**Naive Prediction of Cancer Outcomes using Machine Learning**

Matt Bixley1, Mik Black1  
1Department of Biochemistry, University of Otago, New Zealand  
[matt.bixley@otago.ac.nz](mailto:matt.bixley@otago.ac.nz)  
[mik.black@otago.ac.nz](mailto:mik.black@otago.ac.nz)

### ABSTRACT / INTRODUCTION

Prediction of 5 year cancer outcomes from histology images has been undertaken using machine learning (ML), artificial intelligence (AI) and deep learning techniques, by multiple international research groups, with success for a number of different cancers (e.g., breast and colorectal). A key outcome in this approach is the easy translation of technology to allow pathologists to access the applications in their workflow. An extension to the idea of outcome prediction is to use histology image data to estimate genomic characteristics of a tumour, such as those often derived from gene expression data – examples include molecular subtype, proliferation rate, oncogenic pathway activation, and genomic instability.  
  
Typically the training process involves the hand delineation of 100s if not 1000s of slides to identify regions of interest and remove aberrations to improve accuracy. While some automation has been attempted, here we present a naive approach to estimate the accuracy with minimal human intervention. Currently the work has been applied to stomach cancer slides from The Cancer Genome Atlas (TCGA), using both patient outcome data, and genomic data on the molecular characteristics of the tumour.

### ABOUT THE AUTHOR(S)

### Matt is a Carpentries Instructor and Teaching/Research Fellow at the University of Otago. His research background extends from laboratory and field work through quantitative genetics and bioinformatics. Matt’s current research is on the use of Machine Learning tools to predict cancer outcomes.

Mik received a BSc(Hons) in statistics from the University of Canterbury, and a MSc (mathematical statistics) and PhD (statistics) from Purdue University. After completing his PhD in 2002, Mik returned to New Zealand to work as a lecturer in the Department of Statistics at the University of Auckland. An ongoing involvement in a number of Dunedin-based collaborative genomics projects resulted in a move to the University of Otago in 2006, where he now leads a research group focused on the development and application of statistical methods for the analysis of data from genomics experiments, with a particular emphasis on human disease. Mik has also been heavily involved in major initiatives designed to put in place sustainable national research infrastructure for NZ: Genomics Aotearoa and NZ Genomics Limited for genomics, digital literacy training via The Carpentries, and NeSI (New Zealand eScience Infrastructure) for high performance computing and eResearch.