The Development of a Computational Tool for the Prediction of Cancer Drug Resistance

[1]Lebohang Mashatola and [1]Mandeep Kaur

[1] School of Molecular and Cell Biology, University of the Witwatersrand, Private Bag 3, WITS – 2050, Johannesburg, South Africa Presenter's Email: 681452@students.wits.ac.za
Supervisor: mandeep.kaur@wits.ac.za

Introduction: Cancer is the leading cause of death accounting for more than 10 million deaths annually worldwide. Cancer treatment regimens including chemotherapy, surgery and radiotherapy have proven effective. However, approximately 50% of cancer patients will experience therapy resistance. Unfavourable response to anticancer therapy is responsible for up to 90% of cancer related deaths. Precautionary measures by combination and aggressive therapy in early cancer stages can be taken to reduce the likelihood of drug resistance occurrence. The aim of the research is to develop a computational tool for the detection of drug resistance in cancer.

Methods: Feature selection by performing biomarker discovery was conducted. Anticancer therapy resistant cancer RNA-Seq count data samples from publicly available databases were downloaded from the National Cancer Institute data portal. Co-expression analysis identified genes related to therapy resistance by in silico validation. Machine learning models obtained from Sci-kit Learn were trained and tested on RNA-Seq count data with selected drug resistant genes. Furthermore, anticancer therapy resistant tissue slides were downloaded and used to train convoluted neural networks obtained from Keras. The predictive model trained on RNA-Seq count data as well as that trained on slide tissue data were integrated into a single predictive model. Independent RNA-Seq count test datasets and tissue slide imaging test datasets obtained from the National Center for Tumour Diseases and the Clinical Proteomic Tumor Analysis Consortium were used to evaluate the integrated trained model.

Results: Following co-expression analysis on anticancer therapy resistant samples, co-expressed gene clusters showing poor patient prognosis were used for feature selection for the training datasets. The integrative model shows promise in its ability to predict for anticancer therapy resistant and sensitive samples. Test datasets show up to 99% accuracy and an average of 98% accuracy on cross-validated test datasets. However, independent test datasets show a reduced accuracy of approximately 75% and an average of 63% accuracy on cross-validated test datasets. To improve prediction power, hyperparameter tuning, employing ensemble learning algorithms and alternative data transformation methods will be applied.

Conclusion: The emergence of the 4th industrial revolution has resulted in computational methods being used for decision making. Personalised and precision medicine has relied predominantly on using large computational power to solve complex biological patterns. There is therefore great potential in the clinical use of an anticancer therapy resistant prediction tool. As early detection of therapy failure may flag the use of more aggressive and targeted anticancer therapy which may increase survival rates in cancer patients.

Acknowledgements: This research was financially supported by National Research Foundation (NRF), South Africa (grant number: 129356) to MK, and DAAD-NRF Doctoral Scholarship (No: 123247) to LM.

Keywords: RNA-Seq, co-expression analysis, feature selection, integrative model