# Network inference for cancer subtypes

Supervisor: Sach Mukherjee

**Summary: This project will exploit recent advances in network inference to address a crucial question in cancer biology, that of discovering tumour subtypes.**

Networks of proteins called *signalling networks* operate in concert to control cellular proliferation. Cancer is fundamentally a phenomenon of uncontrolled proliferation, and signalling networks play a central role in the biology of the disease. Human tumours display a remarkable degree of heterogeneity in response to therapy (Weinberg, 2007). Current cancer therapies are predominantly "one size fits all" inasmuch as they do not tailor treatment to the specific characteristics of individual tumours. For this reason, the question of uncovering cancer subtypes, which harbour broadly similar aberrations to underlying signalling networks, and may therefore respond in a similar way to treatment, is a critical one in cancer biology. Discovering subtypes in this sense will aid in developing therapies which are targeted at specific patients on the basis of the molecular characteristics of their particular tumour.

In recent years, a class of probabilistic models called *graphical models* (Jordan, 2004) have emerged as a key approach for the study of biological networks (see e.g. Friedman, 2004). These models use a graph *G* to embody probabilistic relationships between variables and are a powerful tool for studying systems of interacting components. In ongoing work, we are using graphical models, Monte Carlo methods, and Bayesian priors to address key questions in cancer biology (see e.g. Mukherjee & Speed, 2008). A key finding of our recent efforts (Mukherjee et al., 2007; Mukherjee et al., 2009) has been that signalling networks differ in important details between tumour subtypes, and in ways which are therapeutically relevant.

However, to date we have focused on the setting in which tumour subtypes are already known. We would like to move towards automatically discovering cancer subtypes which are coherent in a network sense. **The central aim of this project is explore precisely this idea of cancer subtype discovery using network inference.** The student will benefit from clear direction at the outset, but at same time the project offers ample scope for creative work.

**Skills developed:** The project will develop your understanding of machine learning in general and network inference in particular, and acquaint you with related challenges in cancer biology. The project is ambitious, and aimed at an important, unsolved problem, so an outstanding project may lead to a scientific publication.

**Abilities required:** You will need to have good computational skills and an understanding of statistical inference (at the level of Co902). *Absolutely no prior knowledge of biology is required.*

**PhD opportunities:** The mini-project could lead very naturally to a PhD project; both the specific question and general area are rich in research opportunities.

(Friedman 2004) N. Friedman. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science*, **303**(5659):799-805, 2004.

(Jordan, 2004) M. I. Jordan. Graphical Models. *Statistical Science* **19**:140-155.

(Mukherjee *et al.*, 2007) S. Mukherjee, N. Bayani, N. Shastri, K. Hacker, A. Bhamidipati, E. Gentalen, U. Nguyen, D. Voehringer, K. Voss, T. Yang, P. B. Vander Horn, S. Pelech, T. P. Speed, J. W. Gray and R. M. Neve. ERK-pathway connectivity is dependent on tumor sub-type. *Proceedings of the 98th Annual Meeting of the American Association for Cancer Research, Los Angeles, 2007.*

(Mukherjee & Speed, 2008) Mukherjee, S. & Speed, T. P. Network inference using informative priors. *Proceedings of the National Academy of Sciences,* **105**(38): 14313–14318, 2008.

(Mukherjee *et al.,* 2009) Mukherjee, S., Neve, R.M., Spellman, P.T., Speed, T.P., Gray, J.W. et al. Bayesian analysis reveals distinct ERK pathway connectivity in breast cancer subtypes. In preparation.

(Weinberg, 2007) Weinberg, R. *The Biology of Cancer*. Garland Science, 2007.