8th Workshop on Biomedical and Bioinformatics Challenges for Computer Science – BBC2015

Stefano Beretta[1,2], Mario Cannataro[3,4], and Riccardo Dondi[5]

[1] Dept. of Informatics, Systems and Communication, University of Milano - Bicocca, Milano - Italy
[2] Inst. for Biomedical Technologies, National Research Council, Segrate - Italy
[3] Dept. of Medical and Surgical Sciences, University Magna Græcia of Catanzaro, Catanzaro - Italy
[4] ICAR-CNR, National Research Council, Rende - Italy
[5] Dept. of Human and Social Sciences, University of Bergamo, Bergamo - Italy
stefano.beretta@itb.cnr.it, cannataro@unicz.it, riccardo.dondi@unibg.it

1   Introduction

Life science has been revolutionized by the new developed technologies which now produce an increased amount of complex data, especially influencing fields such as genomics, transcriptomics, and metagenomics. In fact, although the aims of bioinformatics were mainly devoted to the support for the analysis of biological and medical data, like the acquisition, storage, organization, and archiving of data, the recent changes in the nature of the information produced by these new emerging technologies opened new challenges. More precisely, the new needs have made the role of computer science (both theoretical and applied aspects) much more central and critical in all the bioinformatics research directions, requiring the development of new methods and approaches to analyze the new produced data. On the other hand, computational biology was focused on modelling and simulating biomedical processes and systems, with particular attention to the mathematical and computational aspects. Influenced by the advent of the developed technologies, these two disciplines are becoming more and more related each other. In order to tackle the growing complexity associated with emerging and future life science challenges, bioinformatics and computational biology researchers and developers need to explore, develop and apply novel computational concepts, methods, tools and systems. Many of these new approaches are likely to involve advanced and large-scale computing techniques, computational approaches, technologies and infrastructures such as: (i) high-performance architectures and systems (e.g., multicore, GPU); (ii) distributed computing (e.g. grid, cloud, peer-to-peer); (iii) computational simulation (mechanistic, stochastic, multi-model); (iv) algorithms (theoretical and experimental aspects); (v) applied bioinformatics (analysis pipelines, tools, applications); (vi) artificial and computational intelligence (machine learning, agents, evolutionary techniques, bio-inspired methods).

The eighth edition of the Workshop on *Biomedical and Bioinformatics Challenges for Computer Science* (BBC) aimed to bring together scientists in the fields of computer science, bioinformatics and computational biology to discuss the emerging and future directions in these

areas. As in the past, also this year the workshop has been organized in conjunction with the *International Conference on Computational Science* (ICCS).

The first session, *Bioinformatics Application for Medicine (I)*, comprised three full and two short papers. In the first paper *Robust Conclusions in Mass Spectrometry Analysis*, Zoppis et al. proposed a framework to provide robust decisions, when testing hypothesis over a case/control population of proteomic profiles. The approach is based on a graph model for the considered data, on a test of hypotheses of the graph properties and on a reference random graph model, for sources of uncertainties. It is applied to mass spectrometry profiles of the most common type of Renal Cell Carcinoma. In the paper *Modeling of Imaging Mass Spectrometry Data and Testing by Permutation for Biomarkers Discovery in Tissues*, Michal Marczyk et al. provided computational and statistical methods to investigate protein biomarkers. The proposed methodology includes known pre-processing tools, advanced mathematical modeling algorithms and permutation tests, which are applied to identify components differentiating among proteome profiles of tumor and surrounding tissues. The method allows to detect proteins/peptides specific to tumor area, normal epithelium, muscle and saliva gland regions, with high confidence. In the paper *Fuzzy indication of reliability in metagenomics NGS data analysis*, Milko Krachunov et al. introduced modified Hamming and Levenshtein distance functions, which are used as drop-in replacement in NGS analysis procedures. The proposed approach is based on fuzzy sets of reliable bases or an equivalent fuzzy logic, and uses a combination of read similarity and neural networks. In the paper *Pairwise genome comparison workflow in the Cloud using Galaxy*, Oscar Torreno et al. presented the integration of a tools management platform (Galaxy) and a Cloud infrastructure. This integration provides a user-friendly platform that facilitates end-users in performing their experiments, and guarantees the features of a Cloud environment (authentication, security and big-data transfer mechanisms). Pairwise and multiple genome comparison tools are integrated to provide evidence of the usefulness of the proposed approach. In the paper *Iterative Reconstruction from Few-View Projections*, Liubov A. Flores et al. proposed an LSQR-approach, with soft threshold filtering algorithm for few-view image reconstruction based on GPUs. The method helps in dealing with 3D reconstruction of computed tomography imaging.

The second session, *Bioinformatics Application for Medicine (II)*, comprised five full papers about computational methods for bioinformatics analysis of medical data. In the first paper *GoD: An R-Package based on Ontologies for Prioritization of Genes with respect to Diseases*, Cannataro et al. presented an algorithm, called GoD (Gene ranking based On Diseases), to rank a given set of genes exploiting ontology annotations. Genes are ordered based on the semantic similarity computed between the annotations of each gene and those describing the selected disease. The algorithm is applied to Human Phenotype Ontology (HPO), Gene Ontology (GO) and Disease Ontology (DO). In the paper *Large Scale Comparative Visualisation of Regulatory Networks with TRNDiff*, Xin-Yi Chua et al. presented an open source system, called TRNDiff, that supports the comparative analysis and visualization of Transcriptional Regulatory Networks (TRNs) belonging to different genomes. The presented approach allows the rapid identification of functional variations within species, as shown by a multiple TRN analysis of the Fur iron-uptake system of Yersinia. In the paper *Graph Theoretical Analysis of Clarkson Disease GWAS Epistatic Interactions*, Alex Upton et al. provided a graph approach for the study of SNP and gene interaction networks constructed for a Clarkson disease GWAS. The paper identified several genes and SNPs that are potentially involved in Clarkson disease with respect to traditional single SNP analysis. In the paper *Multiple structural clustering of bromodomains of the bromo and extra terminal (BET) proteins highlights subtle differences in their structural dynamics and acetylated leucine binding pocket*, Suryani Lukman et al. con-

sidered the functions of BET proteins, epigenetic readers whose deregulation results in cancer and inflammation. The paper shows that, although BET proteins (BRD2, BRD3, BRD4 and BRDT) are globally similar, there exist subtle differences in the structural dynamics, that could be related to non-subtle functional effects. By applying multiple structural clustering methods, the paper identified representative structures of BET proteins, which are potentially useful for developing potential therapeutic agents. In the paper *Parallel Tools for Simulating the Depolarization Block on a Neural Model*, Salvatore Cuomo et al. presented a computational neural framework to investigate the depolarization block of a CA1 pyramidal cell of hippocampus. An efficient simulation of the framework is provided by means of parallel tools, providing an efficient model simulation.

The third session *Medical Informatics Applications and Tools* comprised four full papers and one oral presentation about medical application of informatics. In the first paper *Using visual analytics to support the integration of expert knowledge in the design of medical models and simulations*, Philippe J. Giabbanelli and Piper J. Jackson described how a Visual Analytics (VA) approach can be useful when dealing with conceptual stage medical models. Different case studies were considered and analyzed using different modelling tools (e.g., system dynamics or network modelling), hence emphasizing the relevance of such an approach for different techniques. In the second paper *Mining Mobile Datasets to Enable the Fine-Grained Stochastic Simulation of Ebola Diffusion*, Nicholas Vogel et al. proposed a model for stochastic simulation of Ebola diffusion. Simulations are based on the interactions of multi-scale models, describing different aspects that could be involved in the Ebola diffusion, including models for cellular level, individual level and workplace and family level, population level, at the abstract governmental policy level. The simulation provided results in alignment with Centers for Disease Control and Prevention predictions. In the third paper *A Novel $O(n)$ Numerical Scheme for ECG Signal Denoising*, Salvatore Cuomo et al. presented a novel numerical scheme for ECG Signal denoising that belongs to Infinite Impulse Response (IIR) noise reduction algorithms, hence with low computational requirements. The presented scheme does not require a direct application of the Fast Fourier Transform and allows for an easy implementation on mobile computing devices. In the fourth paper *Syncytial Nature as Basis for Diversity in Spike Shapes and their Propagation in Detrusor Smooth Muscle*, Shailesh Appukuttan and Rohit Manchanda presented a three-dimensional electrical model of smooth muscle syncytium, based on the compartmental modeling technique. Moreover, an investigation of the origin of the various spike shapes is provided. In the oral presentation *The Potential of Machine Learning for Epileptic Seizures Prediction*, Antonio Dourado et al. discuss the potential of Machine Learning for predicting epileptic seizures. They discuss the use of classifiers to detect the pre-ictal time, happening some minutes before the seizure.

The program co-chairs sincerely thank the Program Committee members and external Reviewers, for the time and expertise they put into the reviewing work, the ICCS organization, for providing the opportunity to arrange the BBC workshop in conjunction with the ICCS 2015 conference, and all the workshop attendees.

June 2015

Stefano Beretta
Mario Cannataro
Riccardo Dondi