Preprocessing Optimization of LCMS Data from Fingerprints

Solveig Østergaard

201407182

Master's thesis in Bioinformatics

Aarhus University, June 2020

Superviser: Palle Villesen, Associate Professor

Table of Contents

Abstract

Fingerprints are used extensively as forensic evidence, and being able to detect and understand the variation in the prints could give us a better chance of identifying individuals in for example criminal cases. One of the variations that could be interesting to study is the degradation of the prints and the possibility of predicting the deposition age for them. This could give us the possibility of establishing if individuals were at a location at a specific time. In this project, fingerprints were deposited at different times, and they were then analyzed with LCMS to detect the differences over time. The focus of the project was the effects of the preprocessing of the LCMS files and how to optimize this step to predict the age of the fingerprints as precisely as possible.

The XCMS package in R was used for the preprocessing, and a manual tuning of several parameters was made. Here, the new version of XCMS was used and three different optimizations were made: one where the best values of some of the predictors were combined, one where the same predictors were optimized sequentially, and one where the predictors were optimized sequentially using the QC samples for the optimization steps. The new version of XCMS turned out to generally perform better than the old version, and all optimizations performed better than a base version of the new XCMS, which means an optimization of the preprocessing step makes sense. The sequential optimization was performing with the best result, while the combined optimization did not perform as well as expected given individual optimizations of the parameters. The QC optimization did not perform as well as the others, but it might be the most interesting result since the optimization still had a considerable effect and the risk of overfitting is as good as nonexistent. Thereby, the QC optimization might be the optimization to use and further study in the future.