Detection of biomarkers: KPC-3 *Klebsiella pneumoniae* method for identification a reference mass

Manuel J. Arroyo[1], Eva Gato[2], Marina Oviaño[2], Luis Mancera[1]

Abstract

Detection of Carbapenemase-Producing Enterobacteriae (CPE) is one of the main goals of microbiology laboratories, as the constitute a health alarm worldwide. Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) has emerged as a promising technology for the rapid and reliable identification of microorganism as bacteria. It could be a rapid and reliable tool for the correct identification of *Klebsiella pneumoniae* KPC-3 resistant strains. Identification of this resistance ability from non carbapenemase producing is priority to define the prognosis and the correct treatment besides infection control measures. The differentiation of this strain from other CPE like K. *pneumoniae* OXA-48 is important to define the epidemiology. Clover MS Data Analysis software is capable to help and find reference peaks which serve as biomarkers of these microorganism which can classify KPC-3 strains from the other ones. This document shows a workflow using the platform which able to classify the KPC-3 K. pneumoniae from the peak search section.

Introduction

Klebsiella pneumoniae is part of ESKAPE group like one of the six pathogens due to their increased resistance to commonly used antibiotics. K. *pneumoniae* have spread globally and represent a serious threat to public health since it´s one of a causative agent of many nosocomial infections nowadays. Rapid detection of these resistant strain has become in a priority in order to improve the handling these infections.

Anna F. Lau et al. tracked the bla_{kpc} carbapenemase gene-bearing in a plasmid in k.*pneumoniae* resistant strain which is associated to a characteristic peak on 11,109m/z in mass spectrometry spectra (Figure 1). The enzyme type of various resistant K. *pneumoniae* show a different prevalence in several geographic areas. In Europe, KPC-producing K. *pneumoniae* is endemic in Italy and Greece, while in other countries OXA-48 family carbapenemases are reported with overall lower prevalence. Nevertheless, KPC family has the most extensive distribution of all carbapenemases associated with Enterobacteriae family and the most clinically significant. The scope of this document is to show how the platform can find biomarkers, in this particular case, the KPC-3 representative MS peak to classify and differentiate KPC-3 K. *pneumoniae* strains from the OXA-48 strains.
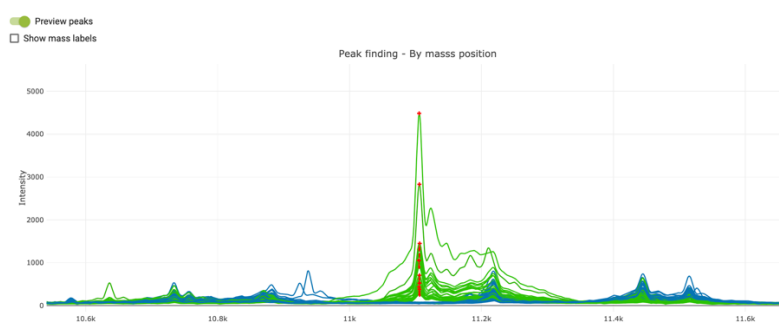

Figure 1. Preview of 11109m/z peak of the samples in Reproducibility test in CLOVER´s platform. It can be seen who the reference MS peak is only present in KPC-3 samples. Green: KPC-3 spectrums; Blue: OXA-48 spectrums.

Materials and methods

As an input data for search biomarkers we used n=82 samples from diverse hospitals (25 KPC-3 and 57 OXA-48 k. pneumoniae). These spectra were obtained by an "in target" extraction method, however there is evidence that other methods like full extraction work too for these analysis (ref). Extender?

Experimental

After creating the study and the experiment in the Clover platform with the samples, the "Reproducibility test" can work with the uploaded data. In this section, first of all, samples must be chosen and labeling per group, two in this case, one label for KPC-3 and other for OXA-48 strains. In the next step the platform allows to apply a noise reduction for all spectrum. For these particular samples, a tophat filter with 0,02 factor and Savitzky-Golay filter with 11 window length

and 3 as a polynomial order were applied as baseline subtraction and smoothing respectively. The mass filtering is the next option in this module. The spectrum obtained in this study were from 2000m/z to 20000m/z. In this case the range of interest is from 3000m/z to 12000m/z due in the lower part of the spectrum there are too much noise and the higher part haven´t relevance information in this example.

Once all the processing is done, we reach to the "Peak distribution" step where potential biomarkers can find.
At this point a mode for finding peaks can be chosen. Assuming we don´t know the peak of interest, we will choose a threshold mode. The optimal parameters for this search and these spectrums were 1000 ppm as linear mass tolerance and 0,005 as prominence. After the search, the platform will have found the peaks that have those characteristics. To finish this step, the peaks must be aligned between the spectrums in order to be correctly compared in the table that shows the platform. For this purpose, we set a 2000 ppm as linear mass tolerance in the options below of "Peak Distribution" field. Now, all peaks with well-defined characteristic and aligned are show in the peak distribution table.
In order to find our biomarker, we need to search a peak (or peaks) that differentiate between groups; either in frequency or intensity. In the "Groups pair comparison" we can select and order by an "appearance comparison" between our two groups (Table 1).

Mass ≑	KPC-3 - OXA-48 ▾
11104.96002	100.00
3690.6597	67.23
5548.22596	64.49
7402.23593	53.89
5247.93239	50.74
6513.66665	47.23
11872.37403	46.46
6134.42561	45.89
3588.90378	43.44
4195.14437	42.25

Table 1. Group pair Comparison table where KPC-3 and OXA-48 strains are compared peak by peak. The second column show the difference between the appearance in these groups for the mass in the first column. The 11104 mass has a difference of 100%.

Here we can see how the platform detect the same MS peak of interest that Anna F. Lau et al. 11,104m/z in our alignment case. The second

column show a different of 100% between the two groups, telling us that that peak is only present in KPC-3 strains and not in OXA-48.
Additionally, in the "Peak Distribution Table" we can search this peak in order to get more information and confirm the previous results (Table 2.)

Mass ≑	Number of spectra ≑	KPC-3			OXA-48		
		Appearance ⓘ ≑	CV ⓘ ≑	Mean ≑	Appearance ≑	CV ≑	Mean ≑
11104.96002	25	25/25 (100.0%) +	94.654 %	963.175	0/57 (0.0%)	-	-

Total samples: 82

Table 2. Peak distribution table after second alignment. Show the searched peak 11104m/z only. This peak is present in the 25 of 25 KPC-3 spectrum versus the non-appearance in OXA-48 group.

In addition, in the "preview peak" mode we can now visually search only our peak of interest too (Figure 1). In this case we can correctly classify these two groups only by one biomarker, but there are cases that we could need more than one. In these cases, the research must try identify these MS peaks ratio which these different tools.

Results

Once biomarker(s) is selected, we can generate a peak matrix with the samples. This time, "Find Peak" mode, we can select specifically our masses of interest in order to prove the efficacy of our biomarker to correctly classify the samples with the help of some platform´s algorithms like a SVM. If those analysis allow the classification of KPC-3, they can be saved as a prediction base for the input of new blind samples.

Conclusión

The CLOVER MS data analysis platform allow search and find MS peaks of interest which can serve as biomarkers for classify and identify correctly any samples. In this example, with this work-flow, the platform has found the same characteristic MS peak for KPC-3 as the study of Anna F. Lau et. al.