Jonathan Alexander Heiss

Dr. sc. hum.

Epigenome-wide search of biomarkers for early detection of colorectal cancer

Fach/Einrichtung: Medizinische Biometrie u. Informatik / DKFZ

Doktorvater: Prof. Dr. med. Hermann Brenner

Colorectal cancer (CRC) is the second most common cancer in Germany and a major public health burden. If CRC is diagnosed at an early stage, prognosis is good. But the majority of CRC cases are diagnosed at later stages that are associated with low survival rates. Early detection of CRC and precursors by screening reduces CRC incidence and mortality. Current screening modalities like colonoscopy or occult blood stool tests suffer from low adherence. The development of blood tests could widen the reach of CRC screening programs. The aim of this work is to evaluate the potential of leukocyte DNA methylation as a biomarker for the early detection of CRC. For this purpose a hypothesis-free epigenome-wide association study is conducted.

Leukocyte DNA methylation is assessed with high-density microarrays (Illumina Infinium HumanMethylation450 BeadChip). In order to analyze data generated by this platform an analytical pipeline is developed, starting with a new normalization method. To test this method the proper design of benchmarks is discussed. Subsequently, the new normalization method is compared to existing ones in several benchmarks. The results show that the new method outperforms the competition. Furthermore, the issue of confounding by leukocyte composition is discussed: the magnitude of confounding for various common outcomes in epidemiological studies (such as smoking, BMI, diabetes, etc.) is quantified and a model to estimate cell proportions from leukocyte DNA methylation data is developed. Cell proportions estimated from this model are compared against cell counts in a large cohort study, showing that the estimates are often more precise than the cell counts. These estimates are then used as covariates in a hypothesis-free epigenome-wide association study: DNA methylation is assessed in leukocyte samples from CRC cases and controls collected prospectively from participants of screening colonoscopy (screening setting) and from clinical CRC cases (clinical setting), additionally. The large differences between the screening setting and the clinical setting are illustrated. Three differentially methylated CpG sites are identified. A diagnostic models based on this marker panel is trained and the model performance in the screening setting is evaluated, achieving a moderate discrimination between cases and controls (c-statistic 0.69). In an additional analysis the potential of the leukocyte composition, estimated from leukocyte DNA methylation levels, for the same purpose, the early detection of CRC, is evaluated. Diagnostic models based on the leukocyte composition achieve similar performance (c-statistic 0.70).

In conclusion, while the identified biomarkers on their own are not competitive to established screening modalities, they might contribute to the development of multi-marker panels. Furthermore, results suggest that discovery and validation of biomarkers should be conducted in studies using a prospective design close to application, because results from studies using a retrospective design might not be representative of marker performance.