Abstract

Defects in the genetic make up of a cell or organism can lead to developmental problems and hereditary disease. In order to be able to discover these aberrations in the genome of patients, certain techniques are used in both clinical routine diagnostics as in genetic research studies to chart the genetic material. Screening of patient tissue by investigating the genetic material in a cell at a molecular level is called Molecular Karyotyping. Recent evolutions in this field allow a significant increase in resolution when determining the chromosomal characteristic of a tissue. In this respect, the determination of deletions and duplications, among other genomic aberrations, happens through a technology called Array CGH (Comparative Genomic Hybridization), in which representative genomic fragments are used as reporters and are spotted on small glass slides in microarrays.

The growing body of array and patient related data bring specific challenges with respect to the identification and pin-pointing of genomic regions linked to inherited or acquired aberrations, and to the automated analysis of all patient related data. Important challenges have emerged regarding the interpretation of the results of this genetic screening technique.

In the context of this doctoral dissertation, one of the main goals has been to find ways to support and enable this interpretation. On the one hand, methods were sought to handle the raw, untreated data that come from array CGH experiments. Where earlier genetic screening methods relied on interpretation through visual inspection, with array CGH, this is no longer possible, and the need of means for analysis and interpretation of large quantities of numeric data has become obvious. On the other hand, methods have been developed through approaches such as automated text analysis on publicly available biomedical literature and on patient data, among other things – to link phenotype traits to genomic regions, and to give clinical meaning to the results from array CGH assays. Finally, the combination of these methods has lead to a platform for storage, analysis, and management of all data related to patients for whom a genetic screening through Array CGH has been performed, including not only raw data but phenotype descriptions and clinical information as well. This platform supports the interpretation of genetic, clinical and phenotype information and allows for genome annotation in both a diagnostic and a research context.