

Sparse Bayesian Learning for Targeted Cancer Therapy

Anti-cancer drugs show remarkable heterogeneity in their efficacy for different patients and tumour types. Drugs which work well for some patients can prove ineffective for others, despite often devastating side-effects. This motivates a need to develop methods by which to predict, ahead of treatment, whether or not a drug is likely to prove effective. Furthermore, a predictor which can be related to underlying tumour biology may also shed light on the as yet poorly understood molecular mechanisms by which heterogeneity in drug response comes about.

The primary aim of this project will be to exploit recent advances in Bayesian machine learning to develop predictors of drug response based on the activity status of proteins involved in key signaling pathways in breast cancer. The analysis would take account of the activation status of individual proteins and most importantly the *interplay* between them. The project will use data pertaining to breast cancer cell-lines described in Neve et al. (2006). The relevant statistical theory is described in Mukherjee & Speed (2007) and involves elements of Bayesian model averaging and Markov chain Monte Carlo. To make a success of the project the student would need to develop an understanding of the statistical methodology, associated computation and some of the relevant biology.

Abilities required: You will need to have an understanding of statistical inference, and at least some acquaintance with Bayesian methods. While no prior knowledge of cancer molecular biology is required, you would need to be ready to learn.

PhD opportunities: The mini-project could lead to a PhD project. A natural extension would be to develop predictive approaches which explicitly take account of what is already known about the structure of relevant biological pathways.

(Neve et al., 2006) R. M. Neve et al., A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes, *Cancer Cell*, 10(6):515-527, 2006

(Mukherjee & Speed 2007) S. Mukherjee and T. P. Speed. Sparse, noisy Boolean functions. Technical report #731, Department of Statistics, U. C. Berkeley, 2007.