

$$\frac{A_05}{47}$$



# NETWORK TOOLS AND APPLICATIONS IN BIOLOGY NETTAB-BBCC 2010 BIOLOGICAL WIKIS

Napoli, 29<sup>th</sup> november – 1<sup>st</sup> december, 2010

*Editors*

Angelo Facchiano  
Paolo Romano



Copyright © MMX  
ARACNE editrice S.r.l.

[www.aracneeditrice.it](http://www.aracneeditrice.it)  
[info@aracneeditrice.it](mailto:info@aracneeditrice.it)

via Raffaele Garofalo, 133/A-B  
00173 Roma  
(06) 93781065

isbn 978-88-548-3658-7

*I diritti di traduzione, di memorizzazione elettronica,  
di riproduzione e di adattamento anche parziale,  
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

*Non sono assolutamente consentite le fotocopie  
senza il permesso scritto dell'Editore.*

I edizione: novembre 2010

## **Co-chairs**

Angelo Facchiano, ISA, National Research Council, Avellino, Italy  
Paolo Romano, National Cancer Research Institute, Genoa, Italy

## **Scientific Committee**

Claudia Angelini, IAC, National Research Council, Napoli, Italy  
Roland Barriot, Laboratoire de Microbiologie et Génétique Moléculaires, Université Paul Sabatier - CNRS, Toulouse, France  
Alex Bateman, Wellcome Trust Sanger Institute, Hinxton, United Kingdom  
Dan Bolser, University of Dundee, Scotland, United Kingdom  
Philip E. Bourne, University of California, San Diego UCSD, La Jolla, USA  
Mike Carasio, Keygene N.V., Wageningen, The Netherlands  
Maria Luisa Chiusano, University of Napoli "Federico II", Napoli, Italy  
Antonio d'Acerno, ISA, National Research Council, Avellino, Italy  
Italia De Feis, IAC, National Research Council, Napoli, Italy  
Angelo Facchiano, ISA, National Research Council, Avellino, Italy  
Carl Herrmann, Université de la Méditerranée, Marseille, France  
Pascal Hingamp, Université de la Méditerranée, Marseille, France  
Eran Hodis, Weizmann Institute of Science, Rehovot, Israel  
Robert Hoehndorf, Department of Genetics, University of Cambridge, United Kingdom  
Robert Hoffmann, Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, USA  
Thomas Kelder, BiGCaT Bioinformatics, University of Maastricht, The Netherlands  
Roderic Page, Ecology and Evolutionary Biology, University of Glasgow, Scotland, United Kingdom  
Alexander Pico, The Gladstone Institutes, University of California, San Francisco, USA  
Jaime Prilusky, Bioinformatics, Weizmann Institute of Science, Rehovot, Israel  
Paolo Romano, Bioinformatics, National Cancer Research Institute, Italy  
Katharina Siorpaes, Semantic Technology Institute (STI), University of Innsbruck, Austria  
Andrew Su, Bioinformatics, Genomics Institute of the Novartis Research Foundation, San Diego, USA  
Joel Sussman, Weizmann Institute of Science, Rehovot, Israel  
Tim Vickers, Department of Molecular Microbiology, Washington University in St. Louis, St. Louis, USA  
Wyeth Wasserman, Department of Medical Genetics, University of British Columbia, Vancouver, Canada

## **Organizing Committee**

Claudia Angelini, IAC, National Research Council, Napoli, Italy  
Maria Luisa Chiusano, University of Napoli "Federico II", Napoli, Italy  
Italia De Feis, IAC, National Research Council, Napoli, Italy  
Angelo Facchiano, ISA, National Research Council, Avellino, Italy  
Paolo Romano, Bioinformatics, National Cancer Research Institute, Genoa, Italy

## **Supporting Institutions**

National Cancer Research Institute, Genova, Italy  
Italian Network for Oncology Bioinformatics (RNBIO)  
National Research Council (CNR):  
- IAC - Institute for Applied Mathematics "Mauro Picone"  
- ISA - Institute of Food Sciences  
- CNR - Bioinformatics Project

Oxford University Press

BITS – Italian Bioinformatics Society



# Index

NETTAB2010: Biological Wikis <i>Paolo Romano</i>	1
NETTAB and BBCC: a successful synergy ? <i>Angelo Facchiano</i>	3
Pros and Cons of Wikipedia for Scientists <i>Alex Bateman</i>	5
SNPedia <i>Mike Carasio</i>	7
TOPSAN: a collaborative annotation environment for structural genomics and beyond <i>Adam Godzik</i>	9
Collaborative publishing with authorship tracking and reputation system – WikiGenes <i>Robert Hoffmann</i>	11
The wiki-based Transcription Factor Encyclopedia and a model for robust community participation <i>Wyeth W. Wasserman and Dimas Yusuf</i>	13
Extending Mediawiki for community annotation <i>Daniel Renfro, Deborah A. Siegele, James C. Hu</i>	21
The Bioinformatics Italian Society (BITS) <i>Manuela Helmer Citterich</i>	25
WikiPathways: community-based curation of biological pathways <i>Alexander R. Pico, Thomas Kelder, Martijn P. van Iersel, Kristina Hanspers, Bruce R. Conklin and Chris Evelo</i>	27
Demonstration of a citation-enabled workflow using the ConceptWiki triple based approach <i>Christine Chichester , Hailiang Mei , Kees Burger , Barend Mons</i>	33
Using A General-Purpose Database Wiki for Biological Database Curation <i>Heiko Müller, Sam Lindley, Joanna Sharman. Tony Harmar , James Cheney, and Peter Buneman</i>	37
PDBWiki: success or failure? Factors for successful community annotation projects <i>Dan M Bolser, Jose M Duarte, Jong Bhak, Henning Stehr</i>	45
The Gene Wiki: Achieving critical mass and mining for novel annotations <i>Andrew Su</i>	51

SBML2SMW: Bridging System Biology with Semantic Web Technologies for Biomedical Knowledge Acquisition and Hypothesis Elicitation <i>Tobias Mathäß, Peter Haase, Hiroaki Kitano, Luca Toldo</i>	53
Repository for the management of tandem MS data <i>Federica Viti, Ivan Merelli, Dario Di Silvestre, Pietro Brunetti, Luciano Milanese, Pierluigi Mauri</i>	59
Mathematical Models for Feature Selection and their Application to Bioinformatics <i>Paola Bertolazzi, Giovanni Felici, Paola Festa</i>	65
Toward an Improved Combinatoric Algorithm <i>Ekaterina Nosova, Francesco Napolitano, Giancarlo Raiconi, Roberto Tagliaferri, Sergio Coccozza, Roberto Amato, Gennaro Miele</i>	71
RNA-seq: from computational challenges to biological insights <i>Valerio Costa, Claudia Angelini, Luciana D'Apice, Margherita Mutarelli, Amelia Casamassimi, Marianna Aprile, Roberta Esposito, Luigi Leone, Aldo Donizetti, Stefania Crispi, Piergiuseppe De Berardinis, Claudio Napoli, Antonio Baldini and Alfredo Ciccodicola</i>	77
Eliciting Fuzzy Knowledge from the PIMA Dastate <i>Antonio d'Acerno, Giuseppe De Pietro, Massimo Esposito</i>	85
Genome Duplication and Gene Annotation: an Example for a Reference Plant Species <i>Alessandra Vigilante, Mara Sangiovanni, Chiara Colantuono, Luigi Frusciantè and Maria Luisa Chiusano</i>	91
Computational prediction and experimental analysis of RNA structures <i>Sonia Varriale, Stefano Giacomelli, Maria Rosaria Coscia, Claudio Rivetti, Umberto Oreste</i>	95
<i>POSTERS</i>	
Micronutrients in WikiPathways database <i>D. Rivero Guedez, J. Bouwman, S. Wopereis, B. van Ommen and D. Cavalieri</i>	101
Mirroring and Synchronization of Biological Wikis: lessons from the GeneWiki <i>Salvatore Loguercio, Benjamin Good, Andrew I. Su</i>	105
A Semantic Wiki for Genotype/Phenotype Annotation in Plant Species <i>Justin Preece, Justin Elser, Pankaj Jaiswal</i>	109
Free System Biology Collaborative Workbench: Design and implementation <i>Abdul Mateen Rajput, Luca Toldo</i>	113
A Microfluidic Device for the Real-Time control of a Synthetic Gene Network <i>E. Orabona, F. Menolascina, L. De Stefano, D. di Bernardo</i>	119



Structural stability and dynamical properties of steric zipper polymorphs: implications for prion strains <i>Francesca Stanzione, Alfonso De Simone, Luciana Esposito, Luigi Vitagliano</i>	123
Molecular Dynamics Simulations of Phenylalanine Hydroxylase Natural Mutants Leading to Hyperphenylalaninemia <i>F. Fraternali, C. Carluccio, A. Fornili, A. Zagari, F. Salvatore</i>	127
Molecular modeling of the ligand binding domain of human Aryl hydrocarbon Receptor (AhR) and its interaction with 2,3,7,8-TCDD <i>Maria Salzano, Anna Marabotti, and Angelo Facchiano</i>	131
Different water permeability in orthologous aquaporins: the case of fish lens aquaporin <i>Luisa Calvanese, Marialuisa Pellegrini-Calace, Romina Oliva</i>	133
CMPC: A Web Application to Visualize Residues-Residue Contacts at the Interface of Protein Complexes <i>Anna Vangone, Romina Oliva, Luigi Cavallo</i>	137
Visual Characterization of Biomedical Texts With Word Entropy <i>Mirko Degli Esposti, Roxana Danger, Paolo Rosso, Sandra García Blasco</i>	139
REPAIRtoire – a database of DNA repair pathways <i>Kaja Milanowska, Joanna Krwawicz, Grzegorz Papaj, Jan Kosiński, Katarzyna Poleszak, Justyna Lesiak, Ewelina Osińska, Kristian Rother, Janusz M. Bujnicki</i>	143
A Statechart based representation for SBML descriptions <i>Fabio Fioravanti, Manuela Helmer Citterich, Enrico Nardelli</i>	147
Italian Network for Oncology Bioinformatics - RNBIO <i>Paolo Romano</i>	151
An Integrated Microbial Wiki <i>Kaushlesh Kumar Yadav, Satyendra kumar Singh and Anurag chaurasia</i>	153



## Joint NETTAB 2010 and BBCC 2010 workshops on Biological Wikis: rationale and organization

Paolo Romano

*Bioinformatics, National Cancer Research Institute of Genova, Italy*

*paolo.romano@istge.it*

It is now clear that wiki systems offer a variety of advantages for the management of biological data and information.

Some of the specific aims of wikis for biology (bio-wikis) include:

- Collaborative development and sharing of knowledge
- Collaborative annotation of database contents
- Collaborative creation of database contents

The collaborative development and sharing of documentation and knowledge allows communities to promote, exploit, discuss and reach consensus on procedures, data, experiences, news, and other varied information. Indeed valuable expertise and interests in special topics are usually distributed, and are rarely concentrated in a unique site or research group.

The collaborative annotation of biological databases is increasingly under consideration because extended and accurate curation of an ever increasing volume of data is extremely costly and time consuming. Such distributed networks can help to improve and extend databases curation beyond that which is possible with, typically, limited numbers of curation staff. It allows users to contribute their expertise and observations independently of database central organizations. Although the contents of the database are collaboratively annotated, the underlying databases themselves are left unchanged.

Furthermore, collaborative creation of community databases can capture emerging structure in rapidly developing fields. These database-wikis are indexes of

biologically relevant data that emerge from focused and rapidly developing communities. They form a stopgap between unstructured discussion in fora and on mailing lists and the 'mature' databases that emerge subsequently. Hence they can be useful for discerning trends and promoting best practices, as well as collect unique and timely information.

In the future, collaborative 'wiki' updates of almost all database contents can be envisaged. However several important issues need to be addressed:

- How reliable are user contributions?
- What format should annotations take?
- How is this information feed back into 'authoritative' databases?

Procedures should be implemented to support assessment of users' contributions, extraction of structured data from annotations, uptake of original data into reference databases.

The authoritativeness of contributions is essential: how can quality be assessed? How do we deal with contributions from casual end users, since these are usually not considered adequate when compared with professional annotation at service centers. Can the open edition model of many wiki systems, e.g. Wikipedia, be as successful in life sciences? Alternatively, should the usual criteria for assessing quality of research be applied, including peer-evaluation of contributions and identification of users?

Crediting users for their contributions (e.g. authorship assignment) may be necessary to stimulate the broadest possible participation. Which further

benefits may be imagined to acknowledge user investment (subscriptions to services, journals, ...)?

Special features may be required to cater for the specificity of biological data: textual information is only a small part of biological data, how can the numerous and heterogeneous biological data types be catered for, e.g. images, plots, diagrams? An adaptation of current wiki systems is needed. A survey of existing systems as well as current development efforts will be useful, so that possible synergies can be identified and supported, as well as ensure a coherent set of interoperable biological wikis.

Topics addressed by the workshop therefore were: i) development tools, including of course Wikimedia and related extensions, Semantic Wikis, and Wiki-coupled CMSs, ii) arising issues for the biomedical domain, like authoritativeness of contributions and sites, quality assessment, users acknowledgements, stimulation of quality contributions, authorships management and reward, scientific production value for contributions, and management of bioinformatics data types, iii) wikis and collaborative systems for genomics, proteomics, metabolomics, and all other – omics disciplines, proteins analysis and visualization, gene and proteins

interactions, metabolic pathways, and oncology research.

The joint NETTAB and BBCC 2010 workshop on "Biological Wikis" promises to be a great meeting for all researchers involved in the exploitation of wikis in biology. It will be possible to discuss ideas, and doubts, with such scientists as Alex Bateman, Alexander Pico, Andrew Su, Dan Bolser, Robert Hoffmann, Thomas Kelder, who accepted to give invited lectures and tutorials, and Mike Cariaso, Adam Godzik, Luca Toldo, Wyeth Wasserman, Daniel Renfro and other who enthusiastically joined the workshop.

It's a great chance to follow smart tutorials and lectures on some of the best known biological wikis, like WikiPathways, WikiGenes, PDBWiki, Gene Wiki, TOPSAN, and on the proficient use of Wikipedia and Semantic Wiki.

And, of course, a great occasion to enjoy Italian lifestyle....

On behalf of my Co-Chair, Angelo Facchiano, and of the many members of the Scientific Committee, to whom we renovate our grateful thanks, we wish all participants, and all readers of this Abstracts' book, a nice and useful workshop and a pleasant stay in Naples.

## NETTAB and BBCC: a successful synergy ?

Angelo Facchiano

*Institute of Food Sciences, National Research Council, Avellino, Italy*

*angelo.facchiano@isa.cnr.it*

NETTAB-BBCC 2010 represents the meet of two communities of researchers and students, with a common interest for bioinformatics. The international workshop NETTAB, and its focus on a frontier application of new informatics technologies in biological sciences, meets the bioinformatics researchers from Campania region. Is it a good idea? This question is in my mind since Paolo and me discussed about this opportunity.

NETTAB workshops series started in 2001, and so this year it reaches its 10th edition. Paolo Romano is the inventor of NETTAB series, and works as the chair and promoter of each edition, for which he involves every year new colleagues to focus on a different topic, to explore the new frontiers of bioinformatics. A different location is chosen every year, with a local team of researchers involved in the scientific and logistic organization, together with an international scientific committee. I was involved in different editions of NETTAB as a member of the scientific committee, and also for the organization of the edition of 2003 and 2005 as a co-chair with Paolo and other colleagues.

The experience in organizing two editions of NETTAB convinced me to involve myself in organizing workshops and conferences, with the aim of creating opportunities for opening the mind to new scenarios, to exchange ideas, opinions and suggestions, to start new collaborations and projects.

In Campania, many research teams are involved in bioinformatics studies, and many others in biomedical area, as well as in mathematics and informatics. However,

no specific event was regularly organized to give them an opportunity of meet and discuss as an unique community. Why not to organize an annual meeting of bioinformatics for the Campania region researchers? This idea was in my mind when I was involved with other colleagues in the organization of the 2007 edition of the annual meeting of the Italian Society of Bioinformatics (Napoli, April 2007). So, the idea was reinforced by the need of informing and attracting researchers from the local community, in view of the national meeting.

This was the birth of the “Bioinformatica e Biologia Computazionale in Campania - BBCC” meeting series. In December 2006, the first edition registered the participation of about 100 researchers and students, an unexpected success and a strong motivation to repeat the meeting as an annual appointment. From 2006 to 2009, the meeting has been held in Avellino, with a similar number of participants every year. The formula was simple: a single-day event, speakers on very different topics, the declared aim of forcing everyone to know what happens, in some cases, on the other side of the same building. And, last but not least, a special attention to students. During these years, many initiatives were organized in the region, with the involvement of different research teams which participate in the BBCC meetings. I feel that this means that the community is really working (and growing up) together, and I consider it a success of BBCC meeting series, and a strong motivation to continue in this initiative, looking for new interesting opportunities. In the past years, colleagues from other Italian regions were

involved, to open our community and search for new suggestions. This year BBCC is at its fifth edition, and the joining with NETTAB offers the opportunity to meet the international community of the Wiki world.

The organization of a three-days international meeting has suggested a different location: Napoli, of course, more easy to reach for foreigners than Avellino.

But, I am still asking to myself: is it a good idea to joint a specialized international meeting, as NETTAB, and a local meeting with a open topics format, as BBCC ?

I hope the answer will be: yes. Because the specialized topic needs to be spread

around. And because the local community needs to meet the international one, and grow up in experience, motivation, ideas.

The last paragraph is to thank the co-organizers of this event, Claudia Angelini, Maria Luisa Chiusano, and Italia De Feis, for helping me in this hard job, despite hundreds of difficulties. And a special thank to Paolo Romano, for sharing with me the managing and all the efforts of the organization.

Enjoy the meeting!

## **Pros and Cons of Wikipedia for Scientists**

Alex Bateman

*Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom*

*agb@sanger.ac.uk*

Wikipedia is the most successful community annotation project of all time. But can it be used successfully for scientific information? I will discuss the strengths as well as the

weaknesses of the Wikipedia approach. In particular I will discuss the issue of getting scientists engaged in editing Wikipedia as well as show examples of spectacular failures.





## SNPedia

Michael Carasio

*Keygene N.V., Wageningen, The Netherlands*

*cariaso@gmail.com*

This talk will summarize SNPedia.com and how it can be used to interpret a human genome.

This unique resource is used by researchers, clinicians and the general public. Anyone with \$500 can now learn a great deal about their own genome. SNPedia features such as Genosets (boolean combinations of alleles) and Magnitude (a subjective measure of community interest) enables annotating

multiple genotyping platforms, and is equally suitable for non-SNP types of genetic variation.

SNPedia content is creative commons licensed and semantic web accessible with RDFs.

A wiki user community helps to verify and improve the annotations. The tools which perform this analysis are user friendly desktop apps freely available for non-commercial use.



## **TOPSAN: a collaborative annotation environment for structural genomics and beyond**

Adam Godzik

*SanfordBurnham Medical Research Institute, La Jolla, California*  
*adam@sanfordburnham.org*

TOPSAN was developed at JCSG, of the NIH Protein Structure Initiative centers to help initiate instantaneous collaborations of distributed participants to work together in exchanging ideas and data on novel proteins. Borrowing some elements of Wiki (distributed, voluntary contributors), peer reviewed manuscripts (quality control, trackable authorship) and a database (access to experimental data from

automated production pipeline and annotations parsed from other annotation resources) we succeeded in developing such collaborations leading to formal publications for almost 100 proteins, but another 900 are still waiting for a serious effort to describe them and enter them into the mainstream of biological literature. TOPSAN is available at <http://topsan.org>



## **Collaborative publishing with authorship tracking and reputation system – WikiGenes**

Robert Hoffmann

*Computational Biology Center, cBIO, Memorial Sloan-Kettering Cancer Center, MSKCC, New York, USA  
hoffmann@cbio.mskcc.org*

WikiGenes is the first collaborative publishing system to unite traditional and collaborative publishing based on exclusive authorship tracking technology. Whereas Wikipedia and conventional wikis have a limited realm of encyclopaedic knowledge by anonymous contributors, the authorship tracking technology in WikiGenes links every sentence and every word to its author, enabling experts to share their latest and original thoughts. For instance, experts from around the world can create authoritative review articles on highly specific subjects years before this knowledge would enter traditional forms of publication.

The way scientific insights are published and communicated has remained essentially unchanged since the days of Charles Darwin. As a result, facts pertaining to specific genes, chemicals and pathologies are scattered over hundreds and even thousands of different articles and

journals. Unfortunately, the wisdom of crowds, so useful in gathering well-established knowledge, is inadequate to integrate these highly specific bits of knowledge into a bigger picture.

The authorship tracking technology in WikiGenes makes it possible that originators of information -scientists themselves- can publish their insights while contributing at the same time to a body of structured, semantically enriched and integrated knowledge. In WikiGenes, authors who invest time and knowledge in their contributions are given due credit, and readers can always know the author of any part of a collaborative article. Instead of dispersing new insights in an arbitrary way over thousands of scientific journals, experts can publish their insights in context, like pieces of a jigsaw puzzle.

WikiGenes is open access and available online at <http://www.wikigenes.org/>.