Rongxi Yang
Dr. Sc. hum

Identification of inherited familial breast cancer risk factors focusing on ultraconserved elements, miRNAs and codon-usage-changing polymorphisms

Promotionsfach: DKFZ (Deutsches Krebsforschungszentrum)
Doktormutter: Prof. Dr. Barbara Burwinkel

Breast cancer is the most common cancer among women, and accounts for one fourth of all female cancers. Familial breast cancer accounts for about 10 % of all breast cancer cases. Inherited genetic factors are one of the most important reasons for the familial aggregation. The mutations in breast cancer high-penetrance genes, *BRCA1* and *BRCA2*, account for about 25 % of all familial cases. Despite the continuous discovery of intermediate- and low-penetrance genes, they only account for about 10 % of familial cases. Most of the familial breast cancer risk factors are still unknown. The identification of further breast cancer susceptibility genes and genetic variants is essential for understanding the etiology of this disease, the development of diagnostics and therapeutic strategies. We therefore mainly focused on familial breast cancer risk/protective factors in genes, especially the SNPs in UCEs, within miRNAs and synonymous SNPs with alternation from preferred codon-usage. Investigation the inherited methylation signatures in familial breast cancer is also a key study point of us.

We first investigated SNPs in ultraconserved elements (UCEs), which are segments of > 200 bp length showing absolute sequence identity between orthologous regions of human, rat and mouse genomes. Recent studies have shown that some UCEs can function as long-range enhancers of flanking genes or are involved in splicing when overlapping with exons. The depletion of UCEs among copy number variations and the significant underrepresentation of SNPs in UCEs have also revealed their evolutional and functional

importance. We investigated the influence of six SNPs in UCEs on familial breast cancer risk. Two out of six SNPs showed an association with familial breast cancer risk. Whereas rs9572903 showed only a significant borderline association, the frequency of the rare [G] allele of rs2056116 was higher in cases than in controls indicating an increased familial breast cancer risk. Comparing with the older age group, the ORs were increased in women younger than 50 years old, pointing to a possible age- or hormone- related effect. For the first time, SNPs in UCEs are shown to be associated with cancer risk.

Second, we analyzed the codon-usage-changing synonymous SNPs in *BRCA1* and *BRCA2* genes. Numerous non-synonymous SNPs in *BRCA1/2* have shown associations with breast cancer risk, whereas the impact of codon-usage-changing synonymous SNPs on breast cancer risk has not been studied yet. Recently, it has been reported that synonymous SNPs leading to an aberration from the preferred codon-usage can have functional effects and consequently be associated with disease. Thus, we searched for SNPs with alternation from preferred codon-usage in *BRCA1/2*. Based on defined criteria, two codon-usage-changing variants, Ser455Ser (1365A>G) and Ser2414Ser (7242A>G), were detected in *BRCA2*, whereas no such variant could be identified in *BRCA1*. We investigated the impact of the two variants on familial breast cancer risk in a large case-control study. As a result, the two codon-usage-changing SNPs showed no association with familial breast cancer risk.

Third, we also focused our study on SNPs within microRNAs and their association with familial breast cancer. MicroRNAs regulate a large number of cellular pathways by the degradation of target mRNAs and/or repression of their translation. Although SNPs in miRNAs target sites have been studied, the effects of SNPs within miRNAs are largely unknown. We first systematically sequenced breast cancer related miRNA genes and their flanking regions to identify/verify SNPs. We analyzed four SNPs within pre-miRNAs or miRNA flanking regions for a putative association with familial breast cancer risk. The SNP rs895819, located in the terminal loop of pre-miRNA-27a, showed a protective effect. The rare [G] allele of rs895819 was found to be less frequent in the cases than in the controls, indicating a reduced familial breast cancer risk. Furthermore, age stratification revealed that

the protective effect was mainly observed in the age group < 50 years old. It has been shown that artificial mutations in the terminal loop of pre-miR-27a can block the maturation process of the miRNA. We hypothesize that the G-variant of rs895819 might impair the maturation of the oncogenic miR-27a and thus, is associated with familial breast cancer risk.

Recently, the SNPs rs11614913 in miR-196a2, rs3746444 in miR-499 and rs2910164 in miR-146a were reported to be associated with increased breast cancer risk. In order to further investigate the effects of these SNPs in European population, we examined the three SNPs in German and Italian study population by case-control studies, and of the two series combined. We also investigated the effects of the three SNPs on age at breast cancer diagnosis. None of the performed analyses showed statistically significant results. In conclusion, we can not find any association between SNPs rs2910164, rs11614913 and rs3746444 and familial breast cancer risk, or age at breast cancer onset in Caucasian population.

Fourth, we also examined the breast cancer related inherited methylation signatures. We analyzed the methylation status of genomic DNA of the peripheral blood from 72 familial breast cancer patients and 30 healthy controls. A genome-wide scan was applied to interrogate more than 27,000 CpG loci covering about 14,000 genes by the Infinium HumanMethylation27 BeadChip. Applying defined stringent criteria, we selected 53 most significant CpG loci by comparing the methylation level of each CpG between the average of cases and the average of controls. The 53 CpGs are located in different genes, two thirds of which are involved in cancer related pathways or considered as methylation markers. This result gives us candidate CpG loci for the further validation by alternative methods and on enlarged sample size.

In conclusion, we investigated the inherited risk factors and their associations with breast cancer. Our results suggested that the polymorphisms in non-coding regions such as in UCEs and pre-miRNAs might be associated with inherited breast cancer risk. It is possible that these SNPs are functional itself or in linkage disequilibrium with functional variants. The accumulation and combination of inherited variants may have a complex impact on

breast cancer at the individual level. In all, the identification of new predisposing variants and variant combinations are important for understanding the etiology of familial breast cancer and might give hints for new clinical diagnosis and protective strategies. The investigation on genetic and epigenetic inherited factors for familial breast cancer risk will be one of our key study points in the future.