

Geena, a tool for MS spectra filtering, averaging and aligning

P. Romano¹✉, A. Profumo², R. Mangerini³, F. Ferri⁴, M. Rocco², F. Boccardo³, A. Facchiano⁵

¹Bioinformatics, IRCCS San Martino University Hospital - IST National Cancer Research Institute, Genova, Italy

²Biopolymers and Proteomics, IRCCS San Martino University Hospital - IST National Cancer Research Institute, Genova, Italy

³Medical Oncology B, IRCCS San Martino University Hospital - IST National Cancer Research Institute, Genova, Italy

⁴Department of Physics and Mathematics, Insubria University, Como, Italy

⁵Bioinformatics, Institute of Food Sciences, National Research Council, Avellino, Italy

Motivations

Mass spectrometry (MS), one of the most recent high-throughput technologies, produces a high volume of data. Many tools exist for MS data management, but little is available for the automation of related procedures. Geena is a new tool that aims at automating some of the fundamental steps involved in the analysis of m/z and abundance data from MALDI/TOF MS experiments. Geena was developed by taking into account the following assumptions:

a. in each spectrum, molecules are present in the form of different isotopic abundances

that can be summed together to give a total abundance value;

b. often, experimental data have to be normalized against an internal standard in order to obtain (semi) quantitative results;

c. experimental data can be affected by background noise. The selection of signals above a modulated threshold built on the spectra profile may be useful;

d. the analysis of sample replicates yields multiple spectra which are different because of marginal errors/changes in the experimental phase only.

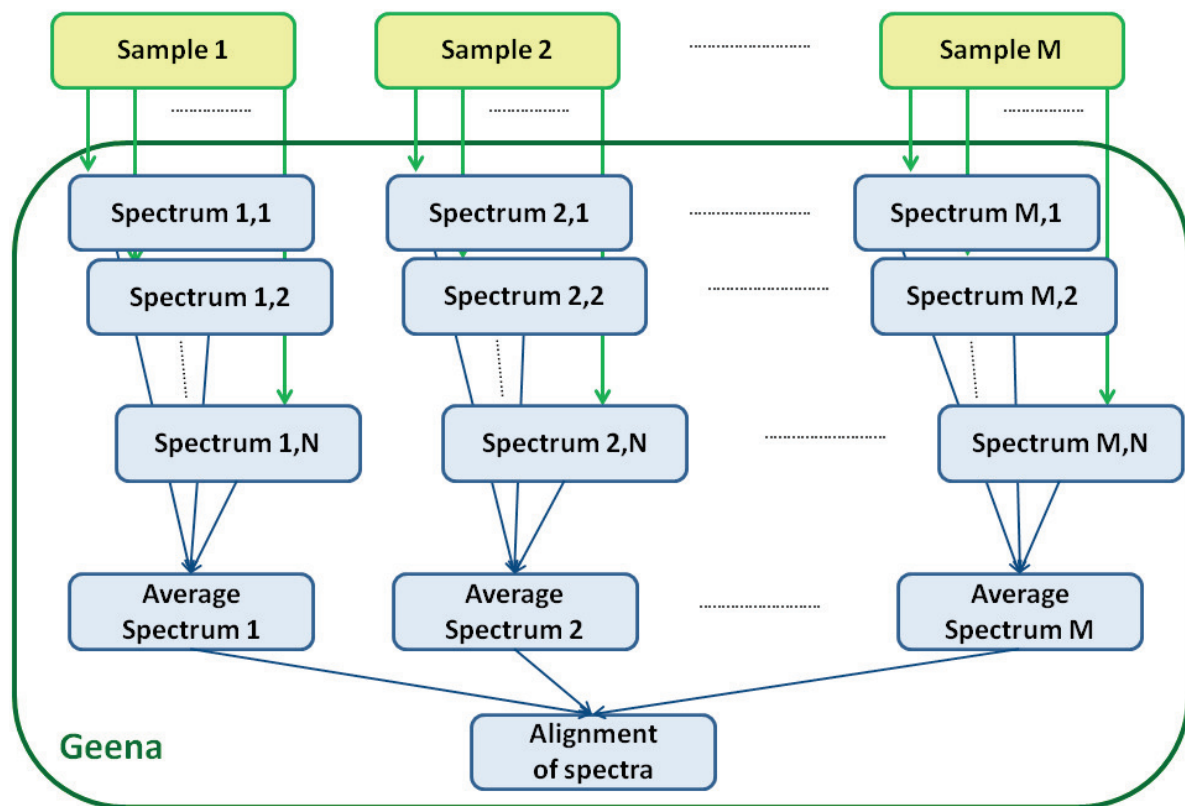


Figure 1. The figure reports an outline of the processing method in Geena.

An average spectrum representative of the sample may be defined by aligning these spectra along the m/z axis and computing mean intensity values from the corresponding abundances. In order to compare single or average spectra obtained from different samples the alignment along the m/z axis is required.

Methods

Geena was written in PHP and partially, for spectra alignment, in perl. Both input and output are managed in simple text files, usually having tab or comma delimited values. Such formats can easily be consumed by MS Excel or any other data management system. Data may be stored on the server in a MySQL database. The processing method is mainly heuristic and it is based on original algorithms. It includes the following steps: a) preprocessing of spectra replicates, which consists in isotopic peaks joining, normalization, and peak selection, b) computing average spectra for replicated analysis of samples, and c) alignment of average spectra.

Results

Geena is a public web server. The input consists in MALDI/TOF MS spectra. The data file is usually uploaded to the server and removed as soon as it has been used. It may include data from many samples, for each of which more spectra replicates can be provided. The output consists both in the averaged spectra from replicates and in the alignment of averaged spectra. The alignment is shown in the results page, while all results are available for downloading from the same page and sent by email, if a valid address is provided. Many parameters are defined. The analysis range is specified by indicating the lower and upper m/z values. The presence of a

normalization peak and its corresponding m/z value must be specified in order to normalize data. The threshold for peak selection is specified by providing abundance threshold values for the upper and lower limits of the analysis range. Further Intermediate thresholds may be specified, in which case a broken line is built by linear interpolation. An alternative method, that is based on data background estimation, is being added to Geena. Isotopic peaks of the same molecule are identified on the basis of the maximum allowed delta between them, i.e. the maximum deviation from expected values to consider a signal as an isotopic abundance of a given peak, and of the maximum number of isotopic replicates. In order to compute average spectra for a given sample, the maximum delta for aligning replicates, i.e. the maximum allowed deviation along the m/z axis between two signals belonging to replicates of the same sample to align them, and the minimum number of signals in replicates, that defines the minimum number of replicates that should contain a signal to include it in the average spectrum, can be specified. Similarly, the maximum delta for aligning average spectra and the minimum number of signals in average spectra can be specified to support alignment of average spectra. The method has been used for the analysis of long-term cryopreserved sera [1].

Availability

<http://bioinformatics.istge.it/geena/>

References

1. Mangerini R, Romano P et al. (2011) The application of atmospheric pressure MALDI to the analysis of long-term cryopreserved serum peptidome. *Analytical Biochemistry* 417, 174-181.