Methylation QC and EWAS Write-up for manuscript

Trent Hawkins

# Full Write-up

Probe Removal

We first removed low quality and cross-reactive probes, as determined by the processing pipeline from SeSAMe ver. 1.15.7.

Data Transformations and Normalizations

We corrected for technical variation in background fluorescence with normal-exponential deconvolution using out-of-band probes (noob). After correcting for that variation, density plots still showed differences between the distributions of type-I and type-II probes. We included further correction for this difference using the quantile normalization procedure performed by matchDesign from SeSAMe. Finally, a non-linear dye bias correction was used to correct for a difference in signal intensity between the red and green color channels.

After quality control, it is standard procedure to estimate cell composition for the sampled tissue type to account for the differences in methylation profile across cell types. No known reference set exists for nasal-epithelial tissue, and there was a concern that reference-free approaches would not be robust to outliers due to a small sample size. To address this source of variability, we referred to Removal of Unwanted Variability for Methylation (RUVm) to define an empirical set of negative control probes before performing RUV-4 (limma, missMethyl). We preserved biological variability due to vape status, age, and sex. We included two normalization factors after inspection of an elbow plot and clustering along the first two principal components.

*Epigenome-Wide Association Study (EWAS)*

We modeled M-values using linear models which adjusted for vape status, recruitment center, sex, age, and the two normalization factors. We finally corrected for epigenomic inflation using *bacon* [32] and for multiple comparisons using FDR adjustment.

# Condensed Write-up

We performed quality control and normalization for methylation probes using the process documented by SeSAMe ver. 1.15.7 (citation for: SeSAMe). This procedure includes removal of low-quality and cross-reactive probes followed by normalizations for background fluorescence and differences in distribution between type I and type II probes.

After quality control, it is standard to estimate cell composition for the collected sample type to account for variation in methylation profile across cell types using a reference set. At the time of this study, there is no known reference set for nasal-epithelial tissue. There was also a concern that reference-free approaches would not be robust to outliers or technical variation due to a small sample size. We used Removal of Unwanted Variability for Methylation (RUVm) to identify an empirical set of negative control probes which could then be used to remove remaining sources of unwanted variation using RUV-4 (citation for: limma, missMethyl). When estimating negative control probes, we preserved biological variability due to vape status, age, and sex. We included two normalization factors after inspection of an elbow plot, a relative log-expression plot, and clustering along the first two principal components of the normalized m-values.

We modeled M-values using linear models which adjusted for vape status, recruitment center, sex, age, and the two normalization factors. T-tests were used to test for a difference in M-values between vape users and non-users, adjusting for the other covariates. We finally corrected for epigenomic inflation using *bacon* (citation for: bacon) and for multiple comparisons using FDR adjustment.