Daniel E. Stange
Dr. med.

High Resolution Genomic Profiling for the Detection of DNA Copy Number Aberrations: Towards the Elucidation of Pathomechanisms in Crohn's Disease and Invasive Breast Cancer

Geboren am 24.1.1980 in Ulm
Staatsexamen am 24.10.2006 an der Universität Heidelberg

Promotionsfach: Deutsches Krebsforschungszentrum (DKFZ)
Doktorvater: Prof. Dr. rer. nat. Peter Lichter

High resolution genomic profiling by matrix-CGH (comparative genomic hybridization) has revolutionized the detection of DNA copy number alterations in human diseases. This thesis comprises the development of a new matrix-CGH microarray consisting of ~ 6500 genomic fragments, the optimization of hybridization techniques and biostatistical analysis as well as the validation of this setting for clinical applications. To this aim a constitutional germ-line disease with homogenous genetic background as well as a sporadic disease with a mixture of normal and aberrant genomes (cells) was analyzed. In the first application, DNA copy numbers were analyzed by using the new matrix-CGH microarray in the germ-line of patients with colonic Crohn´s disease. The only region in the genome that differed between patients and healthy controls was the beta-defensin gene cluster on chromosome arm 8p23.1 Defensins are endogenous antimicrobial peptides protecting the intestinal mucosa against bacterial invasion. The DNA copy number of this gene cluster is known to be highly polymorphic within the normal population. In this thesis it was shown that patients with colonic Crohn´s disease have a shift towards lower copy numbers compared to healthy individuals. In a collaborative study, we could subsequently correlate the low beta-defensin copy number to a diminished mRNA expression of one of the affected genes. Furthermore, it was possible to show that individuals with three or fewer copies have a significantly higher risk of developing colonic Crohn´s disease than individuals with four or more copies (OR: 3.06, 95% CI: 1.46-6.45). These findings provide a genetic basis for the deficient beta-defensin expression, which may underlie the chronic inflammation of colonic Crohn´s disease.

The second application is concerned with the molecular analysis of subtypes of sporadic invasive breast cancer. The two major histological subtypes of invasive breast cancer were analyzed, invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). They differ with regard to presentation, metastatic spread and epidemiological features. To elucidate the genetic basis of these differences, copy number imbalances that differentiate the histological subtypes were identified. Classification algorithms achieved the highest discriminating power between IDC and ILC, when combining the aberration patterns of chromosome arms 1q and 16p, which were significantly more often gained in ILC. These regions were further narrowed down to sub-regions 1q24.2-25.1, 1q25.3-q31.3 and 16p11.2. Located within the minimally gained regions on 1q are two genes, *FMO2* and *PTGS2*, known to be overexpressed in ILC relative to IDC. Assessment by RQ-PCR of four candidate genes in the minimally gained regions on 16p11.2 revealed significant overexpression of *FUS* and *ITGAX* in ILC with 16p copy number gain. Unsupervised hierarchical cluster analysis identified three molecular subgroups that were characterized by different aberration patterns, in particular gain of *MYC* (8q24) and the identified candidate regions on 1q24.2-25.1, 1q25.3-q31.3 and 16p11.2. These genetic subgroups differed with regard to histology, tumor grading, frequency of alterations and estrogen receptor expression. Thus, molecular profiling for the

first time identified DNA copy number imbalances on 1q and 16p as significant classifiers of histological and molecular subgroups.

In conclusion, this thesis includes methodological work in the development and improvement of a new matrix-CGH microarray that subsequently proved to have sufficient sensitivity for the analysis of DNA copy number aberrations of homogenous genomes in germ-line diseases as well as the more demanding analysis of heterogeneous genomes in sporadic disease.