Developing population quarantine methods using Mass spectrometry(MS) data to overcome real time epidemic situations.

Mass spectrometry(MS) is a system used to recognise, distinguish and quantitate atoms in view of their mass-to-charge (m/z) proportion. Mass spectrometry (MS) investigation of proteins identification and evaluate atoms in basic and complex blends has turned out to be important over an expansive scope of fields and applications, including proteomics. High-throughput genomic and proteomic profiling harbour impressive use of MS methods in the systematic investigation of all metabolites present within a biologic system, which are used in biomarker development for many human diseases. These investigation are very helpful when a rapid identification of the sample is required. MS based techniques are capable of distinguishing between different biomarkers.

We have developed a methodology for High throughput MS protein identification and characterisation which can be applied in an epidemic situation. As there are no real epidemic data available big enough to meet the real life scenario, we have generated synthetic data using MSSimulator[1] and induced jitter and error in the generated data to match the real life data. After taking the initial seed samples from peptide atlas [2] the process consists of categorizing proteins by abundance into multiple simulated cohorts. After performing in-silico trypsin digestion to the generated peptides the simulator program(MS) generates (m/z, rt) peak data. The output data is jittered to produce a cohort and randoms peaks are added to induce noise in the data. We have applied the distributed version of clustering algorithm[3] MEDEA (M-Estimator with DEterministic Annealing) on the synthetic data to classify the corresponding population samples based on the biomarker selection. The experiment is concluded with the visualisation of the clusters to verify the clustering.

The methodology can be of great importance in the case of epidemic situation where the data can be quickly generated by Mass spectrometer and processed by the high-throughput clustering pipeline to classify the population based on biomarkers based on mz/rt peaks and later it can be used for effective implementation of quarantine methods.

1) MSSimulator: Simulation of Mass Spectrometry Data; Chris Bielow, Stephan Aiche, Sandro Andreotti, and Knut Reinert; Journal of Proteome Research **2011** 10 (7), 2922-2929; DOI: 10.1021/pr200155f
2) http://www.peptideatlas.org
3) Frühwirth, Rudolf, D. R. Mani, and Saumyadipta Pyne. "Clustering with position-specific constraints on variance: Applying redescending M-estimators to label-free LC-MS data analysis." BMC bioinformatics 12.1 (2011): 358.