NATIONAL BIOINFORMATICS INSTITUTE UNIT

Salvador J. Capella Gutierrez (since June)
Unit Head

Technicians
Andrés Cañada (TS)*, José M. Fernández (TS)*, José M. Rodríguez (TS)*

**Titulado Superior* (Advanced Degree)

OVERVIEW

The Spanish National Bioinformatics Institute (*Instituto Nacional de Bioinformática, INB*) is a component of the National Infrastructure of Biomolecular and Bioinformatics Resources Platform (*Plataforma en Red de Recursos Biomoleculares y Bioinformáticos, PRB2*) of the Spanish National Institute of Health Carlos III (*Instituto de Salud Carlos III, ISCIII*). The INB is the Spanish Node of ELIXIR, the permanent European Infrastructure for Life Sciences. The INB is composed of 10 working nodes distributed across 9 different research centres. The INB Unit at the CNIO undertakes the coordination of the institute. As the coordination node, the goals of the INB Unit are to:

→ Coordinate the Spanish participation in ELIXIR. Promote the implementation and adoption of ELIXIR guidelines among the Spanish bioinformatics community.

"The INB Unit has actively participated in the management of data portals for big research consortia like BLUEPRINT, ICGC, and PanCancer, aiming to understand the genetic bases of cancer."

→ Design, promote and ensure the execution of the INB's scientific-technical and training programmes, undertaken with the support of all nodes.

→ Promote the collaboration between INB nodes and third parties, including research consortia, other research infrastructures, small and medium enterprises (SMEs), and the industry.

RESEARCH HIGHLIGHTS

Despite this Unit's coordination role, the INB actively participates in different Work Packages (WP) of the ELIXIR-EXCELERATE programme. It is also involved in other major projects such as RD-Connect, BLUEPRINT and eTOX. The Unit's contribution can be divided into three main areas:

Data resources and Bio-computing

The storage and processing of data have become fundamental tasks within almost all of the current research projects. Through a collaboration model, the Unit participates in several research projects studying the data requirements and developing solutions to store and process the data. An example of this is the BLUEPRINT data-portal (http://dcc.blueprint-epigenome.eu). BLUEPRINT is a high impact FP7 project aimed at producing epigenomes of haemopoetic cell lines from healthy and non-healthy human donors. In the current version, the data portal provides an epigenomic analysis, obtained from 1,019 samples, to the scientific community. Their associated epigenomes are characterised by: gene and transcript expression (from RNA-Seq experiments), hyper and hypo methylated regions (derived from WGBS experiments), chromatin accessibility (DNAse-Seq), and 7 Histone marks binding activity (ChIP-Seq). Recently, a scientific article was published illustrating the possibilities offered by this portal (Fernandez JM *et al*, 2016).

Infrastructure development

Within the infrastructure development aspect, there is a clear focus on developing a **text-mining infrastructure** for the processing of biomedical texts. The LiMTox system (http://limtox.bioinfo.cnio.es) is the first text mining approach designed to extract associations between compounds and a particular toxicological end point at various levels of granularity and evidence types, all inspired by the content of toxicology reports. During this time, a second end-point (http://melanomamine.bioinfo.cipf.es) has been built using the same system and focusing on the study of different aspects of melanomas.

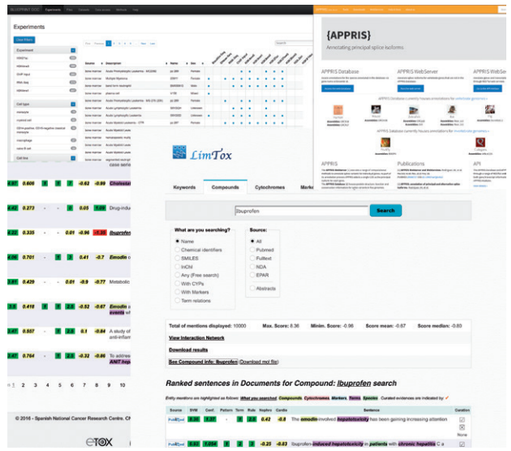

Figure A collage of the different services developed by the INB Unit. Top left shows the experimental information available at BLUEPRINT. Top right shows the information available at the APPRIS home-page. Bottom panels feature information found for *Ibuprofen* using the LimTox system.

End users applications and services

The Unit actively contributes to the creation of an integrated platform that connects databases, registries, biobanks and clinical bioinformatics for research on rare diseases. RD-Connect (http://rd-connect.eu) allows researchers and clinicians to explore the possible genetic causes of these diseases by combining genomic data with phenotypic information from patients across Europe, in a clear attempt to have enough statistical power to support findings. The INB unit has also developed APPRIS (http://appris.bioinfo.cnio.es) and keeps improving it in order to annotate genes and identify the principal isoform of every single gene. Currently, the GENCODE consortium uses APPRIS to annotate the principal human and mouse isoforms. ■

PUBLICATIONS

- Chen L *et al.* (incl. Fernandez JM, Rico D, Valencia A) (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* 167, 1398–1414.
- Abascal F *et al.* (incl. Capella-Gutierrez S, Valencia A) (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology* 17, 251.
- Ecker S *et al.* (incl. Fernandez JM, Rico D, Valencia A) (2016). Genome-wide Analysis of Differential Transcriptional and Epigenetic Variability Across Human Immune Cell Types. *BioRxiv*, 83246.
- Fernández JM, de la Torre V, Richardson D, Royo R, Puiggròs M, Moncunill V, Fragkogianni S, Clarke L; BLUEPRINT Consortium., Flicek P, Rico D, Torrents D, Carrillo de Santa Pau E, Valencia A (2016). The BLUEPRINT Data Analysis Portal. *Cell Syst* 3,491-495.

PUBLICATIONS AT OTHER INSTITUTIONS

- Altenhoff, AM *et al.* (incl. Capella-Gutierrez S) (2016). Standardized benchmarking in the quest for orthologs. *Nat Methods* 13, 425–430.
- Vlasova A *et al.* (incl. Capella-Gutierrez S) (2016). Genome and Transcriptome Analysis of the Mesoamerican Common Bean and the Role of Gene Duplications in Establishing Tissue and Temporal Specialization of Genes. *Genome Biol* 17, 32.
- Sebé-Pedros A, Peña MI, Capella-Gutierrez S, Antó M, Gabaldón T, Ruíz-Trillo I & Sabidó E. (2016). High-throughput proteomics reveals the unicellular roots of animal phosphosignaling and cell differentiation. *Dev Cell* 39, 186-197.