A platform for high-resolution morphology analysis reveals tumor heterogeneity and enables label-free enrichment of target subpopulations

Andreja Jovic[1], Kiran Saini[1], Tiffine Pham[1], Jeanette Mei[1], Michael Phelan[1], Vivian Lu[1], Stephane Boutet[1], Kevin Jacobs[1], Julie Kim[1], Chassidy Johnson[1], Nianzhen Li[1], Mahyar Salek[1], Maddison (Mahdokht) Masaeli[1]

[1]Deepcell Inc. Menlo Park, CA | Contact: jovic@deepcellbio.com

→ Introduction

Methods to study cancer at the single-cell level and capture tumor heterogeneity are crucial to better understand and treat cancer. However, current detection of tumor cells from tissue typically relies on targeting biomarkers (e.g. EpCAM), resulting in isolation of specific tumor cell population subsets used for further molecular and functional analysis. The Deepcell platform performs multi-dimensional morphology analysis of single-cells using deep learning on high-resolution brightfield images captured in microfluidic flow to phenotype and route label-free target cells for collection in real-time. This approach eliminates biomarker-specified isolation of tumor cells and adds deep interpretation of morphology as a dimension for more comprehensive understanding tumor heterogeneity.

In this study, we developed a deep convolutional neural network AI model to identify and route malignant cells from non-small cell lung cancer (NSCLC) dissociated tumor cell (DTC) samples for retrieval. The routed NSCLC cells are label-free, unperturbed, and intact, making them amenable to diverse downstream analyses without the need for sample pre-processing, biomarkers, gating, or bioinformatics capabilities. Further, captured cell images were used to generate multi-dimensional morphological profiles visualized by UMAP to uncover morphologically heterogeneous tumor cell populations that were subsequently retrieved for further molecular analysis. Further work is planned to evaluate the link between the morphological, molecular, and functional characteristics of each subpopulation. Together, the Deepcell platform enables discovery and characterization of cellular phenotypes and integration of morpholomics into other molecular -omics.

→ Deepcell Platform

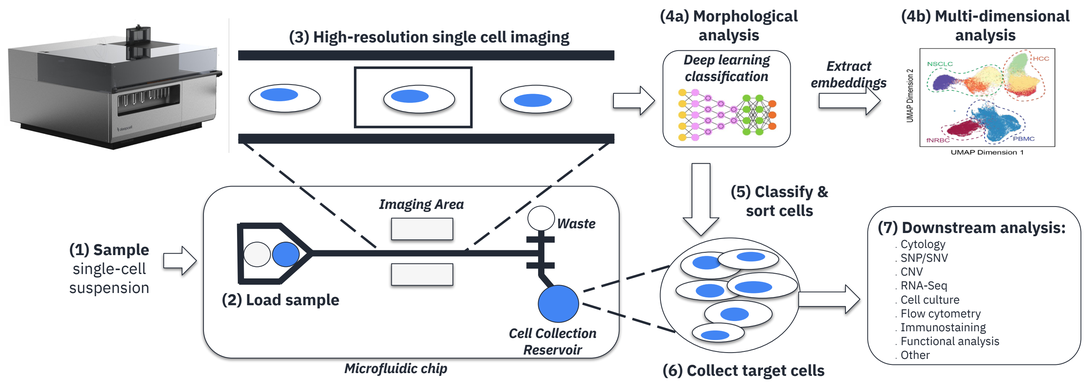
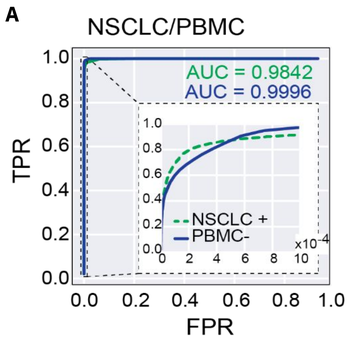


Figure 1. Deepcell workflow. (1) Single-cell sample suspension is **(2)** loaded onto a microfluidic chip and **(3)** as each cell flows past the imaging area, high resolution bright-field images are captured. Images are used by the **(4a)** deep learning AI model for multi-dimensional morphological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

→ Results

NSCLC Tissue AI Model Identifies and Enriches Tumor Cells Based on Morphology

A

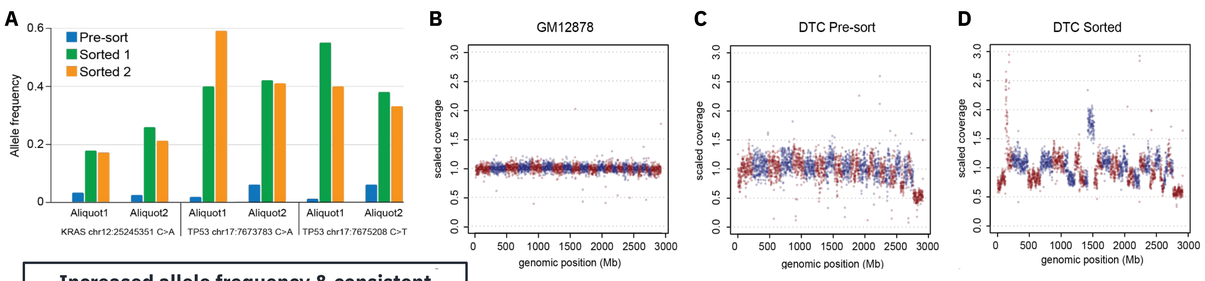


Validation of malignant effusion AI classifier shows high accuracy and concordance

B

DTC sample	Fraction of Malignant Cells		
	scRNA-seq (tumor markers)	FACS (EpCAM+)	AI Classifier (Deepcell)
1	7	1	2
2	9	11	7
3	9	3	10
4	N/D	8	6
5	12	2	16
6	27	11	14
7	N/D	13	16
8	36	22	28
9	N/D	40	27
10	59	47	50

Figure 2. Deepcell NSCLC Tissue AI Model performance.. Pure cell populations are processed on the Deepcell instrument to capture high resolution bright-field images used for AI model training and validation. **(A)** We applied the instrument and NSCLC Tissue AI model to recognize and sort for malignant cells in NSCLC dissociated tissue (DTC) samples. Model performance was evaluated *In silico* and the area under the curve **(AUC) is 0.98. (B)** Close agreement in % tumor cells between scRNA-seq, flow cytometry, and NSCLC Tissue AI Model for 10 DTC samples



Increased allele frequency & consistent capture of mutational heterogeneity in sorted vs. pre-sort cells

Increased sensitivity of CNV in DTC sorted samples vs. pre-sorted

Figure 3. Deepcell platform identifies and enriches for tumor cells from NSCLC DTCs. (A) Mutations identified in patient samples (KRAS, TP53) for two sorting runs with two aliquots reported. **(B-D)** CNV analysis was performed on **(B)** GM12878 cell line (baseline control), **(C)** DTC pre-sorted **(D)** and DTC sorted cells.

Multi-dimensional Morphology Analysis Correlates with Gene Expression Profiles

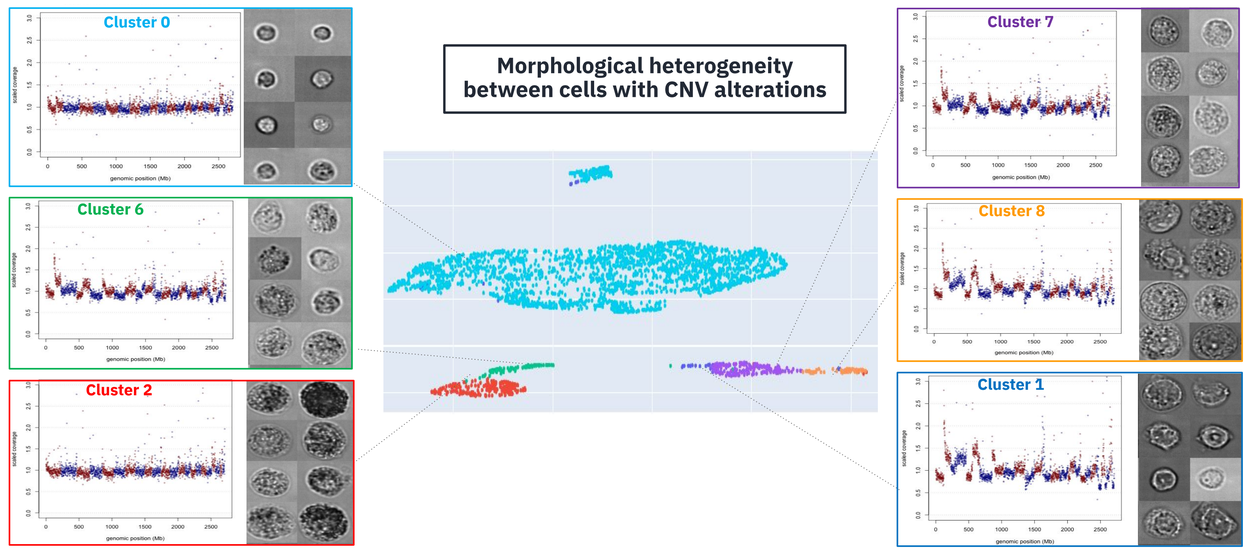


Figure 4. Morphologically distinct clusters of cells from NSCLC tissue exhibit unique CNV patterns and correlate with gene expression profiles. (A) Six morphologically distinct clusters (0, 1, 2, 6, 7, 8) were identified by manual identification. The NSCLC Tissue AI Model was trained to identify and sort for them and CNV analysis revealed unique patterns for each cluster.

Bulk RNA-seq and distinct morphology-based clusters

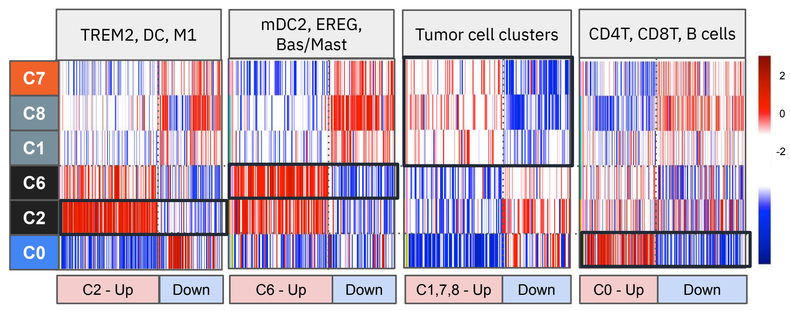
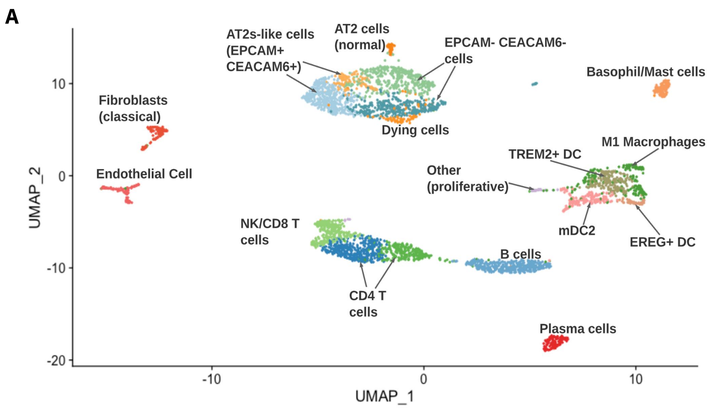


Figure 5. Bulk RNA-seq uncovers the identity of the 6 morphologically distinct clusters. Bulk RNA-seq data identified 4 main cell classes: 3 immune cell populations and 1 tumor cell population. Heatmaps summarize gene signature results associated with the indicated cell types and the respective gene expression profiles for each cluster. Clusters 1, 7, 8 express genes associated with tumor cells; clusters 2 and 6 express genes associated with Macrophages/DCs/Mast/Basophils/EREGs; and Cluster 0 expresses genes associated with T and B cells. Cluster 6 has a gene expression profile corresponding to DCs, EREG and basophils/mast cells. To gain insight into the specific cell compositions of each cluster, we used scRNA-Seq analysis.

A



scRNA-seq annotation and identification of cell classes including EpCAM+ and EpCAM- malignant cells

B

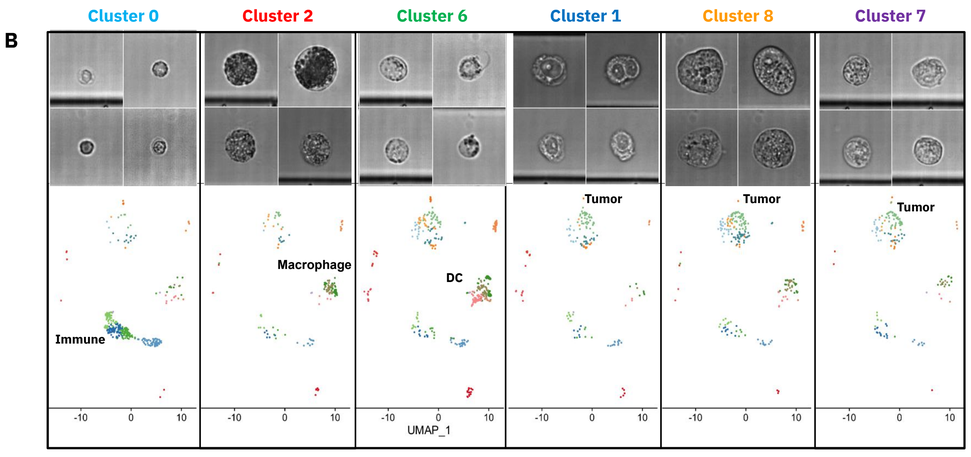


Figure 6. scRNA-seq analysis was performed to characterize the different cell populations covered in each of the morphology clusters. (A) Annotation of the different cell populations was conducted on the aggregated datasets of the cell populations before and after enrichment of the morphology clusters on the Deepcell platform. Epithelial lung cell populations were identified in 5 different clusters: 4 different malignant cell types defined as AT2s-like EPCAM+/CEACAM6+ and EPCAM-/CEACAM6- and one non-malignant AT2 cell population. Additionally, immune cells (NK, CD8/CD4 T, B, plasma, mDC2, TREM2+, EREG+, Mast cells, Basophils and M1 macrophages), fibroblast and endothelial cells were identified. **(B)** Representative images characterizing each morphology cluster are shown and association with the different gene expression cluster identified with scRNA-seq analysis.

Differential expression analysis on scRNA-seq correlates with bulk RNA-seq and associates well with morphology and CNV results identifying normal (immune cells and macrophages) and tumor cells for morphology clusters 0, 2, 1, 7 and 8, respectively. Further work is underway to understand morphology cluster 6.

The Deepcell platform and deep learning enable a new approach to study tumor heterogeneity and integrate cell morphology into single cell multi-omics.

→ Conclusions

- The Deepcell platform images single cells and generates multi-dimensional morphology data based on high resolution brightfield imaging.
- Using the multi-dimensional morphology data obtained on the Deepcell instrument, we trained an AI classifier that identifies and isolates immune and tumor cells from NSCLC samples that are label-free.
- We demonstrated the enrichment of malignant cells from NSCLC tumor tissue using CNV, targeted mutation analysis, bulk RNA-seq, and scRNA-seq analysis. Morphology clusters corresponding to tumor cells displayed increased number of CNV alterations and gene expression profile associated with lung tumorigenesis.
- Deepcell's multi-dimensional data defines distinct morphologic clusters that represent distinct cell morphologies and CNV profiles.
- scRNA-seq and bulk RNA-seq data correlate well with morphology and CNV results indicating that clusters 1, 7 and 8 are tumor cell populations and clusters 0 and 2 are immune cells. Further work is underway to understand cluster 6.
- These results demonstrate the ability to cluster cells with similar gene expression profiles based on morphology alone and highlight the potential power of the technology and cluster-based sorting for capturing new molecular insights in a label-free manner.

deepcell