Topic: The use of deep learning in cell line authorisation for modern drug discovery.

**Highlights:**

- Novel deep learning architectures for cell line authentication.
- Automated cell line quality control methods with spatial-temporal analysis.
- Comprehensive experimental work for the proposed approaches.

**Overview:**

Cross-contamination of human cell lines with other cell lines and misidentification of cell cultures are reported to be 18-36% [1], which costs the drug discovery industry 28 billion USD per year [1]. Contamination may be due to other human cell lines (i.e. intraspecies contaminant) or cell lines of different species (i.e. interspecies contaminant). Use of contaminated cell cultures plays a negative impact on medical exercises in terms of wasted financial resources and time. Presently, to handle the problem of cell line authentication, short tandem repeat (STR) profiling has been commonly used but lacks sufficient stability.

In this project, we propose to develop fully automated and trainable deep learning architectures for cell line authentication and to detect genetic & transcriptional evolution (GTE) using sequences of images over time. This is a multi-disciplinary research project, pulling together expertise in molecular biology and pharmacology, artificial intelligence and computer science at Leicester. The following specific objectives form the core activities of the proposed research programme:

**Objective 1:** For any cell type undergoing characterisation, cell images can be obtained at a given time. This enables the direct image comparison against the benchmark. Transfer learning with VGG16/18 and Inception net were validated for cell images [2]. We will focus on novel multi-task learning [3] to distil key phenotypic factors including cell morphology, cell count, cell cycle and group behaviours. We will test related cell lines, e.g. OVCAR 3 & OVCAR 4. All cell line samples and image data will be provided by our research collaborator AstraZeneca UK, a research-based business at Cambridge.

**Objective 2:** We propose to use the kinetic behaviour of cells over time. It is important to indicate potential GTEs and degradation over time. Kinetic cellular images impose the challenge of changes in cell counts, morphology and cell interactions. We propose to combine Bayesian inference and recurrent neural network (RNN) [4] to understand cell line behaviours via the analysis of a sequence of images. A deep neural network architecture (Objective 1) can be used to capture static phenotypes, where RNNs allow us to understand the history of cell behaviours. We will use A549 cells as a key model of DDR pathways and CRISPR that has shown phenotypic drifts [5].

The expected outcome of the work is a novel automated software tool for cell line authentication. In regard to the dissemination of our work, in the short term, we aim to publish at least three papers on top journals, such as IEEE Trans. On Medical Imaging or Medical Image Analysis, and three papers at top conferences such as CVPR and MICCAI. In the longer term, we may pursue possible commercialisation of the developed system to our research collaborator AstraZeneca, one of the top pharmaceutical companies in the world with the annual revenue of $21.0 billion in 2018.

**Further reading:**


Level: PhD

Supervisors: Dr. Huiyu Zhou, Dr. Yinhai Wang and Prof. Thomas Erlebach.

Subject areas: Biomedical engineering; data analysis; information science; genetics.

In case of any query, please contact Dr. Huiyu Zhou at: hz143@leicester.ac.uk.