AARHUS
UNIVERSITY
BIOINFORMATICS RESEARCH CENTRE (BIRC)

Master's thesis

Predicting the ages of blood samples using targeted metabolomic data

10/1 2023

Sofie Agerbæk

Supervisor:

Palle Villesen,

Associate professor,

Bioinformatics Research Center at Aarhus University

Contents

Abstract

Determining the time since deposition of a blood stain at a crime scene is of great interest to forensic investigators as it sets a time frame for the crime committed and links a suspect to the scene at a certain time. At present many methods have been explored in the pursuit of developing a predictive tool to accurately estimate the age of a blood sample. Here we obtained targeted metabolomic data from blood samples of different ages. Clear age-dependant patterns were visible with principal component analysis, so we proceeded to implement and compare an array of predictive regression models for estimating the ages of the blood samples. With ensemble modelling we achieved errors in the range of ±6 hours for samples below 60 hours of age, while the errors for older samples, from 60 to 800 hours, were in the range of -30 to +70 hours. Classification of samples into the two intervals was implemented both with classifier and regression models, these obtained accuracies of 0.97 and 0.96 respectively. Data kept under varying temperatures, humidities, and locations obtained much higher errors and the predictions of our models on this data was deemed insensible. We concluded that the effect of external factors on the denaturation of blood samples is detrimental. Finally, a set of technical replicates was obtained and the clear batch effect between it and the original data was somewhat reduced through filtering based on metabolite ratio similarity.
We showed promising results for the use of metabolomics in blood age determination but must modify our experimental protocols, so we obtain similar data between batches. Furthermore, the effect of changing conditions must be thought into the modelling, to create robust models suitable for real-life crime scene evidence.

Introduction

Biological evidence, such as bloodstains, is detrimental to forensic investigations. It contains a wealth of imperative information; from personal information on the victim and suspect to pattern analysis to reconstruct the crime. One of the most important and at its start revolutionizing for forensics is genetic profiling, which is used to verify a suspect's identity and link them to the crime scene (Jobling and Gill, 2004). While genetic profiling is a very accurate way of linking a person to a crime scene, it is not informative about the time since evidence deposition. In some cases, suspects will claim to have been present at the crime scene at a time before the crime, where they left the biological evidence in question. To determine the age of the evidence is thus of great interest to investigators. But creating a new technique