Poster 2401

Multi-dimensional morphology analysis enables identification and label-free enrichment of heterogeneous tumor cell populations

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

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

Introduction

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 sort 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 a more comprehensive understanding of tumor heterogeneity.

We developed a deep convolutional neural network AI model to identify and sort malignant cells from non-small cell lung cancer (NSCLC) dissociated tumor cell (DTC) samples for retrieval. The sorted 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. 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.

Deepcell Platform

Deepcell workflow

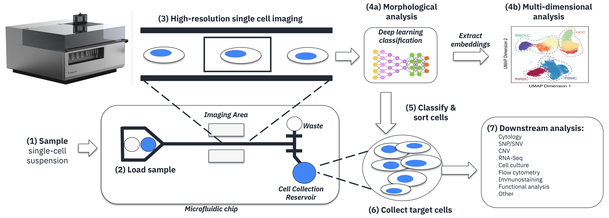


Figure 1. Deepcell workflow. (1) Single-cell sample suspension is **(2)** loaded onto a microfluidic chip and **(3)** as each cell flows past the imaging area, high resolution bright-field images are captured. Images are used by the **(4a)** deep learning AI model for multi-dimensional morphological analysis and/or **(5)** real-time classification and cell routing. The target cells sorted 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.

NSCLC Tissue AI Model Training and Validation

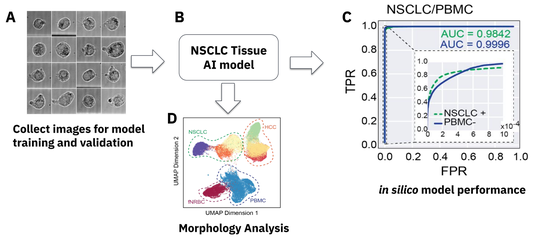


Figure 2. Deepcell NSCLC Tissue AI Model. (A) Pure cell populations are processed on the Deepcell instrument to capture high resolution bright-field images used for AI model training and validation. **(B)** We applied the instrument and NSCLC Tissue AI Model to recognize and sort for malignant cells in NSCLC dissociated tissue (DTC) samples. 21 cell types were used to train the model. **(C)** Model performance was evaluated *in silico* and the area under the curve **(AUC) is 0.98. (D)** The image embeddings can also be extracted and used for UMAP analysis (**Figures 4-5**).

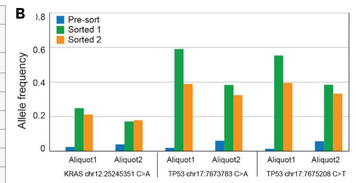
Results

Multi-dimensional Morphology Analysis Correlates with Mutation and Gene Expression

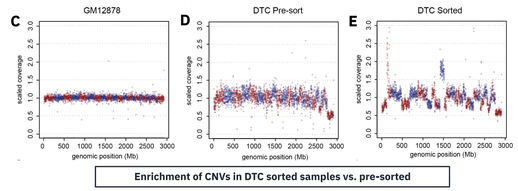
A

DTC sample	Fraction of Malignant Cells: scRNA-seq (tumor markers)	Fraction of Malignant Cells: 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

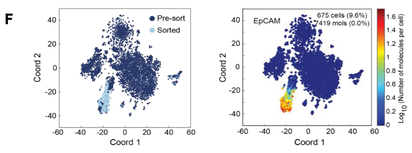
Validation of NSCLC AI classifier shows high accuracy and concordance



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



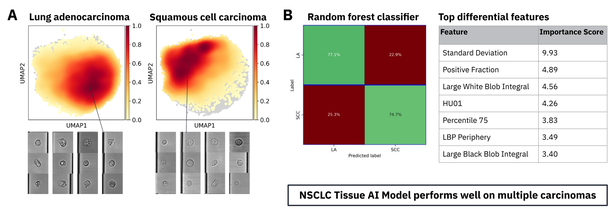
Enrichment of CNVs in DTC sorted samples vs. pre-sorted



scRNA-seq confirms sorted tumor cells overlap with high EpCAM expression.

Figure 3. Deepcell platform identifies and enriches for tumor cells from NSCLC DTCs. (A) Close agreement in % tumor cells between scRNA-seq, flow cytometry, and NSCLC Tissue AI Model for 10 DTC samples **(B)** Mutations identified in patient samples (KRAS, TP53) for two sorting runs with two aliquots reported. **(C-E)** CNV analysis was performed on **(C)** GM12878 cell line (baseline control), **(D)** DTC pre-sorted **(E)** and DTC sorted cells. **(F)** scRNA-seq analysis showed the sorted tumor cells (light blue) had strong overlap with cells expressing the highest levels of EpCAM.

NSCLC Tissue AI Model Accurately Predicts 2 Lung Carcinoma Types Based on Morphology



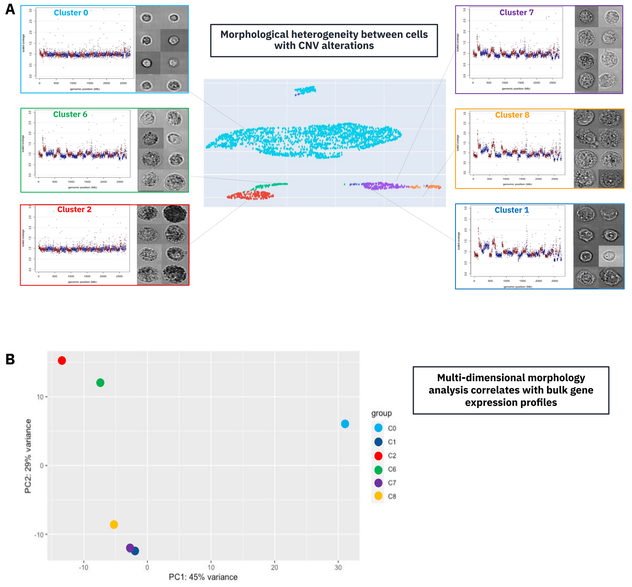
Top differential features

Feature	Importance Score
Standard Deviation	9.93
Positive Fraction	4.89
Large White Blob Integral	4.56
HU01	4.26
Percentile 75	3.83
LBP Periphery	3.49
Large Black Blob Integral	3.40

NSCLC Tissue AI Model performs well on multiple carcinomas

Figure 4. Deepcell AI classifier can distinguish different types of lung carcinomas. (A) Dissociated tumor biopsies from 17 lung adenocarcinomas (LA) and 12 squamous cell carcinoma (SCC) cells were run on the Deepcell platform, and UMAPs of the resulting analysis show overlapping but distinct clusters for LA vs SCC morphologies. Representative images of cells from each cluster are shown. **(B)** A random forest classifier can predict LA vs SCC types with ~75% accuracy, as shown by the confusion matrix. The top morphological features distinguishing the two carcinomas are shown in the table.

Multi-dimensional Morphology Analysis Correlates with Molecular Profiles



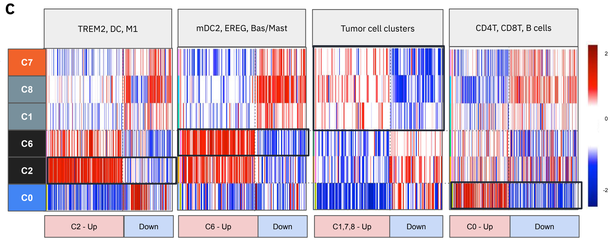


Figure 5. Morphologically distinct clusters of cells from NSCLC tissue exhibit unique CNV patterns and correlate with known gene expression profiles. (A) Six morphologically distinct clusters (0, 1, 2, 6, 7, 8) were identified by manual identification. The NSCLC Tissue AI Model was trained to identify and sort for them and CNV analysis revealed unique patterns for each cluster. **(B)** Bulk RNA-seq analysis followed by principal component analysis (PCA) was performed on each of the clusters. Morphology clusters 2 and 6 and morphology clusters 1, 7 and 8 localize together by gene expression profiles. While morphology clusters 1, 6, 7, and 8 have similar CNV profiles, cluster 6 is morphologically and transcriptionally most similar to cluster 2. **(C)** The 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 malignant 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 mixed gene expression profile corresponding to DCs, EREG and basophils/mast cells.

scRNA-seq Annotation & Identification of Cells Including EpCAM+ and EpCAM- Malignant Cells

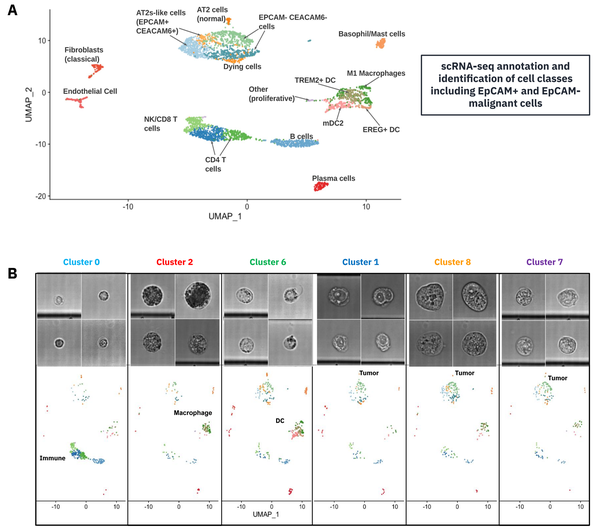


Figure 6. scRNA-seq analysis was performed to characterize the different cell populations covered in each of the morphology clusters. (A) Annotation of the different cell populations was conducted on the aggregated datasets of the cell populations before and after enrichment of the morphology clusters 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.

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

Conclusions

- The Deepcell platform images single cells and generates multi-dimensional morphology data based on label-free imaging.
- 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.
- The NSCLC Tissue AI Model can accurately predict lung adenocarcinoma vs squamous cell carcinoma cells.
- 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.
- 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.