Abstract of the doctoral dissertation

MS Aleksandra Pisarek

„ Development of epigenetic tools for age prediction in human semen samples”

Over the last several years, extensive DNA methylation studies have allowed for the discovery of a number of practical markers that can be used to predict various diseases. One of the most remarkable discoveries was the demonstration of a strong correlation between DNA methylation and human ageing. This finding has led to the numerous applied research studies, the goal of which was to find methylation markers dedicated to accurate estimation of human age. The epigenetic age prediction has clinical significance and is useful in forensic science. Age is an important factor for the development of many diseases, and the accelerated ageing has been associated, for example, with cardiovascular diseases, diabetes, overweight or neurodegenerative diseases. In a police investigation, age is one of the crucial pieces of information that can efficiently reduce the number of potential suspects and thus speed up the investigation process. Many somatic cells and tissues examined in a forensic application have shown that age prediction based on DNA methylation analysis offers high accuracy. Although semen samples are one of the most significant sources of DNA analyzed in forensic laboratories in cases of sexual crimes, the prediction of epigenetic age in semen remains challenging.

The purpose of this doctoral thesis was to evaluate the available markers and tools for the prediction of epigenetic age, select new age estimation markers and develop reliable prediction models for semen samples. The study analyzed data collected from 467 men aged 19 to 60 years. First, the focus was on the evaluation of two available tools for epigenetic age estimation in semen. The first tested calculator analyzes the methylation status of 3 CpG sites using the minisequencing method. For all the investigated markers, the obtained results confirmed a statistically significant correlation of DNA methylation with age. The mean absolute error (MAE) for the analyzed set of semen samples from the Polish population was 6 years. The second evaluated method assesses the level of DNA methylation within 51 loci employing microarray technology. The conducted analyzes showed a significant correlation with age for 98% of the analyzed markers and higher prediction accuracy with an MAE < 3 years. The second goal of the project required large-scale studies and was realized by analysis of 38 semen samples from men aged 24 to 58 years with the Infinium MethylationEPIC v1.0 BeadChip (EPIC) microarray technology. With the use of R packages, 10 candidate CpG sites were selected, which together with the 3 cytosines included in the first of the previously tested tools (*TTC7B*, *FOLH1B* and *LOC401324* genes) were validated on a set of 179 samples from men aged 26 to 57 years. Targeted next-generation sequencing (NGS) was used to develop a protocol to collect methylation data along with the VISAGE Enhanced Tool development. The proposed VISAGE predictor analyzes the DNA methylation level of 6 CpG sites within the 5 genes *FOLH1B*, *SH2B2*,

EXOC3, *IFITM2* and *GALR2*. In the training group the selected markers explained 60% of the age variability. The new age model provides a prediction accuracy with an MAE = 5.1 years in the test group. At the next stage of the research, the collected microarray data was used to develop the optimal predictive model, EPIC-8CpG. The regression model is based on 8 CpG sites from 8 different genetic loci and shows the prediction of epigenetic age in semen with an MAE = 3.2 years. This model has an optimal ratio in terms of the number of markers and accuracy of prediction for forensic applications. As part of the last stage of the research, the data collected for 391 semen samples (individuals aged 19 – 57 years) using the EPIC and NGS technology allowed for a thorough investigation of DNA methylation values obtained for both technologies and the effect of variations on the epigenetic age prediction. The goal of future research into the development of accurate semen age prediction tools is to investigate the influence of lifestyle or semen quality factors on the accuracy of age estimation.