Predicting time-since-deposition of bloodstains using LC-MS/MS and machine learning

Kirstine Lykke Nielsen[1*], Johan Kjeldbjerg Lassen[2], Ida Marie Marquart Løber[1,2], Kim Frisch[1], Tomasz Pawel Czaja[3], Bekzod Khakimov[3], Søren Balling Engelsen[3], Mogens Johannsen[1], Palle Villesen[2,4]

* klyn@forens.au.dk
[1] Department of Forensic Medicine, Aarhus University, Aarhus, Denmark
[2] Bioinformatics Research Center, Aarhus University, Aarhus, Denmark
[3] Department of Food Science, University of Copenhagen, Copenhagen, Denmark
[4] Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

Knowing the age of bloodstains, as determined by the time since their deposition, can be significant in forensic investigations for reconstructing the sequence of events. However, there are currently no established and accepted methods for this purpose. In this research, we introduce a metabolomics approach to identify aging biomarkers, that serve as inputs for machine learning algorithms used to predict the age of bloodstains.

Selection of 54 biomarkers was conducted through untargeted UHPLC-qTOF-MS screening of 150 bloodstains collected from five volunteers. The bloodstains were placed on glass and left to age between 1h and 64 days at 23°C and 40% humidity before metabolite extraction. The biomarkers include a variety of metabolites and their breakdown products for which a targeted and sensitive LC-MS/MS analysis was developed. The number of biomarkers were further reduced to 42 by only selecting the most stable compounds based on technical replicates.

To overcome differences in blood volume and metabolite variations between donors, we utilized ratios between biomarkers to make the data normalization free. Three machine learning models were tested including a general linear model (GLM), a Lasso (least shrinkage and selection operator) model, and a gradient boosting model (lightGBM). Lasso and lightGBM showed the best cross validation accuracy of the training set.

Three independent aging studies with different donors and bloodstain ages were additionally used to test the two model's predictability. These studies introduced variation in temperature (10, 23, and 30°C), humidity (20, 40, and 70%), and blood volume (unknown variation). Mock-up samples stored at different indoor places were likewise measured. The Lasso model showed good predictability of the test sets. Optimization of models and practicalities to incorporate this analysis into forensic practice is currently being investigated.