



Statistical and Data Mining Analysis of Omics and Clinical Data for Cancer Research.[†]

Mario Cannataro,^a Giuseppe Agapito^a

Genomics, proteomics, and interactomics refer to the study of the genome, proteome and interactome of an organism. Such omics disciplines are gaining an increasing interest in the scientific community due to the availability of novel, high throughput platforms for the investigation of the cell machinery, such as mass spectrometry, microarray, next generation sequencing, that are producing an overwhelming amount of experimental omics data. On the other hand, the large volumes of omics data poses new challenges both for the efficient storage and integration of the data and for their efficient preprocessing and analysis. Moreover, both raw experimental data and derived information extracted by raw data are more and more stored in various databases spread all over the Internet, not fully integrated. Thus, managing omics data requires both support and spaces for data storing as well as algorithms and software pipelines for data preprocessing, analysis, and sharing. The resulting scenario comprises a set of methodologies and bioinformatics tools for the management and analysis of omics data stored locally or in geographically distributed biological databases. The talk introduces main omics data (e.g. gene expression and SNPs, mass spectra, protein-protein interactions), and describes some parallel and distributed bioinformatics tools for the preprocessing and statistical and data mining analysis of omics data, including those developed at the Bioinformatics Laboratory of the University Magna Graecia of Catanzaro (micro-CS, DMET-Analyzer, DMET-Miner, OSAnalyzer, coreSNP, GenotypeAnalytics, etc.). Some real cases regarding the statistical and data mining analysis of DMET SNPs datasets for pharmacogenomics studies in cancer research are also presented.

Riferimenti

- 1 M. Cannataro, B. Gaeta, A. Khan, Encyclopedia Of Bioinformatics and Computational Biology, Elsevier Science (2019).
- 2 M. Cannataro, P. Guzzi, Data Management of Protein Interaction Networks, Wiley (2011).
- 3 M. Arbitrio, M. T. Di Martino, F. Scionti, G. Agapito, P. Hiram Guzzi, M. Cannataro, P. Tassone, P. Tagliaferri, DMETTM (Drug Metabolism Enzymes and Transporters): a pharmacogenomic platform for precision medicine, *Oncotarget* 7 (33) (2016) 54028–54050. doi:10.18632/oncotarget.9927.
- 4 P. H. Guzzi, G. Agapito, M. Milano, M. Cannataro, Methodologies and experimental platforms for generating and analysing microarray and mass spectrometry-based omics data to support P4 medicine, *Briefings in bioinformatics* 17 (4) (2015) 553–561.
- 5 M. T. Di Martino, M. Arbitrio, P. H. Guzzi, M. Cannataro, P. Tagliaferri, P. Tassone, Experimental treatment of multiple myeloma in the era of precision medicine, *Expert Review of Precision Medicine and Drug Development* 1 (1) (2016) 37–51.
- 6 P. H. Guzzi, G. Agapito, M. T. Di Martino, M. Arbitrio, P. Tassone, P. Tagliaferri, M. Cannataro, DMET-analyzer: automatic analysis of Affymetrix DMET data, *BMC bioinformatics* 13 (1) (2012) 258.
- 7 G. Agapito, P. H. Guzzi, M. Cannataro, Dmet-miner: Efficient discovery of association rules from pharmacogenomic data, *Journal of biomedical informatics* 56 (2015) 273–283.

^a Data Analytics Research Center & Department of Medical and Surgical Sciences, University “Magna Graecia” of Catanzaro, 88100 Catanzaro, Italy email: cannataro@unicz.it, agapito@unicz.it

Creative Commons Attribution - Non commerciale - Condividi allo stesso modo 4.0 Internazionale

[†] presentato a @ITIM 2019 - 19° Congresso Nazionale Associazione Italiana di Telematica ed Informatica Medica 11-12 Novembre 2019, Matera/Potenza.