Comp 1 Write-Up: Smoking and Epigenetic Age Acceleration

Problem. We investigated whether smoking status (never, former, current) affects epigenetic age, using DNA methylation clocks. Smoking is a well-established health risk, and this analysis tested whether clocks can detect its biological impact on biological aging.

Methods and Rationale. We used the publicly available DNA methylation dataset **GSE50660** (GEO), which includes smoking status annotations: 179 never, 263 former, and 22 current smokers. The dataset was selected because the smoking categories were clearly defined, enabling us to stratify groups and test whether epigenetic clocks capture smoking-related aging effects.

I began with the **Horvath v1 clock**, since it is widely used as a reference and appears in BioLearn examples. Horvath v1 is optimized to predict chronological age, so it seemed a natural starting point. However, both simple and residual acceleration showed minimal group differences, confirmed by visualizations and statistical tests. This prompted a reevaluation: if the clock is not tuned to lifestyle effects, it may miss smoking-related signals. I therefore turned to **GrimAge**, a mortality-trained clock that incorporates DNAm-based surrogates for smoking exposure and other health-related risk factors.

We considered two measures: **Simple acceleration** (DNAm age – chronological age) and **Residual acceleration** (residuals after regressing DNAm age on chronological age). Residuals help control for expected age effects when group age distributions differ. To visualize results, I used **violin + box + scatter plots** (showing samples, distribution, and summary statistics together) and **ridgeline plots** (emphasizing distribution overlap or separation). For statistical testing, I applied **Welch's t-tests**, chosen for robustness under unequal sample sizes and variances.

Results. With Horvath v1, smoking effects were absent: group means differed by <1 year, distributions overlapped, and Welch's tests gave $p > 0.2$. With GrimAge, strong effects emerged: former smokers were on average ~**5.8 years older epigenetically** than never smokers ($p < 0.001$), and current smokers were ~**7.8 years older** ($p < 0.001$). The difference between former and current smokers (~2 years) was not statistically significant.

Representative Figures (GrimAge residual acceleration). *(Additional visualizations are provided in the notebook.)*

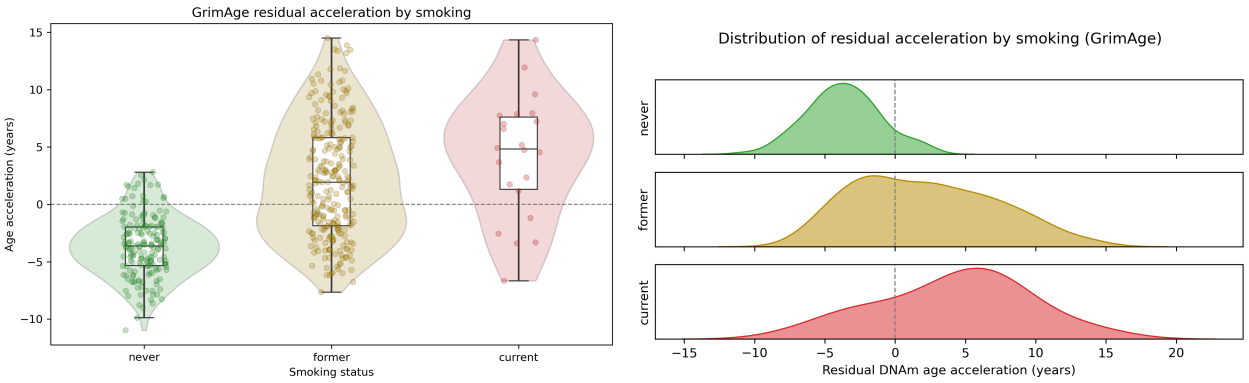

Figure 1. Violin + box + scatter (left) and ridgeline density plots (right) for GrimAge residual acceleration. Together, these complementary views highlight clear upward shifts for former and current smokers compared to never smokers. The violin emphasizes group-level distribution and summary statistics, while the ridgeline clarifies the degree of overlap across distributions.

Summary. Smoking accelerates epigenetic age by ~6–8 years when measured with GrimAge. In contrast, Horvath v1 failed to detect this effect, showing that **clock choice is critical**: starting from a familiar but less suitable model can obscure true biological signals. This analysis illustrates both the biological finding (smoking accelerates aging) and the methodological lesson (appropriate model selection is essential in epigenetic studies).