A Bayesian Clustering Analysis of Breast Cancer Gene Expression

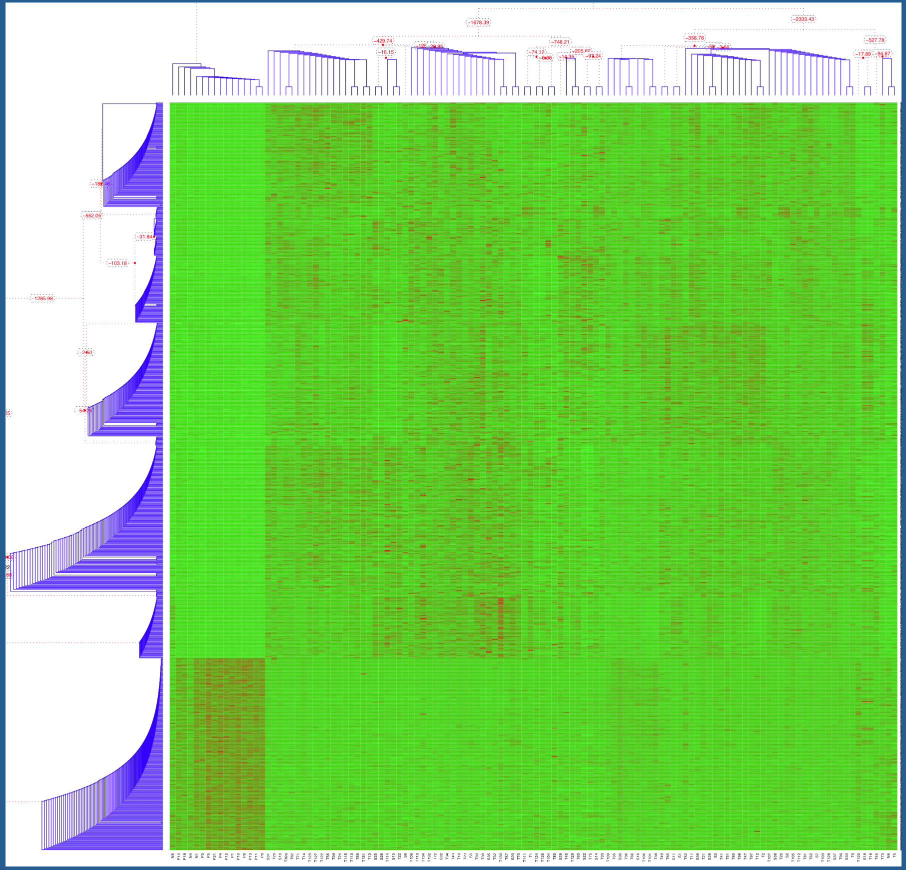
Richard S. Savage[1], Sinéad Aherne[2], Padraig Doolan[2], Colin Clarke[2], Lorraine O'Driscoll[3], Susan Kennedy[4], Martin Clynes[2], David L. Wild[1]

(1) Systems Biology Centre, University of Warwick, UK (2) National Institute for Cellular Biotechnology, Dublin City University, Ireland
(3) School of Pharmacy & Pharmaceutical Sciences, Trinity College Dublin, Ireland, (4) St Vincent's University Hospital, Dublin 4, Ireland

We present results from analysing microarray data from a breast cancer study comprising 104 cancer and 17 normal specimens using a Bayesian Hierarchical Clustering (BHC) algorithm (Savage et al. 2009) .

The BHC model is probabilistic, allowing it to account properly for uncertainty in the data and the use of model selection as a merging criterion allows us to infer in a principled way the most likely number of clusters in the data.

The results are interpreted using literature mining, gene ontology annotation and a biomarker discovery algorithm (Chu et al. 2005). As a case study of these results, we focus on a set of five gene clusters that are over-expressed in a tissue sample group for which cancer patients had particularly poor 7-year relapse-free and survival rates.



EPSRC
MRC Medical Research Council
THE UNIVERSITY OF WARWICK
DCU
NICB
MTCI Molecular Therapeutics for Cancer Ireland