Abstract

In the past decade microarray technology has had a big impact on cancer research. More recently other technologies such as mass spectrometry-based proteomics or array comparative genomic hybridization have emerged as data providers with potentially similar impact. These technologies have a top-down approach in common instead of a bottom-up. Whether it is the genome, transcriptome or proteome that is targeted, each technology attempts to capture its corresponding 'omics' as a whole. Moreover, the data resulting from these technologies potentially hold information on the actual biological reasons why subsets of tumors behave differently, instead of relying on general clinical data or morphological characteristics of a tumor.

In our research, we investigated how omics data can be used to predict diagnosis, prognosis or therapy response in cancer. The large dimensionality of omics data however prohibits direct interpretation and requires dedicated models. Biomedical decision support modeling attempts to tackle this issue and aims to build reliable models. We focused on the use of Bayesian networks as biomedical decision support model. More specifically, we developed a Bayesian network integration framework able to integrate heterogeneous and high-dimensional data. We consider two specific types of data in our framework: patient specific data or entity specific data. We define patient specific data as primary data and entity specific data as secondary data. The latter characterizes entities within each omics layer such as genes in the genome, mRNA in the transcriptome or proteins in the proteome. First, we illustrate Bayesian network modeling on two primary data sources separately: clinical and genomic data. Secondly, we develop algorithms to integrate primary data sources. Finally, we extend the framework to include secondary data sources.

Besides the use of publicly available data and due to the availability of unique data gathered at the University Hospitals Leuven, we applied our framework on two main cancer sites: ovarian cancer and rectal cancer. Our results show the potential of integrating both primary and secondary data sources. Finally, we look into the future and project which research avenues should be pursued to improve the framework.