Deep Learning of Cell Morphologies for Kinome Wide Screening

Mathias Engel[1,2], Xavier Robin[1], Jesper Ferkinghoff-Borg[1], James Longden[1], Rune Linding[1]

***Short Abstract* — Integrating individual and population based cell data, across mulitple experimental methods is crucial in gaining new biological understanding. We present a deep learning probabillity model that can handle the integration of different data types with predictive power. The model is deployed to understand resistance in a screen consisting of kinome-wide RNAi knockdowns throughout multiple cell lines and cancer therapies.**

***Keywords* — cell morphology, resistance, high content screening, kinome, deep learning architectures, RNAi.**

I. Introduction

Today the life science community has several large data producing methods at its disposal, such as next-generation sequencing, mass spectrometry and high content imaging. We can use these techniques to probe behavior and signalling of cells in a systematic and statistically robust manner. Furthermore we can now investigate these cell dynamics in vastly different length scales ranging from genes to full organisms. However although these techniques have led to extremely valuable insights they are often difficult to combine numerically in a predictive probabilistic model. Here we present a new method to handle data integration in combination with a large kinome wide screen aimed at understanding resistance.

II. Data Production

An Opera high content imaging system with cell:explorer robotics (PerkinElmer) was used to conduct multiple kinome wide RNAi knockdown screens. The full image data consists of 2.41 million images distributed on 17 cell lines with 11 anti-cancer therapies. The Opera microscope produces images each containing several hundreds of cells, while still maintaining sufficient resolution to extract cell-specific morphological information (textures, geometries, size). Cells were stained with Hoechst 33342 (nucleus) and Rhodamine phalloidin (F-actin).

Morphological features were extracted using Acapella image analysis software (PerkinElmer). This produced a feature vector containing thousands of entries per cell. This vector was stored in a database to facilitate easy access for the subsequent modeling work.

III. Purpose

It is already known that a connection exist between local signaling networks and morphological phenotypes (Bakal et. al., 2007)[1]. However, this study only aimed at discovering distinct subgroups of morphologies from the feature vectors.

In order to correlate and predict relations between morphologies and various prior information known from the experimental design (cell line, treatment, RNAi) in a bidirectional manner, a different model is needed. This model should be capable of combining uncertanties across different datatypes including both population and cell specfic information.

We will therefore exploit the properties of deep learning architectures (Hinton et. al., 2006)[2,3] to model the probability distribution of the morphological space across multiple conditions. All prior information as well as additional data (e.g. cell count, mRNA expression) can be directly included as input to the model through binary encoded vectors.

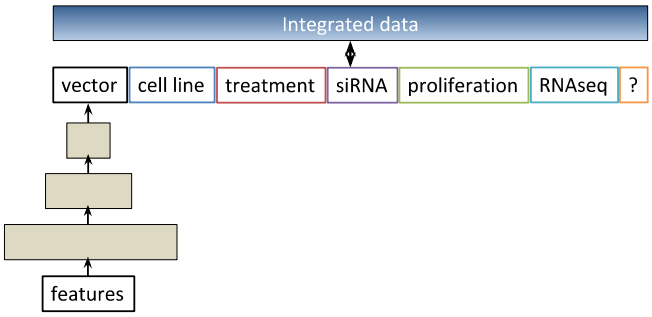

Figure 1. Schematic overview of the data integration in the deep architecture model.

IV. Conclusion

By exploiting deep learning architectures we will model the morphological feature space of multiple cell lines exposed to both kinome-wide RNAi knockdown and different treatment conditions. This will facilitate the systematic identification of the network rearrangements that occur after kinase knockdown, exposure to anti-cancer therapies and combinations thereof.

References

[1] Bakal C, *et al.* (2007). Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology. *Science* ***316***, 1753–1756.

[2] Hinton Ge *et al*, (2006). A Fast Learning Algorithm for Deep Belief Nets. Neural Computation **18**, 1527–1554

[3] Hinton GE, Salakhutdinov RR (2006) Reducing the Dimensionality of Data with Neural Networks. *Science* **313**, 504–507.

[1]Linding Lab, Biotech Research & Innovation Center, UCPH, Denmark.
[2]Niels Bohr Institute, UCPH, Denmark.
This project is funded by the Danish Innovation Fund: 1311-00010B