ProteINSIDE: a web service to computerize an in-depth analysis of functions, sequences, secretions and interactions for proteins

Nicolas Kaspric, Matthieu Matthieu.Reichstadt@inrae.Fr Reichstadt, Brigitte B. Picard, Muriel Bonnet, Jérémy Tournayre

Proceedings

IWBBIO 2014

International Work-Conference on Bioinformatics and Biomedical Engineering

Granada
April, 7-9 2014

volumen 1

Proceedings
IWBBIO 2014
International Work-Conference on Bioinformatics and Biomedical Engineering
volumen 1

Editors and Chairs
Francisco Ortuño
Ignacio Rojas

Deposito Legal: 978- 84-15814-84-9
I.S.B.N: GR 738/2014
Edita e imprime: Copicentro Granada S.L

Preface

We are proud to present the set of final accepted papers for the second edition of the IWBBIO conference "International Work-Conference on Bioinformatics and Biomedical Engineering" held in Granada (Spain) during April 7-9, 2014.

The IWBBIO 2014 (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 disciplines of computer science, mathematics, statistics, biology, bioinformatics, and biomedicine.

The aims of IWBBIO 2014 is to create a friendly environment that could lead to the establishment or strengthening of scientific collaborations and exchanges among attendees, and therefore, IWBBIO 2014 solicited high-quality original research papers (including significant work-in-progress) on any aspect of Bioinformatics, Biomedicine and Biomedical Engineering.

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 life science domain, biomedicine and biomedical engineering were especially encouraged. The list of topics in the successive Call for Papers has 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, resequencing 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,...) 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**. EComputer-assisted surgery. Therapeutic engineering. Interactive 3D modelling. Clinical engineering. Telemedicine. Biosensors and data acquisition. Intelligent instrumentation. Patient Monitoring. Biomedical robotics. Bionanotechnology. 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. **Healthcare 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 2, and on the average 2.8, program committee members or additional reviewer), 197 papers were accepted for oral, poster or virtual presentation, according to the recommendations of reviewers and the authors' preferences.

During IWBBIO 2014 several Special Sessions will be carried out. Special Sessions will be a very useful tool in order to complement the regular program with new and emerging topics of particular interest for the participating community. Special Sessions that emphasize on multi-disciplinary and transversal aspects, as well as cutting-edge topics are especially encouraged and welcome, and in this edition of IWBBIO 2014 are the following:

1. **SS1: Multi-biomarker and informatics in cancer diagnosis** The session will discuss the join research between Hospital Doctors and Informatic Scientists. With the development of life science, the clinical laboratory could provide much more test for patients than before, currently this use reference interval or cut off values as clinical diagnostic standred which could not significantly improve the specificity and sensitivity simultaneously. If the informatic tools could be used to combined analysis multiple clinical test and biomarkers, it will be enhance the using of clinical information from medical laboratory and also could stimulate the translational medicine studies of "omic" technology and theory.
Organizer: Dr. Prof. Yaping Tian, Department of Clinical Biochemistry, Chinese PLA General Hospital, Beijing (China).

2. **SS2: Discovery of non-coding and structured RNAs** Non-coding (nc)RNAs are emerging as some of the most versatile and important biological molecules in the cell. They can act both in cis and in trans to mediate functions as diverse as catalysis, metabolite sensing, regulation of gene expression, epigenetics, chromatin stability, splicing, and more. The explosion of RNA sequences generated by next-gen sequencing and accumulating in various databases represents a massive, and relatively un-explored, "New World" of ncRNAs. The identification of novel ncRNAs from these data is a crucial, but challenging, problem for bioinformatics. Many ncRNAs require that they fold into thermodynamically stable RNA structures to carry out their functions. If these functions are evolutionarily conserved, then structures may also be conserved between related sequences. Thus, identifying thermodynamically stable and conserved RNA structures from sequence data can help identify putative ncRNAs. This session will cover various approaches for modeling RNA structures using thermodynamics, biochemistry, and sequence comparison, identifying homologous RNA sequences/structures, and using these methods to suggest which sequences have likely non-coding functions. This session will also cover the application of these methods to identify ncRNAs in target species.
Organizer: Dr. Walter N. Moss, Yale University and Howard Hughes Medical Institute, New Haven (USA).

3. **SS3: Biological Knowledge Visualization** Biotechnology produces large amounts of data, to be handled and analyzed by computational methods. The scale and complexity of such input data, but also of the analysis results, is large, and can be tackled by visual techniques for representation and interpretation. The focus of this special session is the discussion of approaches that integrate visual intelligence into the analytical process of such biological data. Contributions may range from the development of novel visual

techniques to the successful application of existing ones in biological or biomedical studies.
Organizer: Dr. Rodrigo Santamaria, University of Salamanca (Spain).

4. **SS4: High Performance Computing in Bioinformatics** The goal of this special session is to explore the use of emerging parallel computing architectures as well as High Performance Computing systems (Supercomputers, Clusters, Grids) for the simulation of relevant biological systems and for applications in Bioinformatics, Computational Biology and Computational Chemistry. We welcome papers, not submitted elsewhere for review, with a focus in topics of interest ranging from but not limited to: -Parallel stochastic simulation. -Biological and Numerical parallel computing. -Parallel and distributed architectures. -Emerging processing architectures (e.g. GPUs, Intel Xeon Phi, FPGAs, mixed CPU-GPU or CPU-FPGA, etc). -Parallel Model checking techniques. -Parallel algorithms for biological analysis. -Cluster and Grid Deployment for system biology. -Biologically inspired algorithms.
Organizers: Dr. Horacio Perez-Sanchez and Dr. Jose M. Cecilia, Catholic University of Murcia (UCAM), (Spain).
Dr. Ivan Merelli, Institute for Biomedical Technologies, National Research Council of Italy, Milano (Italy).

5. **SS6: ePathology - Realities and Perspectives** The session will discuss the current research and ongoing activities implemented in the field of ePathology. Digital imaging as well as virtual pathology boards acquire more and more importance. Application of medical and laboratory information management systems for pathology purposes ia also important and actual. Special importance acquires implementation of eLearning technologies for continuous medical education of pathologists as well as introduction and practical application for realization of quality assurance programs in pathology and cytology.
Organizer: Dr. Ekaterine Kldiashvili, Georgian Telemedicine Union (Association), Tbilisi (Georgia).

6. **SS7: Modelling of cellular pathways and disease** This session will discuss applications of mathematical and computational techniques to the representation of biological processes or pathways, with the aim of gaining a mechanistic understanding of cellular functions. A wide range of approaches will be considered, ranging from qualitative network representation to fully quantitative kinetic models. Applications may focus on a specific metabolic, signalling and regulatory pathway, or on the contrary offer genome-scale coverage of processes implicated in a particular biological function or disease. We welcome interdisciplinary papers presenting the integration of modelling techniques with experimental analyses.
Organizers: Dr Jean-Marc Schwartz, Faculty of Life Sciences, University of Manchester (UK).
Prof Marija Krstic-Demonacos, School of Environment & Life Sciences, University of Salford (UK).

7. **SS8: Integration of data, methods and tools in biosciences** With the rise of biotechnology and bioinformatics, a number of problems regarding storing, searching and using biological data occur. The data are scattered over a large number of repositories (public or private) and stored in a large number of different formats. Furthermore, some formats are not adequate for automatic computer processing (such as text documents) and require some kind of preprocessing before they can be input into computer algorithms. This situation makes searching and analyzing the data very difficult, leaving

facts and knowledge on observed biological phenomena hidden in databases and digital collections. The above problems and other similar problems can be overcome only by an integrative bioinformatics approach. The integration can be done across different aspects and on various levels. Topics of interests include (but are not limited to): -database integration techniques, -data acquisition from heterogeneous sources, -genotype-phenotype associations researches, -integrative modeling and analysis of processes and systems in biosciences, -tool integration and workflow systems, -computational infrastructure for biological researches, including laboratory management systems, -biological ontologies and metadata, -integrative data and text mining approaches. Contributions that addresses any other aspect of integrating data, methods, techniques and/or tools, are also welcome. Organizer: Dr. Vesna Pajic, University of Belgrade, Faculty of Agriculture (Serbia).

8. **SS9: Biomaterials in Biomedicine: Computational approaches** The development of materials that can successfully replace biological tissues both in function and appearance, the so-called Biomaterials, is a field of growing interest in Biomedicine. The use of computational methods to determine or predict the physical properties of these materials has become one of the main focuses of the latest research in this area, allowing pre-clinical non-invasive methods for evaluation and testing of this type of materials. Also, theoretical modelling of these materials can expedite research by allowing conducting simulated experiments in order to find the solution in the real system that is being studied in response to changing conditions, and helps obtaining information on optical, mechanical and many other potentially interesting properties. The main topics of interest of this Special Session include (but are not limited to): -Computational methods for assessment and prediction of physical (including mechanical, optical,...) properties of biomaterials; -Application of biomaterials in Biomedicine; -Modelling systems of biomaterials; -Software tools and simulation packages for the evaluation of suitability of biomaterials;
Organizers: Dr. Razvan Ghinea, Department of Optics, University of Granada (Spain).
Dr. Luis Javier Herrera, Department of Computer Architecture and Computer Technology, University of Granada (Spain).
Dr. Maria del Mar Perez, Department of Optics, University of Granada (Spain).

9. **SS10: Effective Soft Computing Methods for Biomedical Signals** The goal of this special session is to elaborate applications of soft computing methods for biomedical signals such as ECG, EMG, and EEG. This session will discuss new hybrid algorithms and investigate effective as well as high performance computing techniques for the classification of biomedical signals for diagnosis or the other applications in Biomedical Engineering and Bioinformatics. We welcome papers, not submitted elsewhere for review, with a focus in topics of interest ranging from but not limited to: -Artificial Neural Networks Algorithms on Biomedical Signals. -Fuzzy Systems and Fuzzy Clustering Algorithms on Biomedical Signals. -Probabilistic Model Algorithms on Biomedical Signals. -Metaheuristic Algorithms on Biomedical Signals. -Hybrid Algorithms on Biomedical Signals.
Organizer: Prof. Dr. Bekir KARLIK, Faculty of Engineering, Selcuk University, Konya, (Turkey)

10. **SS11: Chaperone Therapy for Protein Misfolding Disorders with Brain Dysfunction** Chaperone therapy is a new concept of molecular therapeutic approach, first developed for lysosomal diseases, based on a paradoxical molecular phenomenon involving lysosomal enzyme proteins and their competitive inhibitors as intracellular enhancers

(chaperones). The misfolded mutant enzyme protein is stabilized as a molecular complex with its substrate analogue, chaperone, and transported safely to the lysosome. The complex is automatically dissociated in the lysosome under the acidic condition, the free mutant protein remains stable, and the enzyme activity is expressed. The small chaperone molecule has been confirmed to be delivered to the brain tissue through the blood-brain barrier. This new trial was targeted initially at a few lysosomal diseases, and then has been expanded to many other diseases with pathological protein dysfunction due to structural misfolding. The advantages of this molecular approach over other currently available therapies for genetic and protein misfolding diseases are summarized twofold; first, oral administration to individuals with intractable diseases; and second, delivery to the central nervous system for therapy of brain diseases currently not available with other experimental or clinical trials. In this session we discuss the present status of chaperone therapy research, focusing on chemical compounds, enzyme-chaperone interactions, and animal and human experiments for a few genetic diseases. This therapeutic approach will become a novel and revolutionary scientific and medical strategy in the near future.
Organizer: Dr. Yoshiyuki Suzuki, Tokyo Metropolitan Institute of Medical Science (Japan).

11. **SS13: Computational analysis of gene regulatory elements with next-gen sequencing data** The advent of next-generation sequencing (NGS)-based epigenetic profiling assays opens new perspectives for studying gene regulation. In particular, ChIP-Seq against histone marks and transcription factors (TFs), BS-Seq for DNA-methylation profiling and various protocols for assessing chromatin accessibility allow for a detailed characterization of the chromatin states of tens of thousands of cis-regulatory elements at once. Moreover, high-throughput protocols for sequencing RNA 5'ends allow for genome-wide monitoring of transcription initiation events at single base pair resolution. This session focuses on novel computational approaches to extract insights on gene regulatory mechanism from the vast amounts of epigenetic profiling data that have been accumulated over the last few years. We welcome contributions on topics ranging from but not limited to: -Promoter inference and classification from transcript mapping data. -Functional classification of regulatory elements based on chromatin features. -Methods for identifying differentially modified chromatin regions. -Prediction of nucleosome positioning from DNA sequence and TF binding events -Discovery of genetic variants associated with changes in chromatin state. -Usage of epigenetic profiling data for medical diagnosis and treatment decision
Organizer: Dr. Philipp Bucher, Swiss Federal Institute of Technology in Lausanne (EPFL) (Switzerland).

12. **SS14: Better Oncology Treatment and Patient Outcomes by Using Therapy-Related Symptom Checklists (TRSC/TRSC-C) and a Computerized Two-Way Communication System** This special session will describe and illustrate the development/measurement properties, current use, and future electronic uses of patient-friendly symptom checklists for adults (TRSC) and children (TRSC-C) oncology patients. These checklists can be completed in less than 5 minutes by adults and children/parents, and involve no radical alterations in or increased costs of clinical practice. The adult checklist was developed 1984-1995 and the child checklist 2004-2010. Both checklists now have versions in English, Spanish, Thai, Pilipino, Chinese, and Bahasa Indonesia, and have been used in clinic settings in the USA and other countries. The checklists were originally developed to reduce observed under-documentation of treatment symptoms of

concern to patients in medical records. The hope was that use of these checklists (25 symptoms for adults and 30 for children) would lead to better documentation through integration of the checklists into electronic medical records (EMR), enhance communications among patients and clinicians, and improve health outcomes. Completed research has found high levels of patient/clinician satisfaction with use of the TRSC, no increase in clinic costs, and strong correlations of TRSC/TRSC-C scores with the number of patient symptoms documented and managed, patient functional status, and patient quality of life, A recently published sequential cohort trial with adult outpatients at a Mayo Clinic Health System community cancer center reported that use of the TRSC produced a 7.2 higher covariate adjusted patient quality of life, 116 more symptoms documented and managed, and higher functional status. These results were statistically and substantively significant. This special session will illustrate transitions among design, measurement, application, and informatics required for good patient care. Audience participation will be solicited, and establishment of relationships among the audience for work along the lines presented and discussed in this special session will be encouraged
Organizers: Dr. Arthur R. Williams, Research Associate, CINDRR, US Department of Veterans Affairs, and Professor, Department of Healthcare Policy and Management, College of Public Health, University of South Florida, Tampa (USA)
Dr. Phoebe D. Williams, Professor, School of Nursing, Kansas University Medical Center, Kansas City, Kansas (USA)

13. **SS15: Computational MRI: Theory, Dynamics and Applications** Computational techniques are invaluable to the continued success and development of Magnetic Resonance Imaging (MRI) and to its widespread applications. New processing methods are essential for addressing issues at each stage of MRI techniques. Magnetic Resonance Imaging simulations based on the Bloch NMR equations are of high educational value. They serve as essential tools in basic MRI method development, sequence design and protocol optimization. In this special session, the underlying physical and biomedical models of the Bloch NMR flow equations, their field of applications and possible limitations will be discussed. Magnetization preparation will be simulated in order to tailor sequence protocols for specific applications, exploiting basic spin relaxation as well as advanced Magnetic Resonance contrast mechanisms such as flow and diffusion. The main objective of this special session is to bring together the mathematicians, computer scientists, theoretical physicists, medical scientists and the engineer and apply their tools through high quality paper presentations and contribute to this fast developing and most exciting field of our time without acquiring the most sophisticated equipment. Volume I of a new book titled "Theory, Dynamics and Applications of MRI" will be released and presented at this special session. The book is intended to present basic theory of MRI and develop several fundamental equations which can be invaluable for quantitative and qualitative analysis of NMR magnetizations and signals. Based on this special session, scientists should be able to apply basic MRI methods to solve real life problems using computational methods of their choice.
Organizer: Dr. Omotayo Bamidele Awojoyogbe, Department of Physics, Federal University of Technology, Minna, Niger-State (Nigeria)

14. **SS16: Bioinformatical Approaches to Disordered Proteins** According to the "structure-function-paradigm" a stable, folded structure is a pre-requisite of protein function. However, since the turn of the century it became evident via an increasing number

of known examples that many proteins are able to serve crucial functions in vivo without adopting a well-defined 3D structure. In accordance with the typical functions these Intrinsically Unstructured/Disordered Proteins (IUPs/IDPs) serve in signalling, regulation and transcription, they were found to be at the heart of many diseases such as cancer, neurodegenerative diseases and diabetes, among others. This sparked interest in the focused research of IUPs not only at a basic science level but also in biomedicine/pharmacology. However, due to biological and technical reasons, experimental study of these proteins remains difficult and expensive and therefore the exact biological function, mode of action and biophysical/thermodynamical description of the majority of IUPs remain elusive. Because of these difficulties, bioinformatics tools that target protein disorder play an important role in the identification and characterization of IDPs. This is exemplified by the fact that the majority of our systems-, network- and evolutionary level knowledge of IUPs are based in various bioinformatics prediction methods and analyses. In this session we present the concept of IUPs and focus on the current, state-of-the art bioinformatics approaches, tools and results of analyses concerning protein disorder.
Organizer: Prof. Istvan Simon, Istitute of Enzymology, Research Centre for Natural Sciences (Hungary)

15. **SS18: Stochastic Modelling of Biological Systems** Recently there has been a significant interest in the stochastic approach for modelling biological systems, mainly because experimental data are providing evidences that random events play significant roles in determining the complex behaviours observed in living organisms. The growing amount of evidences arising from experimental observations at the single cell level are showing that fundamental biological processes such as, e.g., fate decision making, gene expression regulation, phenotypic variability are deeply conditioned by random fluctuations at the molecular level. Approaches grounding on computational stochastic modelling and simulations are proving to be useful tools for gaining insights about the role played by random events in determining the global dynamics of biological networks. In these cases deterministic descriptions fail both in predicting the observed random fluctuations and in capturing stochastic-driven phenomena such as stochastic focusing, stochastic switching and multiplicative noise effects. The most popular strategy for building stochastic models of biological phenomena consist in describing the temporal evolution of the considered system as a discrete-state, continuous time Markov process. Recently, an increasing number of works proposing non-Markovian methods is emerging, with the aim of providing more accurate representations of the random events observed experimentally. The need for non-Markovian modeling is also related to the finding of long-range memory and non-ergodicity. In this special session we aim at collecting original research works reporting on topics related to the stochastic modelling of biological systems at different levels, ranging from the inter-molecular scale to the inter-cellular scale, including neural networks. Particular attention will be devoted to those proposals highlighting how the proposed modeling approach allows to gain insights on the investigated biological phenomenon and on the emerging properties of biological networks.
Organizers: Dr. Paolo Paradisi and Dr. Davide Chiarugi, Institute of Science and Technologies of Information (ISTI) - Italian National Research Council (CNR), Pisa (Italy)

This second 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 BioMed Central, e-Health Business Development BULL (España) S.A.,

and the institutions Faculty of Science, Dept. Computer Architecture & Computer Technology and CITIC-UGR from the University of Granada for their support and grants. We wish also to thank to the Editor-in-Chief of different international journal for their interest in editing special issues from 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.

April, 2014
Granada

IWBBIO Editors and Chairs
Francisco Ortuño
Ignacio Rojas

Program and Steering Committee

Irina Abnizova	Wellcome Trust Sanger Institute
Jesus Aguilar	University of Pablo Olavide
Carlos Alberola	University of Valladolid
Patrick Aloy	Institute for Research in Biomedicine (IRB)
Rui Alves	Universitat de Lleida
Miguel Andrade	Max Delbrck Center for Molecular Medicine (MDC)
Eduardo Andres	Spanish National Cancer Research Centre (CNIO)
Jose Andreu	Centre of Biological Research (CIB-CSIC)
Marco Antoniotti	Universit degli Studi di Milano Bicocca
Ana Aransay	Center for Cooperative Research in Biosciences (CIC bioGUNE)
Saul Ares	Spanish National Center for Biotechnology (CNB)
Rubén Armañanzas	Technical University of Madrid (UPM)
Antonia Aránega	University of Granada
Gualberto Asencio Cortés	"Pablo Olavide" University
Bamidele Awojoyogbe	Federal University of Technology in Minna
Jaume Bacardit	University of Nottingham
Esteban Ballestar	Chromatine and Disease Group (IDIBELL)
Gianni Barcaccia	University of Padova
Ugo Bastolla	Research Center of Molecular Biology "Severo Ochoa"
Paul Bates	Cancer Research UK London Research Institute
Oresti Baños	University of Granada
Shomi Bhattacharya	Andalusian Molecular Biology and Regenerative Medicine Centre (CABIMER)
Alfredo Benso	Politechnic University of Turin
Concha Bielza	Technical University of Madrid (UPM)
Jose M. Blanca	Technical University of Valencia (UPV)
Armando Blanco	University of Granada
Ignacio Blanquer	Technical University of Valencia (UPV)
Andrs Bueno	Catholic University of Murcia (UCAM)
Philipp Bucher	Swiss Federal Institute of Technology in Lausanne
Juan Calvete	Biomedicine Institute of Valencia (IBV-CSIC)
Daniela Calvetti	Case Western Reserve University
Mario Cannataro	University Magna Gracia of Catanzaro
Carlos Cano	University of Granada
Emidio Capprioti	University of Alabama at Birmingham
Jose Maria Carazo	Spanish National Center for Biotechnology (CNB)
Pablo Casaseca	University of Salamanca
Nazareth Castellanos	Complutense University of Madrid
Monica Chagoyen	Spanish National Center for Biotechnology (CNB)
M. Gonzalo Claros	Andalusian Bioinformatics Platform
Jean Claverie	National Centre of Scientific Research (CNRS)
Jacques Colinge	Research Center for Molecular Medicine of the Austrian Academy of Sciences
Jose M. Cecilia	Catholic University of Murcia (UCAM)
Ana Conesa	"Principe Felipe" Research Center (CIPF)

Christian Conrad	German Cancer Research Center (DKFZ)
Juan Cruz Cigudosa	Spanish National Cancer Center (CNIO)
Alfonso Marquez Chamorro	Pablo de Olavide University
Davide Chiarugi	Italian National Research Council (CNR)
Guillermo de La Calle	Technical University of Madrid (UPM)
Carmen de Mendoza	"Carlos III" Hospital of Madrid
Coral Del Val	University of Granada
Ramón Díaz Uriarte	University Autnoma of Madrid (UAM)
Jose Díaz	Centre of Biological Research (CIB-CSIC)
Hernan Dopazo	"Principe Felipe" Research Center (CIPF)
Joaquin Dopazo	"Principe Felipe" Research Center (CIPF)
Juergen Eils	German Cancer Research Center (DKFZ)
Ernesto Estrada	University of Strathclyde
Jose Jesus Fernandez	University of Almeria
Sylvain Foissac	Center for Genomic Regulation (CRG)
Jean-Fred Fontaine	Max Delbrck Center for Molecular Medicine (MDC)
Lutz Froenicke	University of California Davis
Hans Gabius	University of Munich
Juan García Gómez	Technical University of Valencia (UPV)
M. Teresa Garca Valverde	Catholic University of Murcia (UCAM)
Razvan Ghinea	University of Granada
Paulino Gmez-Puertas	"Severo Ochoa" Molecular Biology Center
Daniel Gonzalez	University of Vigo
Humberto Gonzalez	University of Santiago de Compostela
Juan Ramón Gonzalez	Center for Research in Environmental Epidemiology (CREAL)
Marina Gordaliza	University of Salamanca
Ananth Grama	Purdue University
Raik Gruenberg	Center for Genomic Regulation (CRG)
Concettina Guerra	University of Padova
Alberto Guillen	University of Granada
Roderic Guigo	Center for Genomic Regulation (CRG)
Stefan Götz	"Principe Felipe" Research Center (CIPF)
Christophe Guyeux	IUT Belfort-Montbeliard
Michael Hackenberg	University of Granada
Kim Henrick	European Bioinformat Institute (EBI)
Vicente Hernández	Politechnic University of Valencia (UPV)
Lynette Hirschman MITRE	MITRE Corporation
Martijn Huynen	Nijmegen Centre for Molecular Life Sciences
Iñaki Inza	University of Basque Country
Ana Maria Ionescu	University of Granada
Richard Jackson	University of Leeds
Andrew Jenkinson	European Bioinformatics Institute (EBI)
Bekir Karlik	Selcuk University
Guzin Kekec	Fatih University
Ekaterine Kldiashvili	Georgian Telemedicine Union
Juliane Klein	University of Tubingen
Martin Krallinger	Spanish Nacional Cancer Research Centre (CNIO)

Natalio Krasnogor	University of Nottingham
Abhay Krishan	Andalusian Molecular Biology and Regenerative Medicine Centre (CABIMER)
Marija Krstic-Demonacos	University of Salford
Sajeesh Kumar	UT Health Science Center
Vipin Kumar	University of Minnesota
Rainer König	German Cancer Research Center (DKFZ)
Pedro Larrañaga Computational Intelligence Group, UPM	Technical University of Madrid (UPM)
David Leader	University of Glasgow
Isaac Lera	University of the Balearic Islands
Miguel A. Lopez Gordo	University of Cadiz
Ernesto Lowy	Center for Genomic Regulation (CRG)
Jose Lozano	University of the Basque Country
Victor Maojo	Politechnic University of Madrid (UPM)
Marc Marti	Spanish National Center of Genomic Analysis
Diego Martín	University of Salamanca
Fernando Martín	Institute of Health Carlos III
Victoria Martín Requena	University of Malaga
Francisco Martínez-Álvarez	University "Pablo Olavide" of Seville
Victor Martínez Gómez	University of Granada
Alexandre Masselot	Swiss Institute of Bioinformatics (CMU)
Marco Masseroli	Polytechnical University of Milano
Jose Mato	Center for Cooperative Research in Biosciences (CIC bioGUNE)
Rune Matthiesen	University of Porto
Miguel Medina	University of Malaga
Ivan Merelli	Italian National Research Council (CNR)
Jordi Mestres	University Pompeu Fabra
Iain Moal	Barcelona Supercomputing Center (BSC)
Antonio Morreale	Center of Molecular Biology "Severo Ochoa"
Walter N. Moss	Yale University and Howard Hughes Medical Institute
Enrique Muro	Max Delbrck Center for Molecular Medicine (MDC)
Carmen Navarro	University of Granada
Isabel Nepomuceno	University of Seville
Michael Ng	Hong Kong Baptist University
Baldomero Oliva	University Pompeu Fabra
José Luis Oliveira	University of Aveiro
Jose Luis Oliver	University of Granada
Salvatore Orlando	University of Venezia
Julio Ortega	University of Granada
Francisco Ortuño	University of Granada
Christos Ouzounis	Centre for Research and Technology Hellas
Erola Pairó	Institute for Bioengineering of Catalonia (IBEC)
Akhilesh Pandey	Johns Hopkins University
Lorena Pantano NA	Institute of Predictive and Personalized Medicine of Cancer (IMPPC)
Vesna Pajic	University of Belgrade

Paolo Paradisi	Italian National Research Council (CNR)
Alberto Pascual	University "Complutense" of Madrid (UCM)
David Pelta	University of Granada
Alexandre Perera	Technical University of Catalonia (UPC)
Mara del Mar Prez	University of Granada
Joanna Polanska	Silesian University of Technology
Javier Perez Florido	Genomics and Bioinformatics Platform of Andalusia (GBPA)
Horacio Perez Sanchez	Catholic University of Murcia (UCAM)
Graziano Pesole	University of Bari
David Pisano CNIO	Spanish National Cancer Center (CNIO)
Alberto Policriti	University of Udine
Héctor Pomares	University of Granada
Pablo Porras	European Bioinformatics Institute (EBI)
Robert Preissner	University of Charit
Alberto Prieto	University of Granada
Carlos Puntonet	University of Granada
Omer Rana	Cardiff University
Jairo Rocha	University of the Balearic Islands
Ignacio Rojas	University of Granada
Julio Rozas	University of Barcelone
Gregorio Rubio	Politechnic University of Valencia (UPV)
Gonzalo Ruiz	University of Granada
Maria Jose Saez Lara	University of Granada
Yvan Saeys	Flanders Institute for Biotechnology (VIB)
Rodrigo Santamaria	University of Salamanca
Pilar Santisteban	Institute of Biomedical Researchs "Alberto Sols"
Javier Santoyo	Andalusian Institute of Human Sequencing
Michael Sattler	Institute of Structural Biology (Helmholtz Zentrum Mnchen)
Peter Schmidtke	University of Barcelona
Vicky Schneider	European Bioinformat Institute (EBI)
Jose Seoane	University of Bristol
Luis Serrano	Center for Genomic Regulation (CRG)
James Sharpe	Centre for Genomics Regulation (CRG)
Vladimir Shulaev	University of North Texas
Jean-Marc Schwartz	University of Manchester
Istvan Simon	Hungarian Research Centre for Natural Sciences
Richard Sinnott	University of Glasgow
Prashanth Suravajhala	Bioclues.org and Bioinformatics.org
Yoshiyuki Suzuki	Tokyo Metropolitan Institute of Medical Science
Li Teng	University of Iowa
Yaping Tian	Chinese PLA General Hospital
Carolina Torres	University of Granada
Huseyin Tombuloglu	Fatih University
Oswaldo Trelles	University of Malaga
Paolo Trunfio	University of Calabria
Shusaku Tsumoto	Shimane University
Renato Umeton	Massachusetts Institute of Technology
Jose Miguel Urquiza	Chromatin and Disease Group (IDIBELL)

Olga Valenzuela	University of Granada
Alfredo Vellido	Technical University of Catalonia (UPC)
Nicola Vitacolonna	University of Udine
Arthur R. Williams	CINDRR and University of South Florida
Phoebe D. Williams	Kansas University Medical Center
Wolfgang Wurst	National Research Center of Environment & Health
Albert Zomaya	University of Sydney

Table of Contents

SS1: Multi-biomarker and informatics in cancer diagnosis

SS6: ePathology - Realities and Perspectives

SS7: Modelling of cellular pathways and disease

Biomedical Data Mining

SS2: Discovery of non-coding and structured RNAs

SS14: Better Oncology Treatment and Patient Outcomes by Using Therapy-Related Symptom Checklists (TRSC/TRSC-C) and a Computerized Two-Way Communication System

High Performance for Sequence Analysis

SS11: Chaperone Therapy for Protein Misfolding Disorders with Brain Dysfunction

SS8: Integration of data, methods and tools in biosciences

SS4: High Performance Computing in Bioinformatics

SS3: Biological Knowledge Visualization

Bioinformatics Tools and Databases

SS13: Computational analysis of gene regulatory elements with next-gen sequencing data

Gene Expression and Microarrays

Biomedical Engineering and eHealth Applications

Ethical Principles in Biotechnology and Bioengineering

High Performance Bioinformatics for Healthcare and Diseases

miRNA Regulation Networks

Computational Approaches for Genomics and NGS

SS16: Bioinformatical Approaches to Disordered Proteins

SS18: Stochastic Modelling of Biological Systems

Poster Session

Virtual Presentation

Protein function easily investigated by genomics data mining using the ProteINSIDE web service

Nicolas KASPRIC[1, 2], Brigitte PICARD[1, 2], Matthieu REICHSTADT[1, 2], Jérémy TOURNAYRE[1, 2] and Muriel BONNET[1, 2]

[1] INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France
[2] Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

Corresponding authors: nicolas.kaspric@clermont.inra.fr & muriel.bonnet@clermont.inra.fr

Abstract. ProteINSIDE is a new workflow to analyse lists of protein or gene identifiers from ruminant species and gather biological information provided by functional annotations, putative secretion of proteins and proteins interactions networks. ProteINSIDE gets results from several software and databases with a single query. From a unique list, ProteINSIDE uses orthologs identifiers within well studied species (Human, Rat or Mouse) to extend analyses and biological information retrieval. ProteINSIDE is freely available at: http://www.proteinside.org.

Keywords: web service, workflow, protein-protein interaction, protein secretion, gene ontology, networks.

1 Introduction

The current challenge for scientists working on the efficiency of ruminant (cattle, sheep or goat) and the quality of their products (meat, milk...) is to understand which genes and proteins control nutrient metabolism and partitioning between tissues or which genes and proteins control tissues growth and physiology [1]. Such questioning leads to the genome annotation, the sequencing and the quantification of gene expression or protein abundance. The quantity of data produced by these genomic and proteomic studies increases continuously [2-4]. There is a necessity to analyse, understand and generate biological information and knowledge from these data [5]. This is possible by using a panel of tools requiring different identifiers (IDs) per protein or gene and time to read and analyse the results. Moreover, most databases (DB) like UniProtKB [6] or NCBI [7] possess a large quantity of information and most of existing bioinformatic tools implemented as web services are specific to one analysis: as the annotation according to the Gene Ontology (GO) [8] or the prediction of signal peptide [9] or the molecular interactions identification [10] and visualization as networks [11, 12]. Many workflows that integrate several analyses are available [13-16]

and are specific to a species (Drosophila, *Arabidopsis thaliana, Escherichia coli*…), and thus are not suitable for the analysis of genomic data from ruminant species. The few workflows working with ruminant data are multispecies, the results are not species-specific and the data source is not available because of the privacy of databases (as the license software Pathway Studio [17]). Other workflows are specialized on the identification of candidate genes related to diseases as ToppGene [18]. Thus, to date there is no workflow dedicated to the integrative analysis of genomic data from ruminant species.

Unlike Human or model species like mouse or rat, ruminant species are less annotated and protein sequences are not always verified. Often, scientists use orthologs with the aim to increase the meaningful biological contexts for proteins. For this purpose biologists query for annotations according to Gene Ontology, the putative secretion of proteins, protein-protein interactions (PPi) and network analysis first in ruminant and then in Human or in rodents. The integration within a workflow of gateways between proteins / genes from ruminants and their orthologs from Human and models species has never been done.

Here we propose ProteINSIDE, a web service dedicated to a systematic and integrative analysis of protein's biological information. ProteINSIDE works using lists of proteins or genes IDs from 6 species (Bovine, Ovine, Caprine, Human, Rat, and Murine) to annotate functions and cellular location, predict secreted proteins, search for interactions between proteins within and/or outside a dataset and allowing cross-species analysis using orthologs.

2 Materials and methods

This section lists necessary equipment, ProteINSIDE resources and describes the dataset used to assess the functionalities of our tool.

2.1 Equipment

ProteINSIDE doesn't require an installation on a computer and the web service is available online at www.proteinside.org by using an internet browser. ProteINSIDE is completely adapted for any internet browser, but for better performances we recommend to use Firefox, Chrome, or Safari.

2.2 Implementation

ProteINSIDE is divided into three parts: the workflow, the database and the web interface. The workflow is a combination of Perl (version 5.10.1; CPAN modules (Comprehensive Perl Archive Network) used and BioPerl [19]) and R (version 3.0.1; with "tnet" package [20]) scripts to query databases, recover protein data, perform calculations and run algorithms for signal peptide predictions and network visualisation. The MySQL database aims to reduce server load and thus stores both available knowledge form major public databases and results (and settings) form queries. The

web interface is programmed in PHP, HTML, and JavaScript. It is devoted to the creation of a new analysis, the view of results and users information with updates.

2.3 ProteINSIDE structure and interface

A flow chart (Fig 1) details the type of analysis (basic or customizable) and the four main queries proposed to the user. Whatever the type of analysis, the workflow uses data from the input file and runs default scripts (basic analysis) or scripts and options selected through the settings (customs analysis). At analysis completion, results are created and uploaded on ProteINSIDE database to decrease web interface treatment duration (results have to be deleted by the user; visitors results are automatically deleted monthly).

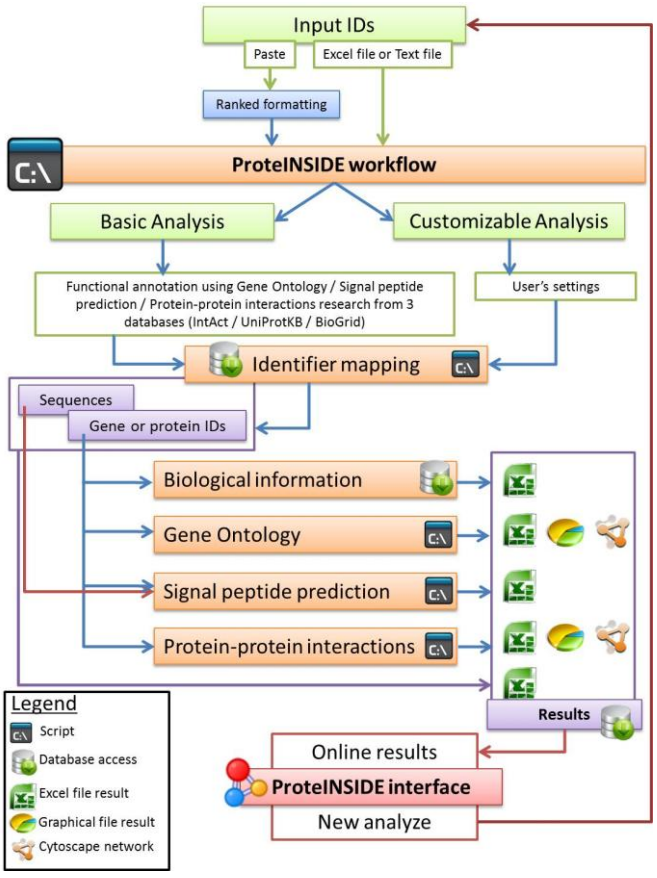


Fig. 1. Flow chart of ProteINSIDE structure. The four modules for querying the available biological information, annotations according to the gene ontology, signal peptide predictions and protein-protein interactions are either all present in the basic analysis or individually selected in the custom analysis.

ProteINSIDE is easily run by biologist through the interface. Registered user or visitor run a new analysis by using the web interface menus “Basic Analysis” (automatic settings) or “Custom Analysis” (user selects the settings):

1. Click on “Basic Analysis” menu on the homepage of ProteINSIDE
2. Fill in “the job name” box
3. Select the species for the analysis (related to the IDs that will be used on this query)
4. Upload your input file or directly paste your IDs
5. Click on the “Run the job” button to submit a new analysis

ProteINSIDE gives a link and an access code to view analysis status and get the results. The analysis status is indicated by the colour of a button: red for “analysis on the waiting list”, yellow for “the analysis is running” and green “analysis completion”. The blue globe is the link to access to the online results views:

1. Click on the blue globe button to view the results (use the trash to delete them)
2. Visualise the results summary produced by selected modules on the first default page
3. Navigate to module’s results pages by clicking on the module’s name on the toolbar menu.

2.4 The input and the output of ProteINSIDE

ProteINSIDE inputs are genes or proteins IDs (e.g. ADIPO or ADIPO_HUMAN) or UniProtKB protein accession numbers (e.g. Q15848). These IDs are uploaded as text tabulated files (extension .tab or .txt) or as Excel files (.xls or .xlsx). The input files have to be ranked as three columns (Fig. 2) because of the database format. Alternatively, the IDs are directly pasted.

	A	B	C
1	1	1	A4D1N9
2	2	2	ADIPO_HUMAN
3	3	3	C1T9A_HUMAN
4	4	4	F261_HUMAN
5	5	5	F262_HUMAN
6	6	6	PFKFB3
7	7	7	PFKFB4

Fig. 2. Example of an input files made using Excel 2010 and formatted for an upload.

The output files are Excel file (.xls), Cytoscape file (.cys or .xgmml), text or FASTA file (.txt or .fa) and pictures (.jpg or .png or .pdf). They are downloadable from the page results of each module of analysis.

2.5 The sample dataset

We created a dataset to assess ProteINSIDE performances. This dataset is composed of the UniProtKB accession numbers of 133 proteins (Table 1): 34 proteins related to the glycolysis cycle, 11 proteins from the respiratory chain, 5 proteins from the tricar-

boxylic acid cycle, 79 hormones or secreted proteins and proteins with very specific functions unrelated to the others. We also included a duplicated protein among proteins of the glycolysis to verify its recognition by ProteINSIDE.

We created this dataset on bovine species, but the numbers of annotations and PPi weren't sufficient for a clear representation of the functionalities of ProteINSIDE. Then, we used the same proteins in Human to test ProteINSIDE with the "Basic" and the "Custom Analysis" (Table 1).

Table 1. Results summary of ProteINSIDE analysis performances. The numbers are the proteins that belong to main pathways in the sample dataset, that are properly annotated by GO terms relevant to glycolysis and tricarboxylic acid (TCA) functions, and that have been predicted as secreted by SignalP for hormones.

Analyses and data	Glycolysis	Hormones	TCA	Analysis time (min)
Dataset	33+1 (duplicate)	79	5	-
Basic Analysis	29	78	3	2
Custom Analysis	33	78	5	10

3 Results and discussion

Here we present the results produced by a "Basic Analysis" and a "Custom Analysis" from our sample dataset, and we discuss the relevance of biological information extracted by ProteINSIDE. All of the 133 proteins were recognized by ProteINSIDE, the protein in duplicate was identified and excluded from the analysis (Fig. 3). Thus, 132 proteins were submitted to the analyses. The numbers of proteins / genes submitted to the analyses, numbers of annotations, PPi and predicted secreted proteins are recorded on the default page following the access to the results (Fig. 3).

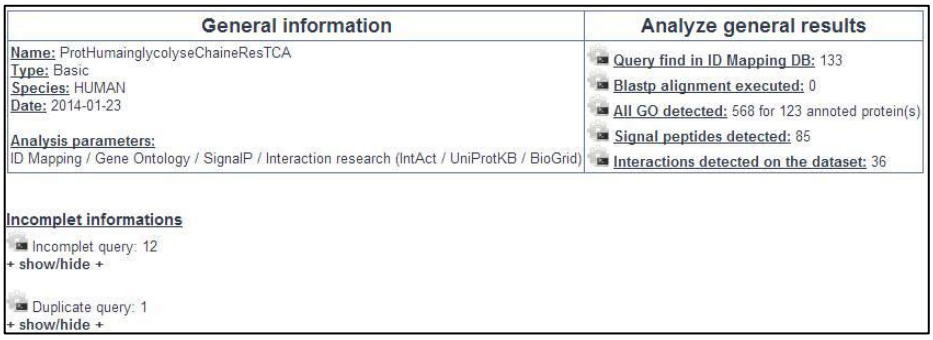


Fig.3. A table is provided by the default page after the access to the results both for a basic and a custom analysis. In addition to general information, the table provide counts of results retrieved by each module that has been run, incomplete query (IDs with missing biological information) and duplicate query.

3.1 Results of the Basic Analysis

The first module of analysis has extracted and summarized, as a downloadable table, other gene or protein IDs, gene or protein names, a summary for the protein function, the gene chromosomal location, information on tissue expression and cellular location, and the species in which orthologs have been identified. Thus on the "ID resume" page of the toolbar menu, a user has access at a glance to several information for a list of genes or their products, and also to direct links with the UniProtKB and the NCBI databases.

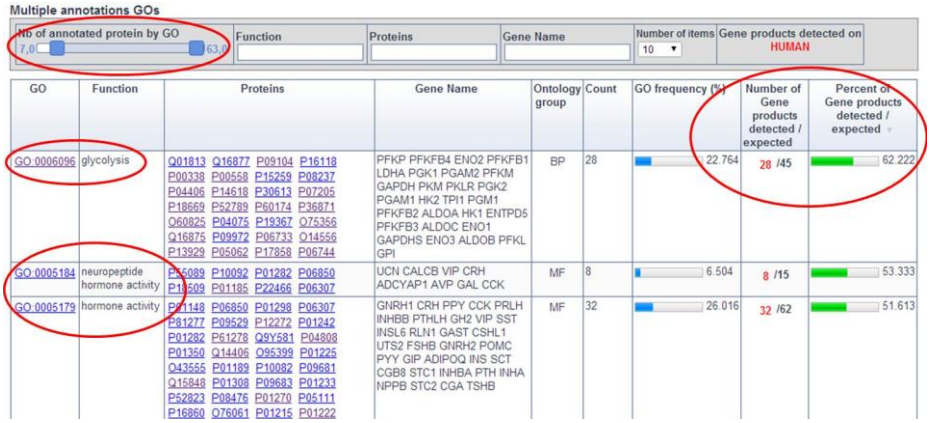
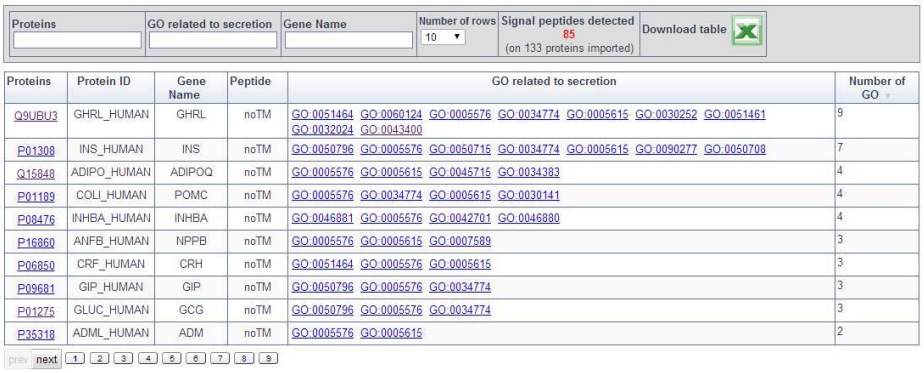


Fig. 4. Results of the functional annotations according to the Gene Ontology are available as dynamic tables. Results can be sorted by: the GO's identifier, the function, protein ID or gene name, the ontology group, the number of annotated proteins or the number of expected gene products.

On the "GO" page of the toolbar menu, we checked the relevance of the annotations extracted by ProteINSIDE by looking for the over-representation of annotations relative to glycolysis and hormones. First, among the 132 proteins submitted, ProteINSIDE annotates 123 proteins with 568 unique GOs (Fig 3). We classed these GO according to the number of proteins annotated by GO and the percentage of gene products detected/expected to identify the most common pathways associated to our sample dataset (Fig. 4). By this way, we retrieved as the most common pathways: glycolysis and hormone activity (about 62% and 52% of expected annotated gene products with these GO in Human, respectively). We have to note a lack of annotations for 12 proteins of the sample dataset, and a lack of annotations relative to glycolysis for 4 proteins (28 of the 33 expected proteins related to the glycolysis were annotated; Table 1). This lack of annotations is related to our choice to use only GO terms that have been agreed by review curator in the "Basic Analysis". This means that the "Basic Analysis" doesn't use GO annotations with IEA (Inferred by Electronic Annotation) evidence code, but the option to use IEA is provided in the custom analysis to extend the annotations.

On the "Secreted protein" page of the toolbar menu, the proteins potentially secreted are listed in a dynamic table (Fig. 3 and 5). From our sample dataset, 85 proteins were predicted as secreted by SignalP [9], among them 78 of the 79 proteins that were expected (Table 1). This lack of perfect prediction of the protein can be explained by the false positive and false negative prediction rates of SignalP, as already evaluated [21]. The prediction of secretion is then confirmed by a search for GOs related to the "secretion" function. Over the 85 predicted secreted proteins, 63 were annotated by GOs related to the "secretion" function. The double query of protein secretion both by the peptide signal prediction from protein sequence and the GO annotation is unique to ProteINSIDE.



Proteins | GO related to secretion | Gene Name | Number of rows 10 | Signal peptides detected 85 (on 133 proteins imported) | Download table

Proteins	Protein ID	Gene Name	Peptide	GO related to secretion	Number of GO
Q9UBU3	GHRL_HUMAN	GHRL	noTM	GO:0051464 GO:0060124 GO:0005576 GO:0034774 GO:0005615 GO:0030252 GO:0051461 GO:0032024 GO:0043400	9
P01308	INS_HUMAN	INS	noTM	GO:0050796 GO:0005576 GO:0050715 GO:0034774 GO:0005615 GO:0090277 GO:0050708	7
Q15848	ADIPO_HUMAN	ADIPOQ	noTM	GO:0005576 GO:0005615 GO:0045715 GO:0034383	4
P01189	COLI_HUMAN	POMC	noTM	GO:0005576 GO:0034774 GO:0005615 GO:0030141	4
P08476	INHBA_HUMAN	INHBA	noTM	GO:0046881 GO:0005576 GO:0042701 GO:0046880	4
P16860	ANFB_HUMAN	NPPB	noTM	GO:0005576 GO:0005615 GO:0007589	3
P06850	CRF_HUMAN	CRH	noTM	GO:0051464 GO:0005576 GO:0005615	3
P09681	GIP_HUMAN	GIP	noTM	GO:0050796 GO:0005576 GO:0034774	3
P01275	GLUC_HUMAN	GCG	noTM	GO:0050796 GO:0005576 GO:0034774	3
P35318	ADML_HUMAN	ADM	noTM	GO:0005576 GO:0005615	2

prev next 1 2 3 4 5 6 7 8 9

Fig. 5. Results as a dynamic table, of the potentially secreted proteins predict by SignalP.

On the "Protein interactions" page of the toolbar menu, proteins within the dataset are linked by the "Interaction detection methods" or reviewed by a curator (clicking on node gives information about the protein and a link to UniProtKB database). We selected to query BioGrid [22], UniProtKB [6] and IntAct [23] because these PPi databases are reviewed by curators and the query of PPi in 2 or 3 PPi database delivered best results [24]. PPi are listed by a dynamic table or viewed as a network (Fig. 6). The interactions research between proteins of our sample dataset has identified 36 PPi that involved 23 different proteins. As expected, PPi within the sample dataset linked proteins known to contribute to the pyruvate dehydrogenase complex (Fig 6A), the complexes IV (Fig 6B) and I (Fig 6C) of the respiratory chain, and also some proteins linked to the glycolysis and the carbohydrate oxidation (Fig 6D and 6E).

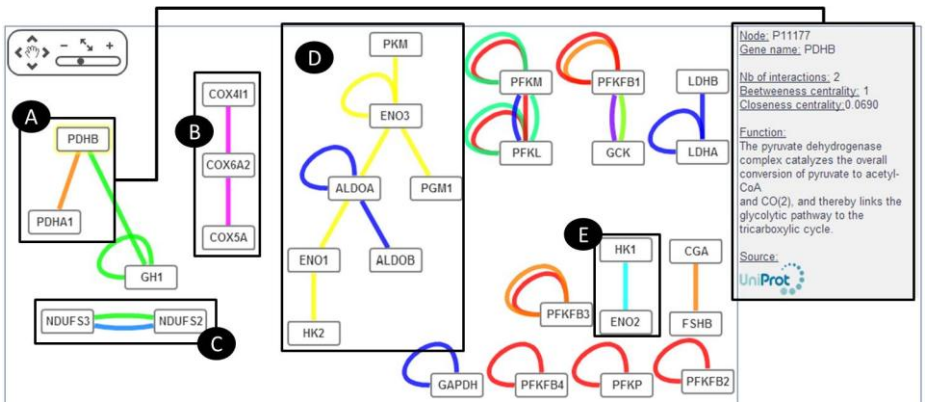


Fig. 6. Network of PPi retrieved by querying BioGrid, UniProtKB and IntAct databases. PPi are between proteins within the sample dataset. Information about a protein is obtained by clicking on a protein/gene or node. Edge colour depends on the detection method of the PPi.

3.2 The Custom Analysis: an added-value provided by the extension of the analysis

We made a "Custom Analysis" using the same major settings as the "Basic Analysis" but we used the proposed additional options:

1. The GO electronic annotation (IEA) evidence codes proposed to extend the annotation
2. GOTree network to view linked GOs
3. The search and the view of the PPi between proteins from our sample dataset and proteins outside the dataset (but still in the same species, here in Human) to extend the network and the biological information.

The use of electronic annotation has increased both the number of annotated proteins (132 rather than 123 without IEA in the basic analysis) and of annotations by around 50% since 1031 unique GOs were retrieved by ProteINSIDE. Over the 33 expected proteins related to the glycolysis, the Custom analysis of ProteINSIDE has annotated 32 proteins with the GO 0006096, glycolysis (Table 1). The GOTree network linked 236 GOs. We have chosen to visualize the GOs of the "Molecular Function" group (Fig. 7). In this visualisation, the dark red colour represents the most common GO associated to our sample dataset. As expected the GO:0005179, nominates "Hormone activity", which is consistent with the over-representation of hormones in our sample dataset. This network has also linked more specific GOs or child terms [25] of the "Hormone activity" GO, as for example "Neuropeptide hormone activity" (GO:0005184).

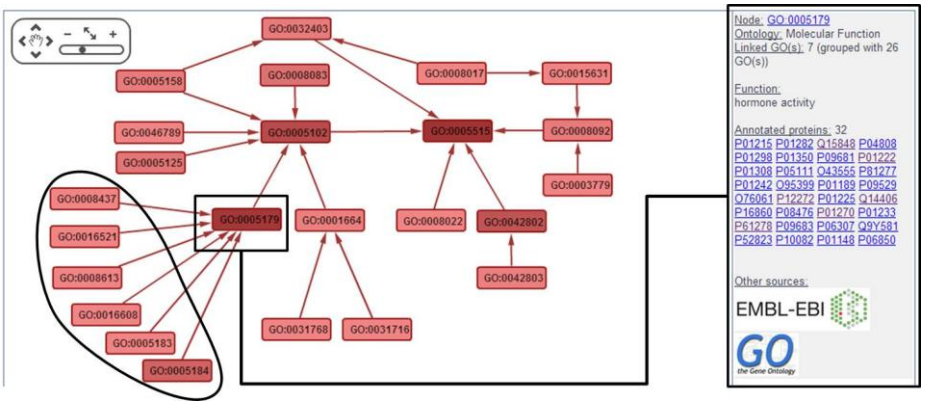


Fig. 7. A Network that links GOs used to annotate proteins of the sample dataset. Red colour is only for the GO terms relative to the Molecular Function. The degree of colour saturation represents the quantity of proteins annotated by a GO. Each edge means that a term A is a subtype of a term B (is_a). Information about a GO is obtained by clicking on the GO or the node.

Lastly, the same proteins as the "Basic Analysis" were predicted as secreted (Table 1). Thanks to the IEA electronic annotation, 82 proteins over the 85 proteins predicted to be secreted by SignalP, were also annotated by GOs related to the "secretion" function. It's 19 more than the 63 of the "Basic Analysis" because of the use of IEA evidence code for GO annotation.

By comparison with "Basic Analysis", "Custom Analysis" searches for PPi between the proteins within and outside the sample dataset by querying up to 28 DB. We chose to query the same 3 DB (BioGrid, UniProtKB and IntAct) to compare with the "Basic Analysis". ProteINSIDE retrieved 616 PPi made by 221 proteins among them 61 from the dataset. We visualized the network of the PPi (Fig. 8) and we retrieved some sub-networks relative to the respiratory chain (Fig. 8A), hormone activity such as signalization by adipokines (Fig. 8B), the growth hormone (Fig. 8C) and thyroid hormones (Fig. 8D), as well as sub-networks relative to glycolysis and carbohydrate metabolism (not highlighted).

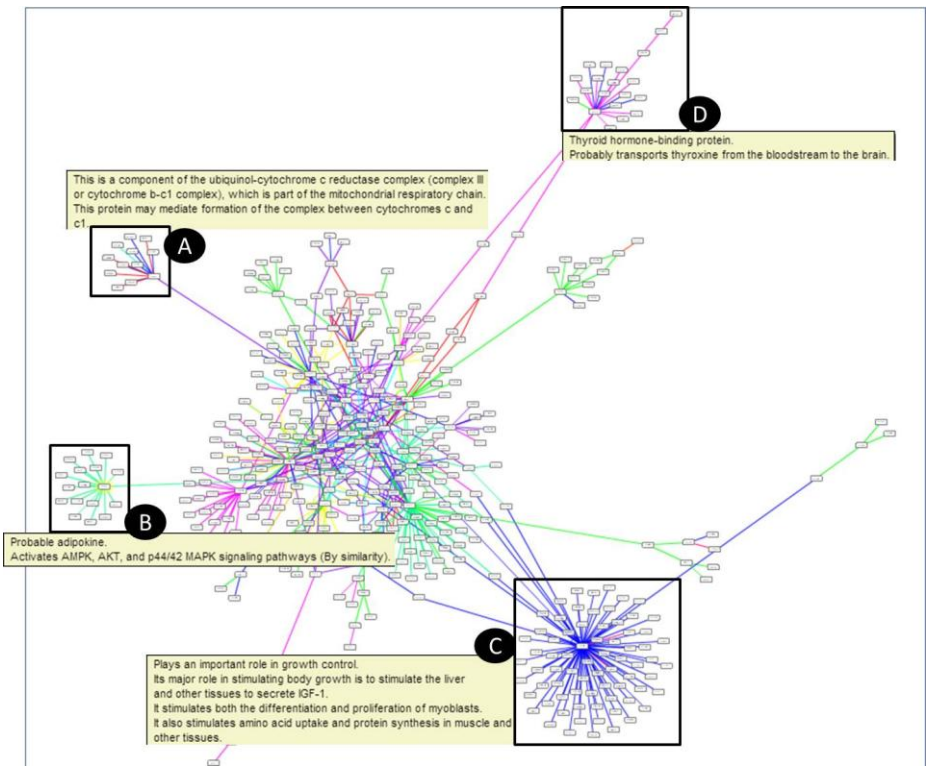


Fig. 8. Network of PPi retrieved by querying the BioGrid, UniProtKB and IntAct DB and using PPi with human proteins outside of the dataset.

4 Conclusion

In this work we present the performances of ProteINSIDE, a new powerful workflow which gathers tools and public databases to retrieve biological information of genes or proteins lists from 6 species (Bovine, Ovine, Caprine, Human, Rat, and Murine). The presented web service has correctly identified a dataset of 133 proteins, has excluded a duplicate query and has retrieved biological information for each protein. According to our dataset, ProteINSIDE properly annotates the proteins related to the glycolysis, the proteins affiliated as hormones, and the putatively secreted proteins. ProteINSIDE has revealed the most common pathways related to our dataset by creating networks from PPI interactions within and outside the dataset and from links between GOs. Each result is easily accessible and downloadable.

ProteINSIDE offers a great support to analyse a large quantity of data from genomic and proteomic studies. ProteINSIDE is also the unique web service that makes all of these analyse using ruminant IDs.

Acknowledgment

This work was supported by the Region Auvergne (FRANCE) and Apis-Gene (FRANCE).

References

1. Bonnet M, Cassar-Malek I, Chilliard Y, Picard B. "Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species". Animal : an international journal of animal bioscience, 4(7), pp. 1093-109. PubMed PMID: 22444612. (2010)
2. Picard B, Cassar-Malek I, Guillemin N, Bonnet M. Quest for Novel Muscle Pathway Biomarkers by Proteomics in Beef Production. In: Moo-Young M, editor. Comprehensive Biotechnology (Second Edition). Burlington: Academic Press; 2011. p. 395-405.
3. Chaze T, Meunier B, Chambon C, Jurie C, Picard B. "Proteome dynamics during contractile and metabolic differentiation of bovine foetal muscle". Animal : an international journal of animal bioscience, 3(7), pp. 980-1000. PubMed PMID: 22444818. (2009)
4. Taga H, Chilliard Y, Meunier B, Chambon C, Picard B, Zingaretti MC, et al. "Cellular and molecular large-scale features of fetal adipose tissue: is bovine perirenal adipose tissue brown?". Journal of cellular physiology, 227(4), pp. 1688-700. PubMed PMID: 21678425. (2012)
5. Woelders H, Te Pas MF, Bannink A, Veerkamp RF, Smits MA. "Systems biology in animal sciences". Animal : an international journal of animal bioscience, 5(7), pp. 1036-47. PubMed PMID: 22440099. (2011)
6. Magrane M, Consortium U. "UniProt Knowledgebase: a hub of integrated protein data". Database : the journal of biological databases and curation, 2011, pp. bar009. PubMed PMID: 21447597. Pubmed Central PMCID: 3070428. (2011)
7. Coordinators NR. "Database resources of the National Center for Biotechnology Information". Nucleic Acids Res, 41(Database issue), pp. D8-D20. PubMed PMID: 23193264. Pubmed Central PMCID: 3531099. (2013)
8. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. "QuickGO: a web-based tool for Gene Ontology searching". Bioinformatics, 25(22), pp. 3045-6. PubMed PMID: 19744993. Pubmed Central PMCID: 2773257. (2009)
9. Petersen TN, Brunak S, von Heijne G, Nielsen H. "SignalP 4.0: discriminating signal peptides from transmembrane regions". Nature methods, 8(10), pp. 785-6. PubMed PMID: 21959131. (2011)
10. Aranda B, Blankenburg H, Kerrien S, Brinkman FS, Ceol A, Chautard E, et al. "PSICQUIC and PSISCORE: accessing and scoring molecular interactions". Nature methods, 8(7), pp. 528-9. PubMed PMID: 21716279. Pubmed Central PMCID: 3246345. (2011)

11. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. "Cytoscape 2.8: new features for data integration and network visualization". Bioinformatics, 27(3), pp. 431-2. PubMed PMID: 21149340. Pubmed Central PMCID: 3031041. (2011)
12. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD. "Cytoscape Web: an interactive web-based network browser". Bioinformatics, 26(18), pp. 2347-8. PubMed PMID: 20656902. Pubmed Central PMCID: 2935447. (2010)
13. Wong AK, Park CY, Greene CS, Bongo LA, Guan Y, Troyanskaya OG. "IMP: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks". Nucleic Acids Res, 40(Web Server issue), pp. W484-90. PubMed PMID: 22684505. Pubmed Central PMCID: 3394282. Epub 2012/06/12. eng. (2012)
14. Pache RA, Ceol A, Aloy P. "NetAligner--a network alignment server to compare complexes, pathways and whole interactomes". Nucleic Acids Res, 40(Web Server issue), pp. W157-61. PubMed PMID: 22618871. Pubmed Central PMCID: 3394252. Epub 2012/05/24. eng. (2012)
15. Renaud Y, Baillif A, Perez JB, Agier M, Mephu Nguifo E, Mirouse V. "DroPNet: a web portal for integrated analysis of Drosophila protein-protein interaction networks". Nucleic Acids Res, 40(Web Server issue), pp. W134-9. PubMed PMID: 22641854. Pubmed Central PMCID: 3394298. Epub 2012/05/30. eng. (2012)
16. Tuncbag N, McCallum S, Huang SS, Fraenkel E. "SteinerNet: a web server for integrating 'omic' data to discover hidden components of response pathways". Nucleic Acids Res, 40(Web Server issue), pp. W505-9. PubMed PMID: 22638579. Pubmed Central PMCID: 3394335. Epub 2012/05/29. eng. (2012)
17. Nikitin A, Egorov S, Daraselia N, Mazo I. "Pathway studio--the analysis and navigation of molecular networks". Bioinformatics, 19(16), pp. 2155-7. PubMed PMID: 14594725. (2003)
18. Chen J, Bardes EE, Aronow BJ, Jegga AG. "ToppGene Suite for gene list enrichment analysis and candidate gene prioritization". Nucleic Acids Res, 37(Web Server issue), pp. W305-11. PubMed PMID: 19465376. Pubmed Central PMCID: 2703978. (2009)
19. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, et al. "The Bioperl toolkit: Perl modules for the life sciences". Genome research, 12(10), pp. 1611-8. PubMed PMID: 12368254. Pubmed Central PMCID: 187536. (2002)
20. T. O. " Structure and Evolution of Weighted Networks". University of London (Queen Mary College), London, UK, pp. 104-22. (2009)
21. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. "Locating proteins in the cell using TargetP, SignalP and related tools". Nat Protoc, 2(4), pp. 953-71. PubMed PMID: 17446895. (2007)
22. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, et al. "The BioGRID interaction database: 2013 update". Nucleic Acids Res, 41(Database issue), pp. D816-23. PubMed PMID: 23203989. Pubmed Central PMCID: 3531226. (2013)
23. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. "The IntAct molecular interaction database in 2012". Nucleic Acids Res, 40(Database

issue), pp. D841-6. PubMed PMID: 22121220. Pubmed Central PMCID: 3245075. (2012)
24. Martha VS, Liu Z, Guo L, Su Z, Ye Y, Fang H, et al. "Constructing a robust protein-protein interaction network by integrating multiple public databases". BMC Bioinformatics, 12 Suppl 10, pp. S7. PubMed PMID: 22165958. Pubmed Central PMCID: 3236850. (2011)
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium". Nature genetics, 25(1), pp. 25-9. PubMed PMID: 10802651. Pubmed Central PMCID: 3037419. (2000)