INFERRING STEM CELL PROTEIN EXPRESSION FROM DYNAMIC COLONY MORPHOLOGY USING MACHINE LEARNING ALGORITHMS

[1]A.Witmer, [2]B. Bhanu, [3]P. Talbot
[1]Department of Bioengineering, University of California, Riverside
[2]Department of Electrical and Computer Engineering, University of California, Riverside
[3]Department of Molecular and Systems Biology, University of California, Riverside
Contact: *adam.witmer@email.ucr.edu*

Introduction: The following abstract details collaborative works between the departments of Bioengineering, Electrical and Computer Engineering, and Molecular and Systems Biology, as well as a proposal for the continuation of the Ph.D thesis of Adam Witmer. These works center around the design of machine learning/ computer vision algorithms for the automated analysis of video microscopy data collected during stem cell toxicology experimentation. The goal of this work is to use deep convolutional neural networks (CCN) to determine the effects of cigarette and cigarette alternatives on human embryonic stem cell growth by modeling dynamic morphological features of stem cell colonies under toxic exposure. Motivation for this project stems from the need for unbiased, standardized quality control and analysis methods for stem cell projects involving live cellular microscopy. The need to characterize dynamic properties of colony development prohibits the use of invasive data collection techniques (i.e. fluorescent staining, genomic/proteomic profiling, flow-cytometry) via sacrifice of cell colonies. To date, two projects have been submitted as conference publications that address two problems relating to this issue: 1. classification of heterogeneous colony populations into homogeneous subclasses, and; 2. generation of synthetic data to supplement indispensable experimental data points. These projects, summarized below with supporting evidence, have provided insights into both the effects of toxicants on diseased stem cell colonies and the design of experimental protocol for the successful completion of the proposed projects.

Materials and Methods: Data for these projects comes from the laboratory of Dr. Prue Talbot in the form of time-lapse, phase contrast microscopy videos of induced pluripotent stem cells (iPSC) expressing the Huntington's disease phenotype. This data was collected to test the hypothesis that exposure to nicotine during culture has a neurogenic/-protective effect on the development of diseased iPSC development. It is noted that upon visual observation of colony growth over a 48 hour period, that pluripotent colonies begin to change towards downstream progenitor and differentiated lineages including neuron like formations. These projects address the need to automatically quantify changes that are difficult to determine using by-hand analysis methods.

Results and Discussion: For this purpose, a patch based classification method employing CNN is implemented to localize four morphologically homogeneous colony subtypes composing contiguous heterogeneous populations. Results suggest a positive correlation between concentration of nicotine exposure and overall increase in colony area classified as downstream lineages over time (Figure 1). Secondly, a synthetic dataset is generated using Generative Adversarial Networks (GAN) to supplement the number of real biological data points that must be irreplaceably used to train the CNN. It is noted that the addition of 500 synthetic images to a limited real dataset increases the confidence of a the classifier in its probabilistic predictions by 2% as well as slightly increasing the true positive rate of classification of real images.

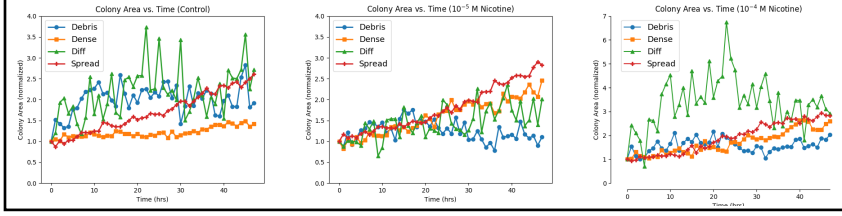

Figure 1: Growth curves of colony area show a positive correlation between nicotine exposure and downstream lineage differentiation, as well as a reduction in debris colony formation, suggesting that nicotine has both protective and neurogenic effects in this disease model

Conclusions: Findings have proven the efficacy of CNN for the analysis of stem cell toxicology experiments and have influenced the synergistic design of cell culture and model based experimental protocols for the continuation of the thesis work. The goal of the proposed work is to track colony development and protein expression using only live cell colony morphology that is validated via proteomic profiling and immunocytochemistry. This work will be used to uncover the effects of cigarette toxins on underlying stem cell biology, and can be expanded to include other experimental protocols involving stem cells such as differentiation or therapeutic drug testing.