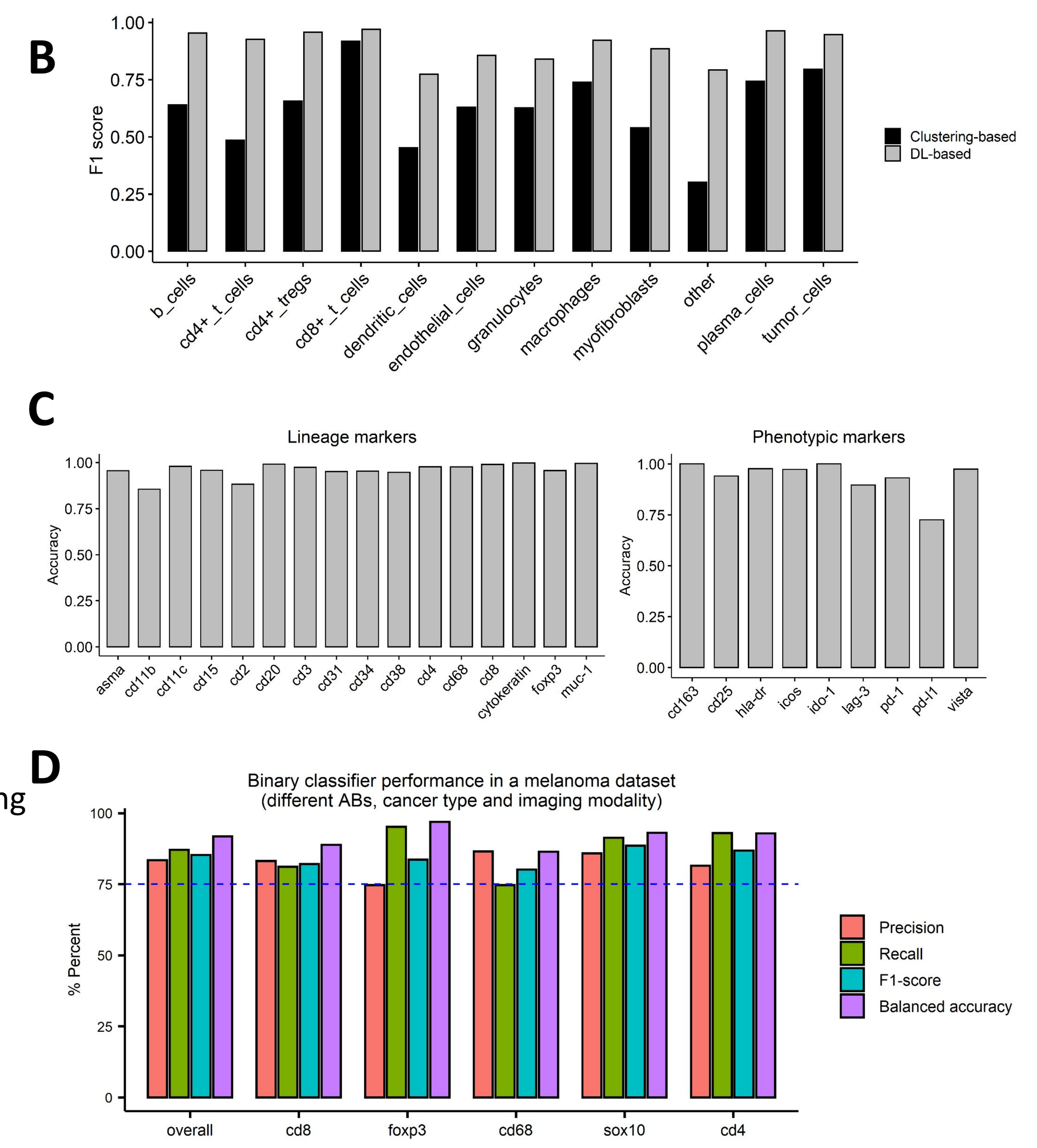

B
F1 score
Clustering-based
DL-based
b_cells cd4+_t_cells cd4+_tregs cd8+_t_cells dendritic_cells endothelial_cells granulocytes macrophages myofibroblasts other plasma_cells tumor_cells
C
Lineage markers
Accuracy
asma cd11b cd11c cd15 cd2 cd20 cd3 cd31 cd34 cd38 cd4 cd68 cd8 cytokeratin foxp3 muc-1
Phenotypic markers
cd163 cd25 hla-dr icos ido-1 lag-3 pd-1 pd-l1 vista
D
Binary classifier performance in a melanoma dataset
(different ABs, cancer type and imaging modality)
% Percent
Precision
Recall
F1-score
Balanced accuracy
overall cd8 foxp3 cd68 sox10 cd4

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 codex images, in a melanoma cohort stained by Phenoimager.

- 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+ (p=0.01) and VISTA+ (p=0.004) plasma cells in the plasma-enriched neighborhood were among the top features associated with good prognosis (Fig. 3B). This demonstrates that the functional state of plasma cells within the plasma-enriched neighborhood dictates the prognosis of CRC patients.
- 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.

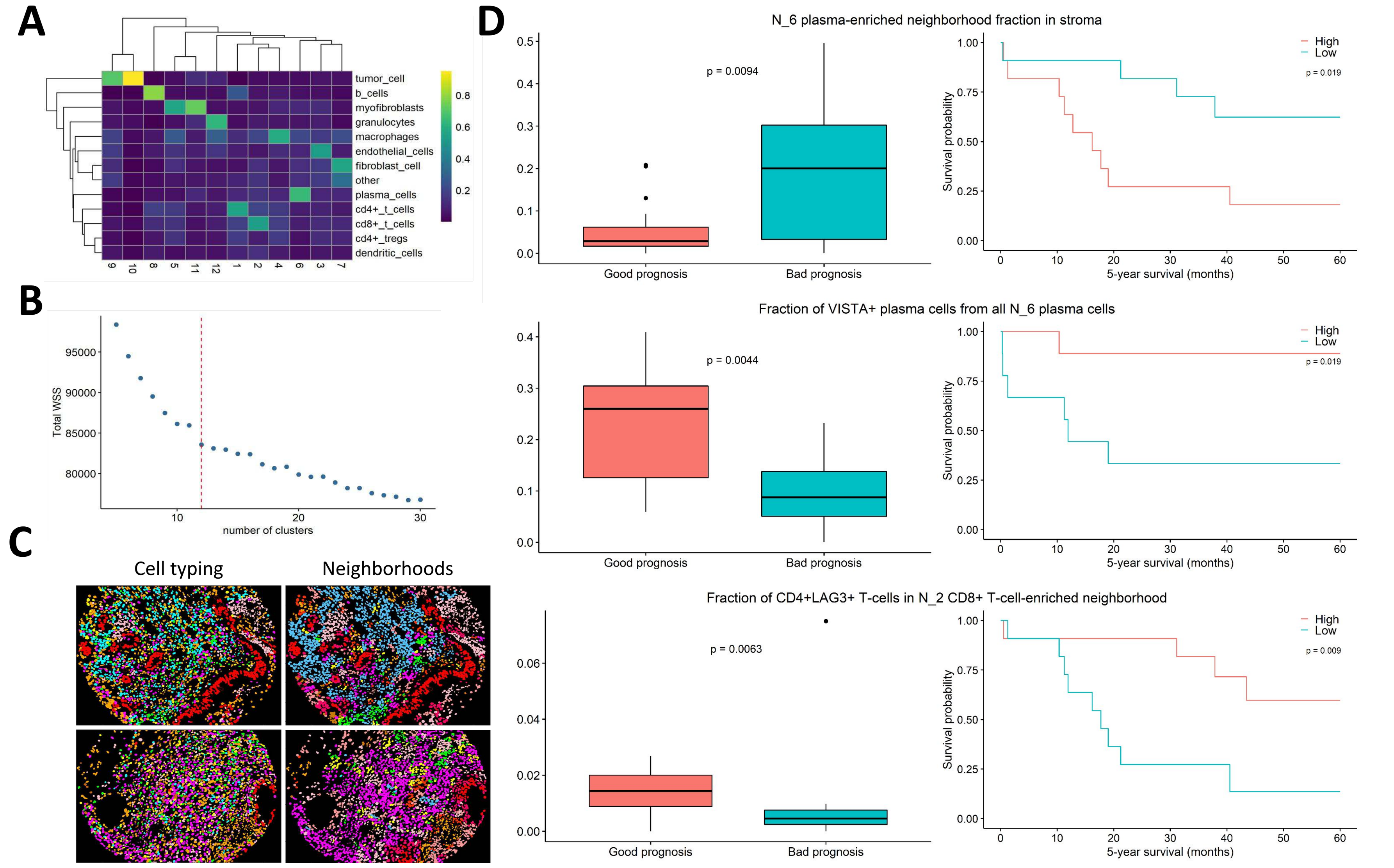

A
tumor_cell
b_cells
myofibroblasts
granulocytes
macrophages
endothelial_cells
fibroblast_cell
other
plasma_cells
cd4+_t_cells
cd8+_t_cells
cd4+_tregs
dendritic_cells
B
Total WSS
number of clusters
C
Cell typing
Neighborhoods
D
N_6 plasma-enriched neighborhood fraction in stroma
p = 0.0094
Good prognosis
Bad prognosis
High
Low
p = 0.019
Survival probability
5-year survival (months)
Fraction of VISTA+ plasma cells from all N_6 plasma cells
p = 0.0044
p = 0.019
Fraction of CD4+LAG3+ T-cells in N_2 CD8+ T-cell-enriched neighborhood
p = 0.0063
p = 0.009

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.

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.