“Lifestyle-dependent epigenetic signatures and the impact of lifestyle on epigenetic age acceleration”

Joanna Rudnicka

Abstract

Lifestyle encompasses the broad spectrum of daily habits and choices that collectively affect overall health and well-being. Studies show that these factors have an impact on a biological level by influencing the epigenome, particularly by altering DNA methylation patterns. Epigenome-wide association studies (EWAS) have provided long lists of CpGs associated with smoking, and also other individual habits, such as alcohol consumption, BMI (lifestyle outcome) and physical activity. Importantly, some of these changes maintain and can be detectable in blood cells for many years, while others are rapidly reversible. In forensics, the ability to reliably predict lifestyle from trace DNA based on selected markers can be highly informative in characterizing an unknown trace donor, and thus useful in guiding investigations. The identification of behavioral signatures in the epigenome is also of medical significance. DNA methylation changes that persist in the human body can pose a risk of habit-related diseases. Therefore, knowledge of lifestyle-triggered changes in genes and molecular pathways would allow a better understanding of the lifestyle impact on general health and aging processes.

The first part of this dissertation examines lifestyle-derived DNA methylation changes, focusing on identifying epigenetic signatures and exploring their influence on epigenetic aging. The analyses were conducted on two cohorts: Cohort 1, comprising 755 living individuals aged 20 – 87 years, and Cohort 2, including 200 individuals aged 30 – 60 years (collected post mortem). Comprehensive EWAS analyses were performed for each lifestyle factor, employing two analytical models: a basic model adjusted for age and sex, and an extended model incorporating additional covariates: immune cell composition, sociodemographic variables, and lifestyle habits.

Significant associations between lifestyle factors and DNA methylation were identified. For smoking, 51 CpG markers reached epigenome-wide significance ($P < 9.42 \times 10^{-8}$). The most significant marker, cg05575921 in the *AHRR* gene, exhibited an extraordinarily high association with smoking ($P = 2.13 \times 10^{-38}$), making it the strongest marker identified across all studied lifestyle factors. The identified markers enabled the development of a robust only

3 CpGs-based classifier for predicting smoking status, achieving high accuracy (AUC = 0.788). For alcohol abuse, a notably larger set of 368 markers were discovered representing the most epigenetically impactful lifestyle factor studied. For BMI, 5 significant CpG markers were identified, including *CPT1A* and *ABCG1*, both associated with lipid metabolism and inflammatory responses. Additionally, single CpG markers were identified for physical activity (studied in a group of bodybuilders), coffee consumption, and sleep duration at a suggestive EWAS significance threshold ($P < 5 \times 10^{-6}$). While these markers did not reach epigenome-wide significance, they provide a foundation for further exploration of the subtle epigenetic effects of these lifestyle factors. Functional analyses across all lifestyle factors revealed that the identified markers were enriched in biological pathways related to lipid metabolism, immune regulation, inflammatory processes, and cellular stress responses. Alcohol consumption and smoking were the most impactful, particularly in accelerating epigenetic aging as assessed by clocks such as GrimAge, which proved highly sensitive to these exposures. BMI, sleep duration, coffee intake, and physical activity showed more nuanced effects, with their impact varying depending on the clock used.

The second part of the dissertation focuses on testing high-throughput sequencing (HTS) technologies for targeted DNA methylation analysis, specifically Ion AmpliSeq and SureSelect, compared to epigenome-wide method using EPIC microarrays. Ion AmpliSeq demonstrated superior sensitivity, accurately quantifying methylation levels with DNA inputs as low as 25 ng, making it highly suitable for forensic applications where sample amounts are often limited. SureSelect allowed for the analysis of large genomic regions (up to 24 Mb) but required minimum DNA input of 500 ng, restricting its utility for low-input scenarios. EPIC microarrays provided comprehensive methylation profiling but were less effective for targeted analyses due to fixed marker positions and high DNA input. Comparative read depth and repeatability assessments showed that Ion AmpliSeq maintained consistent target coverage (100% at $\geq 50\times$ read depth) and achieved the highest correlation between replicates (Spearman r = 0.993).

Overall, this study confirmed well-established epigenetic markers of lifestyle factors, and identified novel markers, significantly expanding understanding of how lifestyle influence DNA methylation and the rate of epigenetic aging. These findings also represent a critical step toward developing robust DNA methylation-based prediction tools for forensic and clinical research. By bridging lifestyle behaviors and DNA methylation, this dissertation lays the groundwork for medical applications, enabling better disease risks prediction and personalized interventions for healthy aging.