Linking Cell Morphology to Cell Identity Using Machine Learning

Sydney Alderfer[1], Jiangyu Sun[2], Lubna Tahtamouni[2], and Ashok Prasad[1,3]

***Short Abstract* — Cell morphology and cytoskeletal organization can provide great insight into cell health and disease states, as the cytoskeleton plays essential roles in countless cell processes, from cell division to migration to signaling. In cancer cells, cytoskeletal dynamics, cytoskeletal filament organization, and overall cell morphology are known to be altered substantially. This project used a small fluorescence microscopy image dataset of retinal pigment epithelial (RPE) cells to investigate the effectiveness of convolutional neural networks (CNNs) to distinguish between normal and oncogenically transformed cells and between different subtypes of transformed cells of the same cell line. We found that cell morphology was a sensitive signature of cell identity, and thus could be a very useful method for assaying cell phenotype.**

***Keywords* — Morphological screening, single cell imaging, machine learning**

I. PURPOSE

Our central hypothesis is that cell shape and phenotypic state of the cell are linked, and thus useful information regarding phenotype of the cell is buried in the cell shape [1]. One first step towards deciphering the information hidden in cellular morphology is to estimate how well morphology can distinguish between different cellular phenotypes, as well as the same cells in different experimental conditions.

Machine learning methods are a useful technique to use for answering this question. Suitably complex machine learning architectures can work as universal approximators and help us find the underlying unknown function that links morphology to cellular identity and phenotype.

In this work we used Convolutional Neural Networks with transfer learning on a database of cellular images collected by our group.

II. MATERIALS AND METHODS

We used retinal pigment epithelium cells (ARPE-19) that were transformed into cancerous cells using Ras, MEF or Akt oncogenic mutations. We also collected images of three triple negative breast cancer cell lines. Cells were cultured and imaged as described previously [2].

III. USING CONVOLUTIONAL NEURAL NETS

In recent years convolutional neural nets (CNNs) have emerged as an important machine learning technique for analyzing images. CNNs take the image data itself as the input training data and therefore do not need the images to be preprocessed for feature extraction, but instead automatically recognizes features hierarchically. They have also proved useful in feature identification in complex data. However, CNNs typically require significant data to be trained. A technique called transfer learning seeks to overcome this by re-purposing CNNs trained on large datasets to perform classification tasks on new datasets.

Here we modified ResNet-50 [3], a 50-layer deep CNN that was pre-trained on the ImageNet database. Model performance was evaluated using 10-fold cross validation. The data was divided into a training set and a validation set representing 90% and 10% of the data, respectively. The process was repeated 10 times, using a 10% holdout. The validation accuracies for each of the ten folds were averaged to produce an overall estimated accuracy score for the model.

IV. RESULTS AND CONCLUSION

We found that the CNN could accurately identify cancer cells from the parental cell line with an accuracy of about 95% or better at the single cell level, significantly better than with artificial neural networks (ANNs) [2]. Furthermore, CNNs could not only accurately distinguish each transformed cell line from the parent cell line but also from the other transformed cell lines, indicating that each single oncogenic mutation produces a unique signature in actin morphology [4].

REFERENCES

[1] Prasad, Ashok and Alizadeh, Elaheh, Cell form and function: interpreting and controlling the shape of adherent cells, Trends in Biotechnology, 37(4), 347-357, April 01, (2019)

[2] Elaheh Alizadeh, Jordan Castle, Analia Quirk, Cameron Taylor, Wenlong Xu, Ashok Prasad. Cellular morphological features are predictive markers of cancer cell state. Computers in Biology and Medicine, Nov;126:104044 (2020).

[3] Kaiming He et. al. Deep Residual Learning for Image Recognition, arXiv:1512.03385v1

[4] Sydney Alderfer, Jiangyu Sun, Lubna Tahtamouni, Ashok Prasad *In Preparation.*

Acknowledgements: This work was partially funded by an OVPR Research Continuity Grant, 2021-22..

[1]Department of Chemical and Biological Engineering and The School of Biomedical Engineering, Colorado State University, Fort Collins, CO

[2]Department of Biochemistry, Colorado State University, Fort Collins CO

[3]Corresponding author. Email: ashok.prasad@colostate.edu