Towards personalized treatment of cancer

Tomas Jacobsson[*]

May 2012

Abstract

Every year about 55,000 people are diagnosed with cancer and about 20,000 people die from the disease in Sweden. Statistically, one of three Swedes will suffer from cancer at some time in their life. The significance of reducing cancer mortality can hardly be overstated. Most patients can be cured by cancer surgery and radiation of the primary tumor if no metastasis have occurred. It is therefore very important to reduce the tumor spreading at an early stage. To reduce the tumor spreading, chemotherapy is normally used to inhibit the production of new cancer cells.

During the last decade it has been realized that cancer is a heterogeneous disease. This suggests that we need better methods to match molecular tumor characteristics with an optimal drug combination for each patient. To study the potential of developing such methods, data from cancer cell lines was used to test two cases: Case 1) To optimize the treatment based on the patient's molecular tumor profile. Case 2) To develop accurate prediction models for drug screening to help make cancer drug discovery more efficient.

By using different techniques within the field of chemometrics (the intersection of chemistry and statistics) it was possible to integrate gene expression data (describing the characteristics of a cancer tumor) and chemical data (describing the properties of a chemical compound) to predict the concentration level needed for a chemical compound to inhibit the cell line growth with 50%. Such a prediction model is of direct use in Case 1 and Case 2, where it can be used for predicting the optimal drug treatment based on the patient's molecular tumor profile and for predicting the effect of a new drug candidate, respectively.

In both cases, the best model for predicting the concentration level needed for a chemical compound to inhibit the cell line growth with 50% was achieved with random forest. The variables describing the chemical compounds were of high importance when predictions were made. Most importantly, it was also found that the gene expression data, describing the cancer cell line, adds significant information, indicating that cancer treatment should be personalized.

---

[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: jacobsson_tomas.hotmail.com. Supervisor: Jan-Olov Persson.