

Preface

We are proud to present the set of final accepted full papers for the third edition of the IWBBIO conference “International Work-Conference on Bioinformatics and Biomedical Engineering” held in Granada (Spain) during April 26–28, 2017.

The IWBBIO 2017 (International Work-Conference on Bioinformatics and Biomedical Engineering) seeks to provide a discussion forum for scientists, engineers, educators, and students about the latest ideas and realizations in the foundations, theory, models, and applications for interdisciplinary and multidisciplinary research encompassing the disciplines of computer science, mathematics, statistics, biology, bioinformatics, and biomedicine.

The aim of IWBBIO is to create a friendly environment that could lead to the establishment or strengthening of scientific collaborations and exchanges among attendees, and, therefore, IWBBIO 2017 solicited high-quality original research papers (including significant work in progress) on any aspect of bioinformatics, biomedicine, and biomedical engineering.

The following topics were especially encouraged: new computational techniques and methods in machine learning; data mining; text analysis; pattern recognition; data integration; genomics and evolution; next-generation sequencing data; protein and RNA structure; protein function and proteomics; medical informatics and translational bioinformatics; computational systems biology; modelling and simulation and their application in the life science domain, biomedicine, and biomedical engineering. The list of topics in the successive call for papers also evolved, resulting in the following list for the present edition:

1. **Computational proteomics.** Analysis of protein–protein interactions. Protein structure modelling. Analysis of protein functionality. Quantitative proteomics and PTMs. Clinical proteomics. Protein annotation. Data mining in proteomics.
2. **Next-generation sequencing and sequence analysis.** De novo sequencing, re-sequencing, and assembly. Expression estimation. Alternative splicing discovery. Pathway analysis. Chip-seq and RNA-Seq analysis. Metagenomics. SNPs prediction.
3. **High performance in bioinformatics.** Parallelization for biomedical analysis. Biomedical and biological databases. Data mining and biological text processing. Large-scale biomedical data integration. Biological and medical ontologies. Novel architecture and technologies (GPU, P2P, Grid etc.) for bioinformatics.
4. **Biomedicine.** Biomedical Computing. Personalized medicine. Nanomedicine. Medical education. Collaborative medicine. Biomedical signal analysis. Biomedicine in industry and society. Electrotherapy and radiotherapy.
5. **Biomedical engineering.** Computer-assisted surgery. Therapeutic engineering. Interactive 3D modelling. Clinical engineering. Telemedicine. Biosensors and data acquisition. Intelligent instrumentation. Patient monitoring. Biomedical robotics. Bio-nanotechnology. Genetic engineering.

6. **Computational systems for modelling biological processes.** Inference of biological networks. Machine learning in bioinformatics. Classification for biomedical data. Microarray data analysis. Simulation and visualization of biological systems. Molecular evolution and phylogenetic modelling.
7. **Health care and diseases.** Computational support for clinical decisions. Image visualization and signal analysis. Disease control and diagnosis. Genome-phenome analysis. Biomarker identification. Drug design. Computational immunology.
8. **E-health.** E-health technology and devices. E-health information processing. Telemedicine/e-health application and services. Medical image processing. Video techniques for medical images. Integration of classical medicine and e-health.

After a careful peer-review and evaluation process (each submission was reviewed by at least two, and on average 3.2, Program Committee members or additional reviewers), 120 papers were accepted for oral, poster, or virtual presentation, according to the recommendations of the reviewers and the authors' preferences, and to be included in the LNBI proceedings.

During IWBBIO 2017, several Special Sessions were carried out. Special Sessions are a very useful tool to complement the regular program with new and emerging topics of particular interest for the participating community. Special Sessions that emphasize multi-disciplinary and transversal aspects as well as cutting-edge topics were especially encouraged and welcomed, and in IWBBIO 2017 they were the following:

– **SS1: Advances in Computational Intelligence for Critical Care**

Decision-making in health care in clinical environments is often made on the basis of multiple parameters and in the context of patient presentation, which includes the setting and the specific conditions related to the reason for admission and the procedures involved. The data used in clinical decision-making may originate from manifold sources and at multiple scales: devices in and around the patient, laboratory, blood tests, omics analyses, medical images, and ancillary information available both prior to and during the hospitalization.

Arguably, one of the most data-dependent clinical environments is the intensive care unit (ICU). The ICU environment cares for acutely ill patients. Many patients within ICU environments, and particularly surgical intensive care units (SICU), are technologically dependent on the life-sustaining devices that surround them. Some of these patients are indeed dependent for their very survival on technologies such as infusion pumps, mechanical ventilators, catheters and so on. Beyond treatment, assessment of prognosis in critical care and patient stratification combining different data sources are extremely important in a patient-centric environment.

With the advent and quick uptake of omics technologies in critical care, the use of data-based approaches for assistance in diagnosis and prognosis becomes paramount. New approaches to data analysis are thus required, and some of the most interesting ones currently stem from the fields of computational intelligence (CI) and machine learning (ML). This session is particularly interested in the proposal of novel CI and ML approaches and in the discussion of the challenges for the application of the existing ones to problems in critical care.

Topics that are of interest to this session include (but are not necessarily limited to):

- Novel applications of existing CI and ML and advanced statistical methods in critical care
- Novel CI and ML techniques for critical care
- CI and ML-based methods to improve model interpretability in a critical care context, including data/model visualization techniques
- Novel CI and ML techniques for dealing with non-structured and heterogeneous data formats in critical care

Organizers: Dr. Alfredo Vellido, PhD, Department of Computer Science, Universitat Politècnica de Catalunya, BarcelonaTECH (UPC), Barcelona (Spain). Dr. Vicent Ribas, eHealth Department, EURECAT Technology Centre of Catalonia, Barcelona, Barcelona (Spain).

– SS2: Time-Lapse Experiments and Multivariate Biostatistics

Biological samples are evolving in time, phases, periods, behavior. To be able to understand the dynamics, we need to perform time lapse experiments. Today's technique and measurement devices allow us to monitor numerous parameters in semi-controlled environments during the experiment. The increase of measured data is enormous. The interpretation requires both qualitative and quantitative analysis. There are useful methods of biostatistics, multivariate data analysis, and artificial intelligence, namely, neural networks, genetic algorithms, and agent-based modeling, respectively.

In this special section we will provide a discussion on broad examples from time-lapse experimental design through information and data acquisition, using methods of bioinformatics, biophysics, biostatistics, and artificial intelligence. The aim of this section is to present the possible increase in data interpretation and related methods.

Organizer: Dr. Jan Urban, Laboratory of Signal and Image Processing, Institute of Complex Systems, Faculty of Fisheries and Protection of Waters, University of South Bohemia.

– SS3: Half-Day GATB Tutorial. The Genome Analysis Toolbox with de Bruijn Graph

The GATB programming day is an educational event organized by the GATB team. This free event is open to everyone who is familiar in C++ programming and wants to learn how to create NGS data analysis software.

The tutorial has a focus on the high-performance GATB-Core library and is taught by the developers of this library.

During this half-day tutorial, some of the following topics are explored:

- A theoretical introduction to GATB: the basic concepts.
- GATB-Core practical coding session 1: I/O operations on reads files.
- The GATB de Bruijn graph API.
- GATB-Core practical coding session 2: k-mer and graph APIs in action.
- GATB-Core practical coding session 3: writing a short read corrector tool.
- Q and A session: obtain answers from GATB experts.

Organizers: Dr. Dominique Lavenier, GenScale Team Leader, Inria/IRISA, Campus de Beaulieu, Rennes, France.

Dr. Patrick Durand, Inria, Genscale Team, Campus de Beaulieu, Rennes, France.

– **SS4: Medical Planning: Operating Theater Design and Its Impact on Cost, Area, and Workflow**

The design of operating rooms is one of the most complicated tasks of hospital design because of its characteristics and requirements. Patients, staff, and tools should have determined passes through the operating suite. Many hospitals assume that the operating suite is the most important unit in the hospital because of its high-revenue. Arch design of these suites is a very critical point in solving an optimized problem in spaces, workflow of clean, dirty, and patient in/out in addition to staff together with their relations with adjacent departments. In this session, we illustrate the most common designs of operating suites and select the most suitable one to satisfy the effectiveness of the operating suite, maximizing throughput, minimizing the costs, and decreasing the required spaces related to available resources/possibilities. The design should comply with country guidelines, infection control rules, occupational safety and health, and satisfy the maximum benefits for patients and staff. A comparative study was performed on 15 hospitals and it recorded that the single input–output technique is the best design.

Motivation and objectives for the session: Operating suite design is a very critical task owing to its impact. Biomedical engineers should participate in the design and review the workflow and available functions.

Organizer: Dr. Khaled El-Sayed, Department of Electrical and Medical Engineering at Benha University, Egypt

– **SS5: Challenges Representing Large-Scale Biological Data**

Visualization models have been shown to be remarkably important in the interpretation of datasets across many fields of study. In the context of bioinformatics and computational biology various tools have been proposed to visualize molecular data and help understand how biological systems work. Despite that, several challenges still persist when faced with complex and dynamic data and major advances are required to correctly manage the multiple dimensions of the data.

The aim of this special session is to bring together researchers to present recent and ongoing research activities related to advances in visualization techniques and tools, focused on any major molecular biology problem, with the aim of allowing for the exchange and sharing of proposed ideas and experiences with novel visualization metaphors.

Topics of interest include:

- Genome and sequence data
- Omics data (transcriptomics, proteomics, metabolomics)
- Biological networks and pathways
- Time-series data
- Biomedical ontologies
- Macromolecular complexes
- Phylogenetic data

- Biomarker discovery
- Integration of image and omics data for systems biology
- Modeling and simulation of dynamical processes

Organizer: Prof. Joel P. Arrais FCTUC – University of Coimbra, Portugal.

– **SS6: Omics of Space Travelled Microbes – Bioinformatics and Biomedical Aspects**

The National Research Council (NRC) Committee for the Decadal Survey on Biological and Physical Sciences in Space reported that “microbial species that are uncommon, or that have significantly increased or decreased in number, can be studied in a “microbial observatory” on the International Space Station (ISS).” As part of the microbial observatory effort the NRC decadal survey committee suggested that NASA should: “(a) capitalize on the technological maturity, low cost, and speed of genomic analyses and the rapid generation time of microbes to monitor the evolution of microbial genomic changes in response to the selective pressures present in the spaceflight environment; (b) study changes in microbial populations from the skin and feces of the astronauts, plant and plant growth media, and environmental samples taken from surfaces and the atmosphere of the ISS; and (c) establish an experimental program targeted at understanding the influence of the spaceflight environment on defined microbial populations.”

The proposed session discusses state-of-the-art molecular techniques, bioinformatics tools, and their benefit in answering the astronauts and others who live in closed systems.

Organizer: Dr. Kasthuri Venkateswaran, Senior Research Scientist, California Institute of Technology, Jet Propulsion Laboratory, Biotechnology and Planetary Protection Group, Pasadena, CA

– **SS7: Data-Driven Biology – New Tools, Techniques, and Resources**

Advances in sequencing techniques have accelerated data generation at diverse regulatory levels in an unprecedented way. The challenge now is to integrate these data to understand regulation at a systems level. As the sequencing technologies evolve, new tools and resources follow, revealing new aspects of complex biological systems.

This special session brings together experts from computational biology and machine learning to present recent advances in the development of new tools and resources using next-generation sequencing data including novel emerging fields such as single-cell transcriptomics. The session features an invited speaker and three/four short talks. To promote emerging leaders of the field, we select invited speakers who have gained their independence in recent years.

Organizer: Dr. Joshi Anagha, Division of Developmental Biology at the Roslin Institute, University of Edinburgh

– **SS8: Smart Sensor and Sensor-Network Architectures**

There is a significant demand for tools and services supporting rehabilitation, well-being and healthy life styles while reducing the level of intrusiveness as well as increasing real-time available and reliable results. For example, self-monitoring

applications need to be improved to move beyond tracking exercise habits and capture a more comprehensive digital footprint of human behavior. This session focuses on primary parameter capturing devices and networks demonstrating advances in sensor development including a customized algorithmic shell research to support diagnostic decisions. Target domains are, for example, continuous differentiating between mental and physical stress, blood pressure monitoring, sleep quality monitoring, HRV etc.

Organizers: Prof. Dr. Natividad Martinez, Internet of Things Laboratory, Reutlingen University, Germany.

Prof. Dr. Juan Antonio Ortega, University of Seville, Spain.

Prof. Dr. Ralf Seepold, Ubiquitous Computing Lab, HTWG Konstanz, Germany.

– **SS9: High-Throughput Bioinformatic Tools for Genomics**

Genomics is concerned with the sequencing and analysis of an organism's genome. It is involved in the understanding of how every single gene can affect the entire genome. This goal is mainly afforded using the current, cost-effective, high-throughput sequencing technologies. These technologies produce a huge amount of data that usually require high-performance computing solutions and opens new ways for the study of genomics, but also transcriptomics, gene expression, and systems biology, among others. The continuous improvements and broader applications on sequencing technologies is producing an ongoing demand for improved high-throughput bioinformatics tools.

Therefore, we invite authors to submit original research, new tools or pipelines, or their update, and review articles on topics helping in the study of genomics in the wider sense, such as (but not limited to):

- Tools for data pre-processing (quality control and filtering)
- Tools for sequence mapping
- Tools for de novo assembly
- Tools for quality check of sequence assembling
- Tools for the comparison of two read libraries without an external reference
- Tools for genomic variants (such as variant calling or variant annotation)
- Tools for functional annotation: identification of domains, orthologues, genetic markers, controlled vocabulary (GO, KEGG, InterPro)
- Tools for biological enrichment in non-model organisms
- Tools for gene expression studies
- Tools for Chip-Seq data
- Tools for “big-data” analyses
- Tools for handling and editing complex workflows and pipelines
- Databases for bioinformatics

Organizers: Prof. M. Gonzalo Claros, Department of Molecular Biology and Biochemistry, University of Málaga, Spain.

Dr. Javier Pérez Florido, Bioinformatics Research Area, Fundación Progreso y Salud, Seville, Spain.

– **SS10: Systems Biology Approaches to Decipher Long Noncoding RNA–Protein Associations**

Long noncoding RNAs (lncRNAs) make up large amounts of the RNA and total genomic repertoire. Studies on the functional characterization of lncRNAs have resulted in data on interactions with their RNA peers, DNA, or proteins. Although there has been an increase in evidence on the link between lncRNAs and diverse human diseases, there is a dearth of lncRNA–protein association studies. Additionally, existing methods do not provide theories about the possible molecular causes of such associations linking to diseases. How such regulatory interactions between classes of lncRNAs and proteins would have a significant influence on the organism and disease remains a challenge. A good number of bioinformatics approaches have arisen in the recent past exploring these challenges. The idea behind this session is to bring together the wide gamut of researchers who have worked on these methods across different organisms.

The following are the sub-topics of the proposed session, which we would like to call for papers.

- lncRNAs in genomes: annotation and curation
- lncRNA–protein interactions leading to important diseases: systems Biology approaches
- Identifying lncRNAs with respect to their mechanism and dysregulation in diseases
- lncRNA databases and webserver
- Machine learning approaches and prediction servers

Organizer: Prashanth Suravajhala, PhD, Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, India

– **SS11: Gamified Rehabilitation for Disabled People**

Gamification is a hot topic in many areas as it aims at motivating people to do things driven by different innate needs like the wish to accomplish tasks, to compete against others, or to gain something. These and other motivators are efficiently applied in computer games and could be extraordinarily useful in ensuring that patients perform their daily exercises regularly and have fun.

The idea of exergames (exercise games) is not new; the literature reveals that much effort has already been made and with great success. Nevertheless, most applications have been developed for special problems or diseases (i.e., stroke, parkinson, cerebral palsy etc.) and are not generally applicable. In general, people suffering from severe disabilities and chronic diseases are rarely addressed as a target group. Also, the focus is generally set on the medical achievements, which is correct, but the next step would be to enhance the fun factor because no tool is of much use if the patient is not using it because of boredom or demotivation.

The objective of this special session is to gather new ideas about the combination of need and fun, i.e., find ways to create exercise platforms that fit everybody's needs, provide access to the therapist for monitoring and configuration, while the patients benefit physically and mentally when having a good time.

Target groups would be people of:

- All ages, while focusing on younger people, who can be involved more easily but are less addressed in the literature
- All diseases, while focusing on chronic illness and severe disabilities (e.g., muscle dystrophies and atrophies)

The contributions should show advances in at least one of the following areas:

- Adaptability to users with all kinds of problems (e.g., possibility to configure the limbs used to play or playing with facial movements, wheelchair and standing modes, coping with muscle weaknesses etc.)
- Implementation of gaming techniques and special motivators
- Physical or mental exercises, aimed at rehabilitation or daily practice
- Understanding the users, awareness of their level of motivation, fatigue or progress and react accordingly

Organizer: Dr. Martina Eckert, Associate Professor at University of Madrid, Spain.

– **SS12: Modelling of Glucose Dynamics for Diabetes**

Diabetes is the eighth most common cause of death, while its treatment relies on technology to process continuously measured glucose levels.

Organizer: Dr. Tomas Koutny, Faculty of Applied Sciences, University of West Bohemia

– **SS13: Biological Network Analysis in Multi-omics Data Integration**

In many biological applications, multiple data types may be produced to determine the genetics, epigenetics, and microbiome affecting gene regulation and metabolism. Although producing multiple data types should provide a more complete description of the processes under study due to multiple factors such as study design (synchronization of data production, number of samples, varying conditions), the analysis may leave more unfulfilled promises than synergy expected from the wealth of data.

In this session, some of the following challenges are addressed:

- How to conduct meaningful meta-analysis on historical data.
- How to use biological knowledge (represented in reproducible and interoperable manner) in the analysis of large and sparse data sets more effectively.
- How to fill the gap between hypothesis-driven mechanistic studies, e.g., applying modelling to very well studied biochemical processes and data-driven hypothesis-free approaches. How omics data can help.
- Beyond meta-transcriptomics and metagenomics: integration and interpretation of microbiome and host data.

We would like to bring together communities concerned with these topics to present state-of-the-art and current cutting-edge developments, preferably work under construction or published within the past year.

- Objective 1: presentation and discussion of newest methods
- Objective 2: round-table discussions on the topics highlighted above and other related topics suggested by session participants

An additional topic that does not fit the proposed session but that I would love to see addressed is: How to improve open access to the data that is not next-generation sequencing (e.g., metabolomics, proteomics, plant phenotyping). For this an active participation of journal editors would be necessary to discuss opportunities to change journal publication policies.

Organizer: Dr. Wiktor Jurkowski, Jurkowski Group, Earlham Institute, Norwich Research Park, UK

– **SS14: Oncological Big Data and New Mathematical Tools**

Current scientific methods produce various omics data sets covering many cellular functions. However, these data sets are commonly processed separately owing to limited ways in how to connect different omics data together for a meaningful analysis. Moreover, it is currently a problem to integrate such data into mathematical models. We are entering the new era of biological research where the main problem is not to obtain the data but to process and analyze them. In this regard, a strong mathematical approach can be very effective (see J. Gunawardena's essay "Models in biology: accurate descriptions of our pathetic thinking," BMC Biology 2014, 12:29).

In this Special Session we focus on big data (omics and biological pathways) related to oncological research. General biological processes that are relevant to cancer can also be studied. Mathematical tools basically mean statistical learning (data mining, inference, prediction), modeling, and simulation. We want to place a special emphasis on causality. Closely tied to mathematical tools, efficient computational tools can be considered.

Organizers: Dr. Gregorio Rubio, Instituto de Matematica Multidisciplinar, Universitat Politecnica de Valencia, Valencia, Spain. Dr. Rafael Villanueva, Instituto de Matematica Multidisciplinar, Universitat Politecnica de Valencia, Valencia, Spain.

In this edition of IWBBIO, we were honored to have the following invited speakers:

1. Prof. Roderic Guigo, Coordinator of Bioinformatics and Genomics at Centre de Regulacio Genomica (CRG). Head of the Computational Biology of RNA Processing Group. Universitat Pompeu Fabra, Barcelona, Spain
2. Prof. Joaquin Dopazo, Director of the Computational Genomics Department, Centro de Investigación Príncipe Felipe- CIPF, Valencia, Spain
3. Prof. Jose Antonio Lorente, Director of Centre for Genomics and Oncological Research (GENYO). Professor of Legal and Forensic Medicine, University of Granada, Spain

It is important to note, that for the sake of consistency and readability of the book, the presented papers are classified under 16 chapters. The organization of the papers is in two volumes arranged following the topics list included in the call for papers. The first

volume (LNBI 10208), entitled “Advances in Computational Intelligence: Part I” is divided into seven main parts and includes the contributions on:

1. Advances in computational intelligence for critical care
2. Bioinformatics for health care and diseases
3. Biomedical engineering
4. Biomedical image analysis
5. Biomedical signal analysis
6. Biomedicine
7. Challenges representing large-scale biological data

The second volume (LNBI 10209), entitled “Advances in Computational Intelligence: Part II” is divided into nine main parts and includes the contributions on:

1. Computational genomics
2. Computational proteomics
3. Computational systems for modelling biological processes
4. Data-driven biology: new tools, techniques, and resources
5. E-health
6. High-throughput bioinformatic tools for genomics
7. Oncological big data and new mathematical tools
8. Smart Sensor and sensor-network architectures
9. Time-lapse experiments and multivariate biostatistics

This fifth edition of IWBBIO was organized by the Universidad de Granada together with the Spanish Chapter of the IEEE Computational Intelligence Society. We wish to thank to our main sponsor and the Faculty of Science, Department of Computer Architecture and Computer Technology and CITIC-UGR, from the University of Granada, for their support and grants. We also wish to thank the Editors-in-Chief of different international journals for their interest in having special issues with the best papers of IWBBIO.

We would also like to express our gratitude to the members of the different committees for their support, collaboration, and good work. We especially thank the local Organizing Committee, Program Committee, the reviewers, and special session organizers. Finally, we want to thank Springer, and especially Alfred Hofmann and Anna Kramer, for their continuous support and cooperation.

April 2017

Ignacio Rojas
Francisco Ortuño

Bioinformatics and Biomedical Engineering
5th International Work-Conference, IWBBIO 2017,
Granada, Spain, April 26–28, 2017, Proceedings, Part II
Rojas, I.; Ortuño, F. (Eds.)
2017, XXXVI, 741 p. 275 illus., Softcover
ISBN: 978-3-319-56153-0