Poster 2407

Classification and enrichment of carcinoma cells in real-time based on high-dimensional morphology enables increased sensitivity of molecular and cytology analysis

Anastasia Mavropoulos[1], Janifer Cruz[1], Jordan Nieto[1], Kevin B. Jacobs[1], Kiran Saini[1], Amy Wong-Thai[1], Wes Austin[1], Julie Kim[1], Jeanette Mei[1], Weibo Yu[2], Vivian Lu[1], Chassidy Johnson[1], Stephane C. Boutet[1], Thomas D. Lee[2], Nianzhen Li[1], Jianyu Rao[2], Mahyar Salek[1], Maddison (Mahdokht) Masaeli[1]

[1] Deepcell Inc., Menlo Park, CA | [2] University of California, Los Angeles (UCLA), Los Angeles, CA

Introduction

Morphology is a fundamental cell property associated with cell identity, state, function, and disease. Current methods to quantify morphology lack high discriminative power and are generally destructive to cells, limiting data that can be extracted from samples. Here, we applied artificial intelligence (AI) and computer vision to characterize cells in malignant effusion fluids, which represent an accessible tumor source for cytopathology characterization and biomarker discovery. Using the Deepcell platform, we performed real-time deep learning interpretation of high-content morphology data at single-cell resolution and sorted classified carcinoma cells for validation using molecular assays. This application of deep learning on brightfield cell images enabled multi-dimensional morphological characterization of cells present in malignant effusion fluid samples.

UMAP clusters were associated with quantitative computer vision morphometrics linked to carcinoma cell morphology, including cell size, shape, and texture features. To validate the AI model predictions, carcinoma cells were sorted for collection and further characterization using copy number variation (CNV), targeted sequencing, and scRNA-Seq analyses. Results verified accurate carcinoma predictions by the AI model and enrichment of carcinoma cells, as demonstrated by identified mutations that were previously undetectable in whole patient samples. This combination of label-free imaging, deep learning, quantitative computer vision morphometrics, and cell sorting helps leverage potentially actionable morphological information in tumor cells.

Deepcell Platform

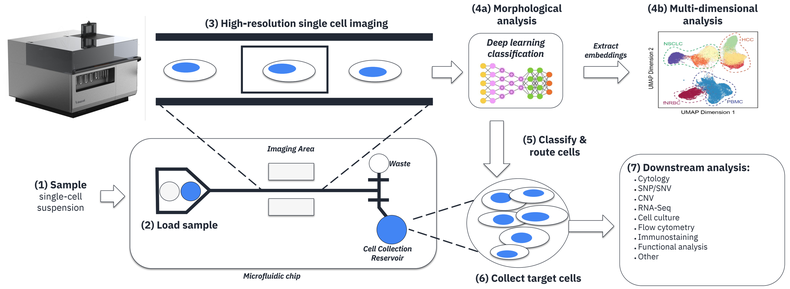

Figure 1. Deepcell workflow. (1) Single-cell sample suspension is **(2)** loaded onto a microfluidic chip and **(3)** as each cell flows past the imaging area, high resolution bright-field images are captured. Images are used by the **(4a)** deep learning AI model for multi-dimensional morphological analysis and/or **(5)** real-time classification and cell sorting. The sorted target cells can be **(6)** collected and processed for **(7)** downstream analysis. **(4b)** The image embeddings can be extracted (in addition or independently to sorting) and analyzed to generate multi-dimensional morphological profiles that can be visualized by UMAP and used to discover morphologically heterogeneous cell populations.

Results

Malignant Effusion AI Model Training & Validation

A

Model Training Set		
Cell Type	**# Samples**	**# Images**
Breast Carcinoma	4*	735,194
CD45+ (Leukocytes + others)	3	535,480
CD45- (Mesothelial + others)	2	129,124
Histiocytes	7	1,198,112
Total Training Sets	**16**	**2,597,910**

* 2 breast carcinoma specimens + 2 breast cell lines = 4

B

Model Validation Set		
Cell Type	**# Samples**	**# Images**
Breast Carcinoma	1	9,212
CD45+ (Leukocytes + others)	14	383,616
CD45- (Mesothelial + others)	15	69,184
Histiocytes	13	655,884
Breast Carcinoma Cell Lines	2	584,416
Mesothelial Cell Lines	1	328,712
Total Validation Sets	**46**	**2,031,024**

C

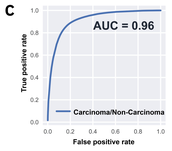

Figure 2. Deepcell malignant effusion AI Model training and validation. FACs-sorted cells from effusion samples were used to capture high resolution bright-field images for malignant effusion AI model training and validation. **(A)** Summary table of biological replicates (patient samples and cell lines) and number of images used for model training and **(B)** *in silico* model validation. **(C)** The performance of the model to identify carcinoma cells from malignant effusion samples was evaluated by *in silico* validation and the area under the curve **(AUC) is 0.96**.

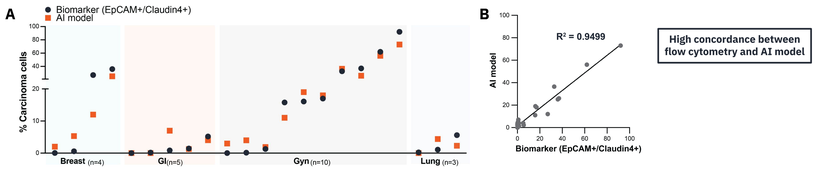

Figure 3. Malignant effusion AI model performance on clinical samples. (A) Flow cytometry (EpCAM+/Caludin4+; gray) compared to AI model (orange) predictions of carcinoma cell abundance in patient-derived breast, GI, gynecological, and lung cancer effusion samples. **(B)** Performance of AI model is highly concordant with biomarker-base flow cytometry assessments, with R-squared correlation value determined as 0.949.

Multi-dimensional Morphology Analysis

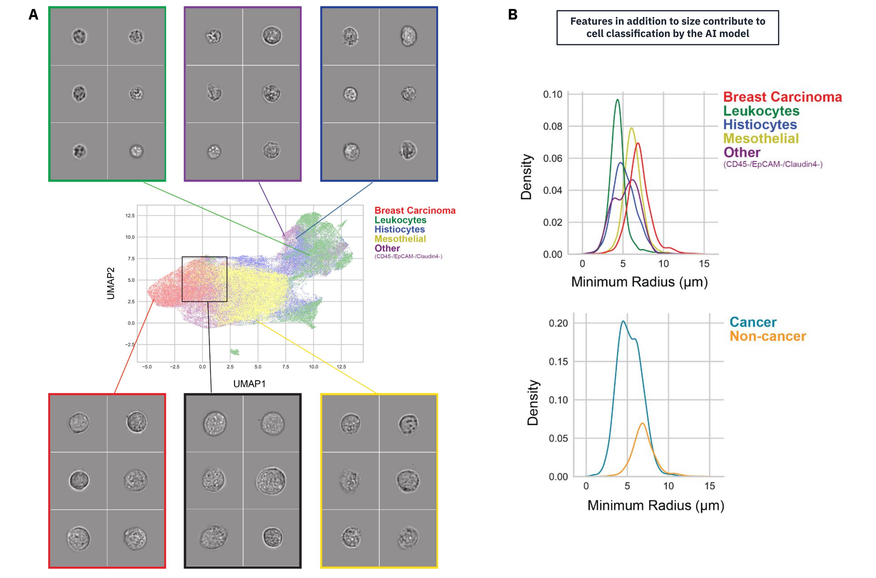

Figure 4. UMAP of AI embeddings extracted from high-content images of the validation dataset. (A) FACS purified cells from clinical samples were imaged on the Deepcell platform and plotted by UMAP (labels indicated). Clustering patterns are based on morphology information detected by the AI model, with cells (dots) clustering closer being more visually and morphologically similar. Annotation of cell types are based on pure populations obtained by flow cytometry "ground truth". Shared morphologies between mesothelial and carcinoma cells are historically challenging for cytologists to distinguish and a common cause of atypia/equivocal diagnosis, and are shown here to have overlapping morphological features as detected by AI. **(B)** Deep learning embeddings are augmented with computer vision derived features and cell radius results are shown to provide insight into morphological differences between cell classes.

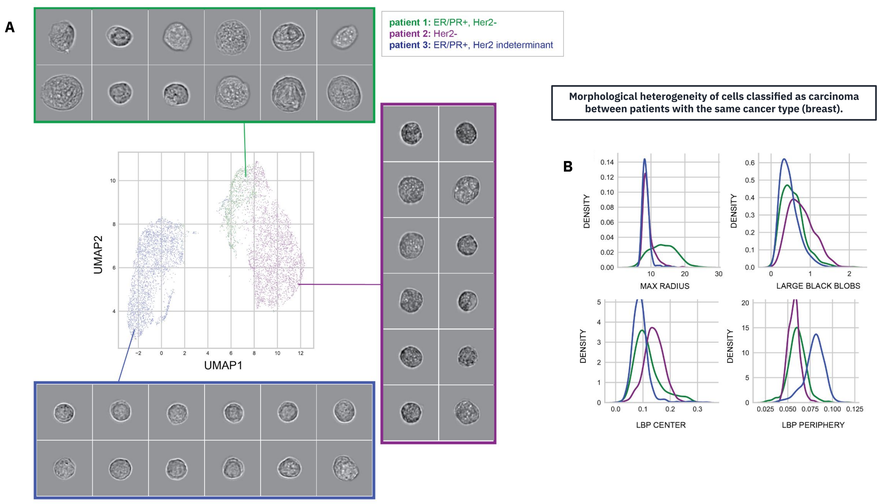

Figure 5. UMAP of carcinoma cells from 3 different breast cancer patient effusion samples. (A) UMAP plot of AI model carcinoma cell predictions derived from captured images of unlabeled effusion samples from 3 different breast cancer patients. **(B)** Density histogram plots by patient, as quantified by computer vision morphometric measures of texture (LBP periphery, LBP center, and large black blobs) and size (radius).

Validation of AI Predictions with Molecular Assays

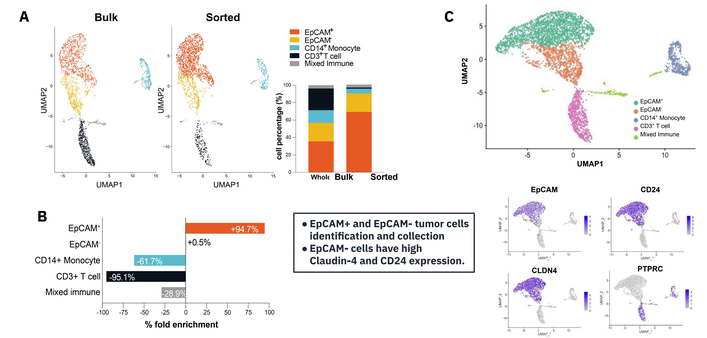

Figure 6. scRNA-seq analysis validates cell populations of interest (carcinoma cells) in sorted ovarian cancer effusion sample. (A) Annotation of different cell populations was conducted on aggregated datasets before and after imaging and sorting on the Deepcell platform. 5 cell populations were identified: EpCAM+ and EpCAM- tumor cells and monocyte, T cell, and other immune cell populations. **(B)** Cell type and respective percent of the total cell population for pre- and post-sorted sample. **(C)** Pseudo-color gene expression level of EpCAM, CLDN4 (Claudin-4), CD24 (stem cell marker), and PTPRC (CD45) expression demonstrating that EpCAM- cells express epithelial marker (Claudin-4) and not lymphocyte marker (PTPRC; CD45) and are epithelial cells.

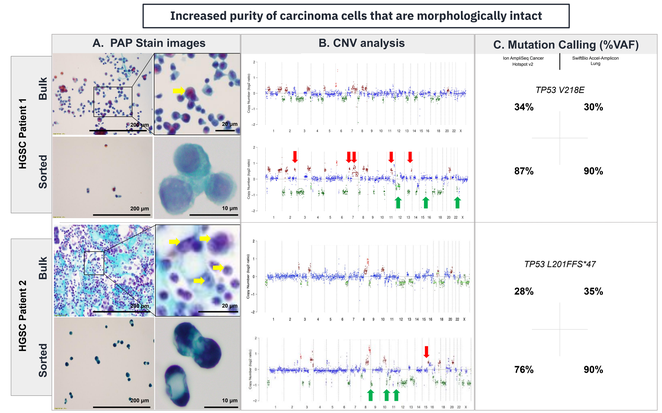

Figure 7. Case studies of two sorted ovarian HGSC body fluid samples. To further validate the AI predictions: **(A)** PAP stain was performed on pre- and post-sorted patient samples. Higher purity of malignant cells was observed in sorted samples, as confirmed by a board certified pathologist. Cells retained intact architecture and showed higher malignant cell purity compared to bulk patient samples. Scale bars as indicated. **(B)** WGS followed by CNV analysis was performed and increased amplitude of deletions and amplifications were observed, including some identified in sorted cells but not in whole samples (arrows: amplifications, red; deletions, green). **(C)** Targeted sequencing identified two somatic p53 mutations, with respective allele frequencies.

Conclusions

- We developed, trained, and validated an AI model to classify carcinoma cells from malignant effusions and used computer vision to add biological explainability to morphological differences in patient samples, and validated carcinoma predictions with downstream molecular assays.
- Model performance was tested on clinical samples and carcinoma frequency was concordant to flow cytometry results.
- Shared and differing morphologies of cells residing in malignant fluids are captured by AI and are consistent with known morphological heterogeneity observed by cytologists. Cell size distributions are similar between cancer and non-cancer cells, suggesting other features contribute to morphological heterogeneity.
- AI predictions of carcinoma cells from malignant effusions were validated by scRNA-Seq, CNV profiling, targeted mutation analysis, and cytological analysis.