

National Bioinformatics Institute *Core Unit*

66

Scientific Report 2010 *cnio*



Alfonso Valencia

Unit Head

Alfonso Valencia is a biologist with formal training in population genetics and biophysics which he received from the *Universidad Complutense de Madrid*. He was awarded his PhD in 1988 at the *Universidad Autónoma de Madrid*.

He was a Visiting Scientist at the American Red Cross Laboratory in 1987 and from 1989-1994 was a Postdoctoral Fellow at the laboratory of C. Sander at the European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.

In 1994 Alfonso Valencia set up the Protein Design Group at the *Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CSIC)* in Madrid where he was appointed as Research Professor in 2005.

He is a Member of the European Molecular Biology Organisation (EMBO), Founder and former Vice President of the International Society for Computational Biology where he has been Chair of the Systems Biology and/or Text Mining Tracks of the main Computational Biology Annual Conference (ISMB) since 2003. He was honoured as ISCB-Fellow in 2010.

Alfonso Valencia serves on the Scientific Advisory Board of the European Molecular Biology Laboratory; the Swiss Institute for Bioinformatics, Biozentrum, Basel; the INTERPRO database; the Spanish Grant Evaluation Agency (ANEP); as well as the Steering Committee of the European Science Foundation Programme on Functional Genomics (2006-2011).

Alfonso Valencia is Co-Executive Editor of *Bioinformatics*, serves on the Editorial Board of *EMBO Journal* and *EMBO Reports*, among others. He is the Director of the Spanish National Bioinformatics Institute (INB).

Summary

The Spanish National Bioinformatics Institute, *Instituto Nacional de Bioinformática (INB)*, has now transferred from the *Genoma España* Foundation to the *Instituto de Salud Carlos III*, where it will remain part of the central service's infrastructure.

The INB will continue to provide bioinformatics support to Genome projects prioritising those related to human health. The virtual structure of the Institute integrates 9 nodes covering the main areas of Bioinformatics from gene discovery to drug design. The CNIO hosts the Central Node as well as that specialised in genome scale protein function annotation.

Main Objectives

- Generate and supply bioinformatic solutions to genome projects with particular emphasis on those relating to human health
- Collaborate with national and international genomic projects
- Support the development of bioinformatics and computational biology in Spain





Technicians: Andrés Cañada, Guillermo Comesaña, Víctor de la Torre, José M. Fernández, Allan Orozco and José M. Rodríguez.

Highlights

During this year the INB Core Unit has focused on the following function annotation systems:

- TreeDet, a system for the prediction of functional sites based on sequence information that has incorporated new methods (Rausell A. et al., *PNAS* 2010).
- FireDB/FireStar, a database and web interface for the extrapolation of binding sites from known protein structures that has been used for the predictions of protein binding sites in the context of the CASP challenge (Asilomar Conference, California USA, December 2010). FireDB/FireStar is also part of the APRIS systems. APRIS is a system for the annotation of principal splicing variants developed under the ENCODE project. APRIS results are used for the annotation of the human genome in collaboration with the HAVANA annotation team at the Sanger Institute (UK).
- The BioCreative Metaserver (BCMS), a text system for the annotation of protein interactions in biological texts by collecting text-mining results from distributed servers. The BCMS metaserver is the core of the BioCreative II.5 challenge (Leitner F. et al., *Nat Biotechnol* 2010, *IEEE/ACM Trans Comput Biol Bioinform* 2010) and of the BioCreative III edition (evaluation workshop, Bethesda USA, September 2010).
- GOPHER, a system to handle the automatic evaluation of bioinformatics methods. This first implementation addresses the evaluation of function prediction systems. In the future the system will be extended to the evaluation of bioinformatics methods in other areas.

Overall, the INB Unit aims to integrate these and other annotation facilities within an integral platform accessible to all kinds of users. Currently MADAS is being used in the European TARPOL, pSYSMO and EMERGENCE projects. Some MADAS functionalities are integrated in the MIT "Repository of Parts", a central resource for the synthetic biology research community.

During 2010 the INB Unit has transferred a collection of more than 400 INB web services to the centralised European registry, coordinated by the EMBRACE project (Pettifer S. et al., *Nuclac Acids Res* 2010) and has also contributed to the efforts of the international community to integrate web resources (Katayama T. et al., *J Biomed Semantics* 2010). In terms of databases, the INB Unit has continued to implement PANGENES – an application to handle cancer epidemiological information developed in collaboration with the CNIO Genetic and Molecular Epidemiology Group. INB also provides support for the database structure of a Spanish project within the International Cancer Genome Consortium (ICGC, *Nature* 2010).

Finally, the INB Unit provides technical support to the CNIO's Next-Generation Sequencing (NGS) pipeline, particularly in massive data handling and organisation.

Publications

International Cancer Genome Consortium (2010). International network of cancer genome projects. *Nature* 464, 993-998.

Leitner F, Chatr-aryamontri A, Mardis SA, Ceol A, Krallinger M, Licata L, Hirschman L, Cesareni G, Valencia A (2010). The FEBS Letters/BioCreative II.5 experiment: making biological information accessible. *Nat Biotechnol* 28, 897-899.

Rausell A, Juan D, Pazos F, Valencia A (2010). Protein interactions and ligand binding: from protein subfamilies to functional specificity. *Proc Natl Acad Sci USA* 107, 1995-2000.

Pettifer S, et al. (2010). The EMBRACE web service collection. *Nucleic Acids Res* 38, W683-W688.

Leitner F, Mardis SA, Krallinger M, Cesareni G, Hirschman LA, Valencia A (2010). An Overview of BioCreative II.5. *IEEE/ACM Trans Comput Biol Bioinform* 7, 385-399.

Katayama T, et al. (2010). The DBCLS BioHackathon: standardization and interoperability for bioinformatics web services and workflows. The DBCLS BioHackathon Consortium. *J Biomed Semantics* 1, 8.

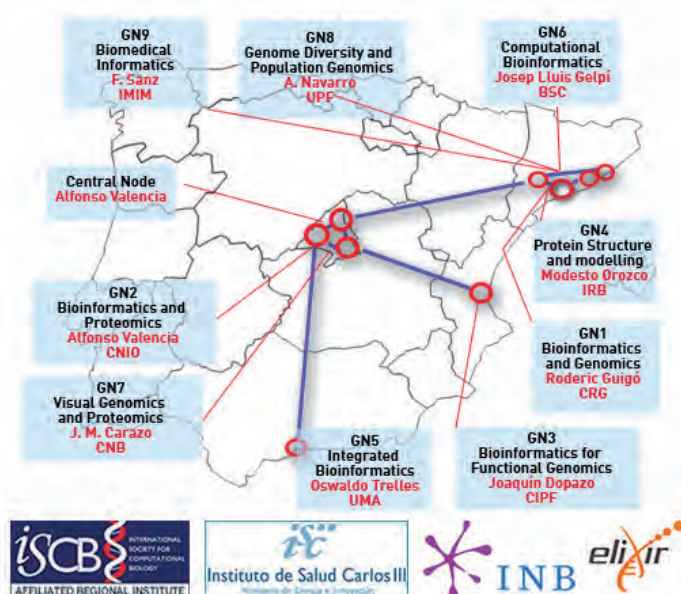


Figure: Organisation of the Spanish National Bioinformatics Institute (INB).