A novice's guide towards processing untargeted ^{1}H-NMR metabolomics data

Jessica van der Westhuizen,

North-West University, Potchefstroom, South Africa

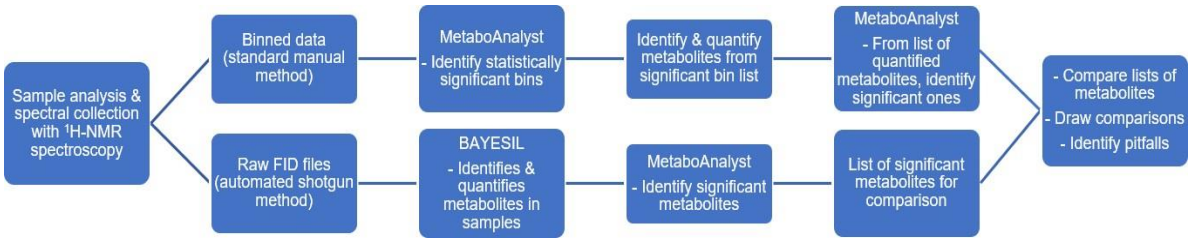

Figure 1: A summary of the approach followed in this study

Introduction: The processing of ^{1}H-NMR metabolomics data is difficult, and even more so when extracted from complex biological matrices such as urine. For this reason, the automation of ^{1}H-NMR spectral data profiling has been gaining attention, with numerous new online software rapidly emerging. For the purposes of this research, two approaches toward processing of untargeted ^{1}H-NMR metabolomics data were used and subsequently compared, to obtain a metabolite profile from highly complex treated and untreated TB urine samples.

Methodology: The samples used for the purposes of this research were TB urine samples before treatment (week 0), after 2 weeks of treatment and healthy control samples. These samples were prepared and analysed to obtain ^{1}H-NMR spectral data. The two data processing methods used in this study were: 1) The standard manual (binning) ^{1}H-NMR data processing method used at the Centre for Human Metabolomics at the NWU, and 2) an automated shotgun method using BAYESIL, an online ^{1}H-NMR data profiling tool. For both methods, the online metabolomics suite MetaboAnalyst was used to carry out statistical analyses. Both BAYESIL and MetaboAnalyst were selected as the online tools as they are well-established. A comparison was drawn between these approaches to understand the balance between user-expertise and high-throughput processing, as well as the resulting impact on the insight gained. The groups were purposefully selected to compare performance in the presence of exogeneous metabolites from pathogens and medication.

Results: Looking at the diseased TB profile (healthy controls vs week 0), 32 and 9 important metabolites were identified via the manual and BAYESIL methods, respectively, with 3 common metabolites. The assessment of TB treatment (week 0 vs week 2) identified 14 and 6 unique important metabolites via the manual and BAYESIL methods, respectively.

Discussion and conclusion: BAYESIL was less time-consuming and required little to no user-expertise when compared to the manual method, with some obvious drawbacks. BAYESIL does not have an algorithm designed for urine and its available library is still limited, hampering its ability to identify all metabolites in a sample. Hence, the metabolites identified were approximately 16 – 100% identifiable. The results therefore suggest that BAYESIL is not yet capable of producing a comprehensive and accurate metabolic profile for urine samples. That said, BAYESIL is well-suited for shotgun metabolic profiling to obtain a bird-eye view of the metabolic picture. This may be the only alternative when an NMR specialist is not available. In scenarios where a comprehensive metabolic profile is desired; novel or exogeneous metabolites are of interest; or the biological matrix is complex it is best to use the standard manual method. BAYESIL may be used in parallel to aid in the identification of some metabolites. During this study, some guideposts were noted to aid future novice users of ^{1}H-NMR metabolomics data.

Keywords: ^{1}H-NMR metabolomics; Untargeted; BAYESIL; Automated processing; Metabolite profiling

This work is based on research supported by the National Research Foundation of South Africa (Grant UID 130913)

National Research Foundation