Spatial analytics of the tumor microenvironment on double stained immunohistochemistry images using deep learning

Máté Levente Nagy, Harry Nunns, Naiyun Zhou, Ben Ovadia, Anna Juncker-Jensen, Karen Yamamoto, Qingyan Au, Ainura Kyshtoobayeva, Josette William Ragheb
NeoGenomics Laboratories, Aliso Viejo, CA

Background

Spatial locations of immune cells in the tumor microenvironment (TME) have been shown to correlate with clinical outcome in different cancers. A worse patient outcome has been reported in oral squamous cancer for individuals with an increased number of Tregs within 30 μm of CD8+ cells [1]. Likewise, the spatial relationship between CD8+ and PD-L1+ cells has become an area of interest as a possible indication for the response to PD-L1-inhibition. It was recently shown in a retrospective study that the combined assessment of CD8 and PD-L1 in NSCLC tumors outperformed CD8 or PD-L1 alone as prognostic markers for predicting treatment with immune checkpoint inhibitors [2].

As reported in these cases, quantifying spatial relationships between two biomarkers can provide clinical and/or biological insights. Spatial analytics of the TME requires accurate cell segmentation and classification. To that end, NeoGenomics has developed an IHC assay for combined CD8 and 22C3/PD-L1 staining, as well as a deep learning pipeline to automatically identify, segment, and classify CD8+ and PD-L1+ cells from whole slide IHC images. We see a 95% concordance of the number of CD8+ and PD-L1+ cells detected in our Dual IHC assay with serial sections stained either for CD8 or PD-L1 alone. Our analysis results are comparable to pathologist scoring of Dual IHC.

We integrated our previously-reported cell segmentation and classification workflow used for MultiOmyx™ data (multiplexed IF images) [3] with Indica HALO® [4] to analyze whole slide IHC images double stained using two distinct chromogens. In this study, we performed IHC double staining and cell classification analysis of a CD8 / PD-L1 assay in NSCLC tumors. In addition to cell segmentation and classification image outputs, we also generate cell-level and slide-level tables with various cell morphological information, phenotype counts and densities, and biomarker intensity values that can be used to automatically define H-score measures for each biomarker. Additionally, advanced spatial analytics is performed to calculate the average distance between cells of various phenotypes, and spatial clustering patterns of different phenotypes in the TME (i.e. CD8+ and PD-L1+ cells). These analyses enable investigation of numerous complex cell interactions in TMEs.

NeoGenomics quantitative Dual IHC assay is compatible with any two biomarkers of interest even if they are expressed on the same cell as long as the sub-cellular localization of the markers is different. The combination of double staining for CD8 and PD-L1, and quantifying spatial relationships between the two biomarkers is capable of providing the needed context to guide experimental decisions.

Workflow Steps

The framework consists of 11 major steps:
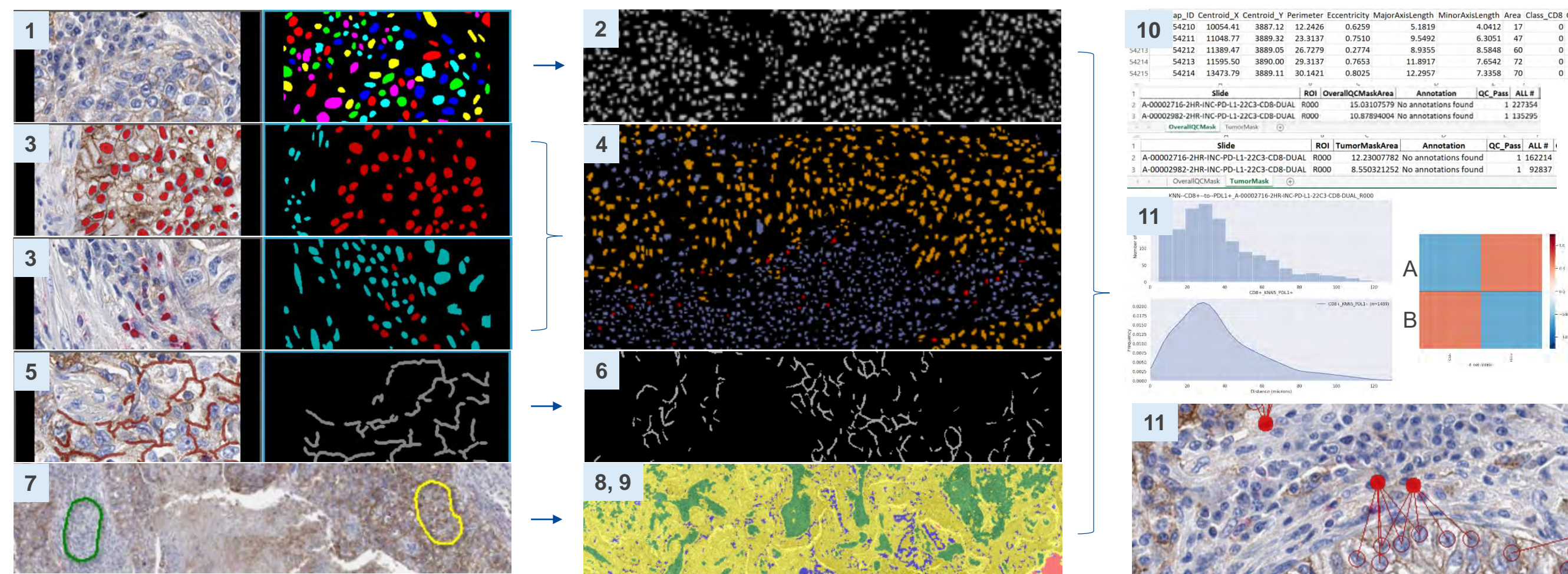
Figure 1: Workflow:
1. Manual annotation and training of a mask-RCNN [5] model for nuclear segmentation using selected IHC image crops
2. Application of the trained network from (1) on full IHC image to delineate individual cells and assign a unique label to each
3. Manual annotation and training of a CNN [6] model for each biomarker (CD8 and PD-L1) for cell classification
4. Application of the models from (3) on the full IHC image to classify each cell as CD8, PD-L1, or negative
5. Manual annotation and training of a CNN (U-Net) [7] model for PD-L1 membrane detection
6. Application of the models from (5) on the full IHC image to identify PD-L1 membrane regions
7. Manual annotation and training of HALO®'s random forest classification module to segment tissue, tumor, necrosis, and glass regions
8. Application of the model from (7) on the full IHC image to segment it into the regions defined by the model
9. Exporting of the annotation regions from HALO® to assign region information for each cell identified
10. Running post-classification analysis to capture cell classification density, intensity values (H-score), and spatial information
11. Visualization of results using heat maps, plots, and graphs

Classification Overlays and Spatial Analysis

The primary way to understand the workflow output is via overlays of the original IHC stains and the corresponding classification label maps and tables. In addition to cell classification, tumor segmentation is also performed to aid in better understanding of the tumor micro-environment.
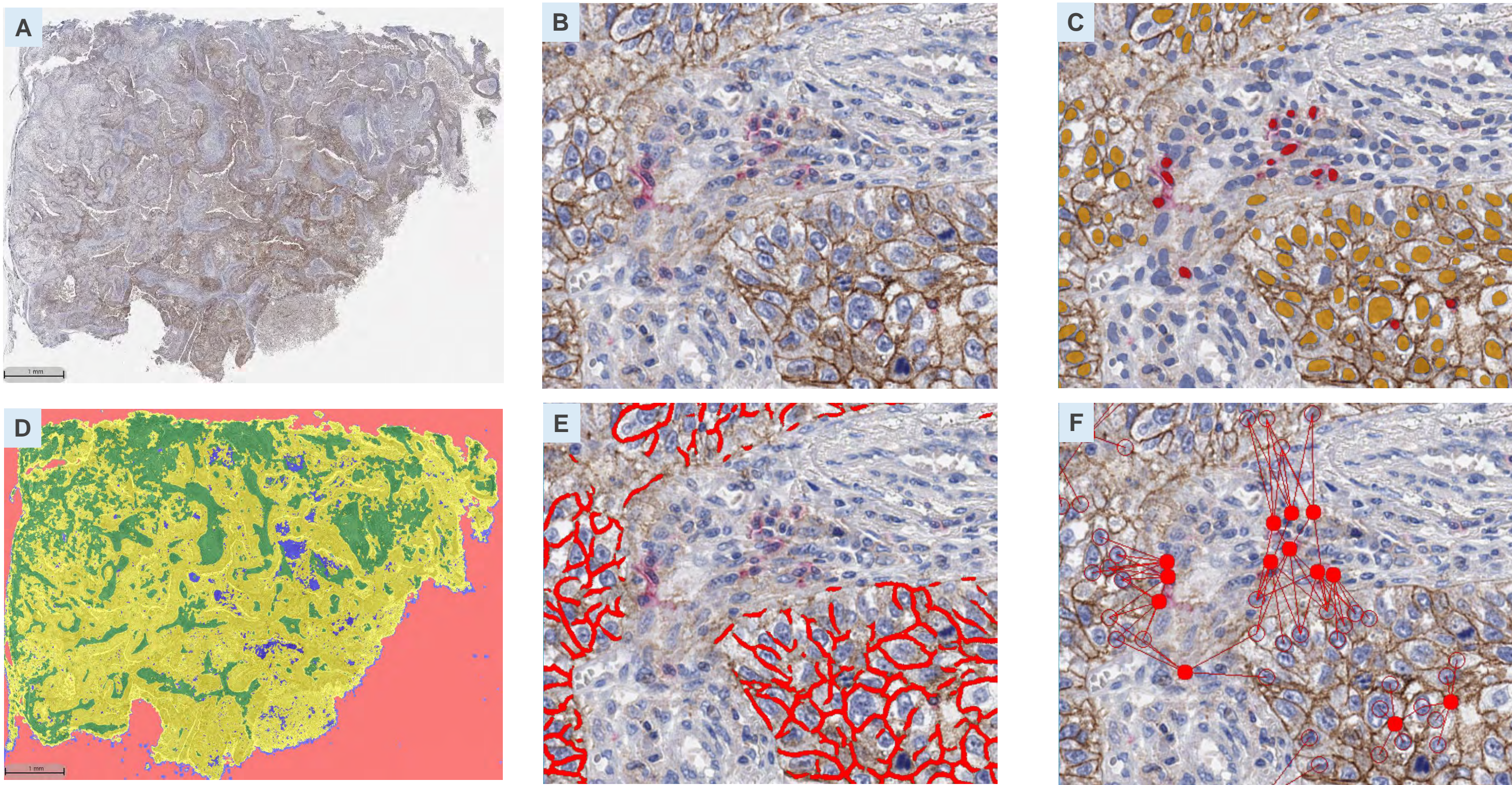
Figure 2: A: Dual IHC for CD8 and 22C3/PD-L1, B: Close-up of a region from A, C: Cell classification overlay (blue: negative, pink: CD8, beige: PD-L1), D: Region classification overlay from Indica HALO® (red: glass, blue: necrosis, green: stroma, yellow: tumor), E: PD-L1 stain mask overlay, F: Spatial connectivity graph overlay between CD8 cells (solid) and the five closest PD-L1 tumor cells (hollow)

Various data visualization techniques may aid in the comparison between samples or regions with regards to their cell-type densities, total cell counts for CD8 and PD-L1, and spatial relations between CD8 and PD-L1 cells.
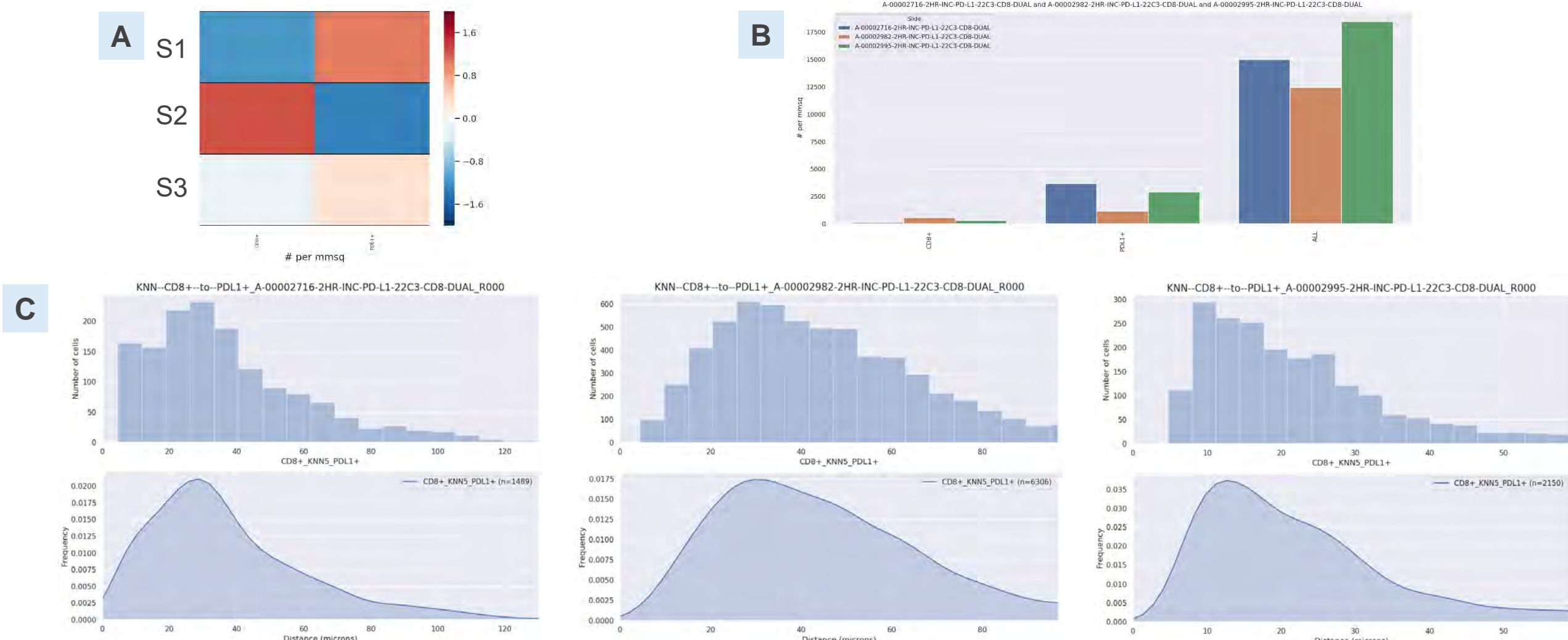
Figure 3: A: Heat map comparing CD8 and PD-L1 densities of three samples, B: Bar plot comparing cell counts for CD8 and PD-L1 for three samples, C: K-nearest neighbor analysis of CD8 to PD-L1 proximity in three samples

Tabular Data Output

Two levels of tabular data help explain the visual data within the overlays: one at the cell level, and one at the slide level, each useful in answering different types of questions about the samples. The cell-level data is useful for answering complex spatial questions, whereas the slide-level data is useful for high-level sample comparison.

Table 1: A: Cell-level Quantification Table, B: Slide-level Summary Table with a tab for the whole slide, and a tab for just the tumor region

Summary

NeoGenomics has developed a new Dual IHC assay for CD8 and 22C3/PD-L1, followed by subsequent advanced spatial analysis using state-of-the-art deep learning methods, and leveraging Indica's powerful and easy-to-use HALO® software. This workflow may be used to compare various patient cohorts for exploratory research to help better guide future research questions and directions.

The final assay and analysis output consists of:

- Dual IHC with CD8 (pink) and 22C3/PD-L1 (brown)
- Nuclear segmentation map with unique labels for each cell
- Region segmentation and classification from Indica HALO®
- Cell classification overlays for CD8 and PD-L1
- Tabular data at the cell level including stain intensity values, cell locations and morphologies, and cell classifications
- Tabular data at the slide level including region areas, cell counts and densities, and H-scores
- Various data visualization overlays to help better explain final results
- Graphs, plots, and charts to drive home major study findings

References

1. Feng, Zipei, et al. "Multiparametric immune profiling in HPV–oral squamous cell cancer." JCI insight 2.14 (2017).
2. Steele, Keith E., et al. "Measuring multiple parameters of CD8+ tumor-infiltrating lymphocytes in human cancers by image analysis." Journal for immunotherapy of cancer 6.1 (2018): 20.
3. Nagy, Mate Levente, et al. "Quantitative image profiling of the tumor microenvironment on double stained immunohistochemistry images using deep learning." (2019): e14619-e14619.
4. https://indicalab.com/halo/
5. He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.
6. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.
7. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

For any questions, please email mate.nagy@neogenomics.com