

# Curriculum Vitae

**Name:** Daniel Sobral

**Birth Date:** 18th August 1979

**Nationality:** Portuguese

**E-mail:** [daniel.elanus.sobral@gmail.com](mailto:daniel.elanus.sobral@gmail.com)

## **Professional Experience**

**2014-:** Head of the Bioinformatics Unit at Instituto Gulbenkian de Ciencia.

**2012-2013:** Next Generation Sequencing data analyst in the Bioinformatics Unit, at Instituto Gulbenkian de Ciencia. I have been collaborating in several research projects within the institute. Examples of projects include the analysis of splice defects in a *Drosophila melanogaster* mutant using RNA-Seq, detecting genomic variation in *E. Coli* resequencing data, differential expression of miRNAs in *Arabidopsis thaliana* using miRNA-Seq, and analysis of histone dynamics using ChIP-Seq in a human cell line.

**2010-2011:** EnSEMBL developer in the Functional Genomics Group at the European Bioinformatics Institute. I've been developing a pipeline to integrate \*-seq data towards building genomic annotations of putative regulatory regions within the ENSEMBL project (Flicek et al., 2011). I handled raw sequence data from large-scale projects like ENCODE (Dunham et al., 2012), performed basic quality tests (eg. read quality, duplicates, etc), aligned the reads to the genome using late-generation aligners and processed the alignments to find regions in the genome enriched in reads (peak-calling). I also performed quality checks on the obtained peaks (eg. by looking for appropriate TF motifs). I've worked within a large compute farm, using shell scripting, Perl and MySQL. I've also updated Ensembl databases and API supporting new data.

**2001-2004: Research Assistant, Institute for Systems and Computer Engineering: Research and Development (INESC-ID), within the Intelligent Agents and Synthetic Characters Group (GAIPS). Projects:** Virtual ICT for Empathic Characters (Victec).

**2002-2004: Monitor (Lab Tutor),** Computer Science Department, IST.

**Nov. 2002: Lab Tutor,** Matlab Introductory Course, Mechanical Engineering Department, IST.

## **Workshops, Courses, Student Supervision**

**2012-:** Lectures on Bioinformatics in several Masters and Doctoral Programs

**April 2012:** Organized an NGS Sequencing Workshop at IGC.

**2010-2011:** Teaching the Ensembl Regulation section within Ensembl Developers Workshops

**2008-2009:** Supervised a Masters student at Patrick Lemaire's lab

**April 2002:** Member of the organizing committee of the IXth Computer Science Week, IST.

## **Education**

**2005-2009:** PhD in Bioinformatics at the *University of Aix-Marseille II*, under the supervision of Dr. Patrick Lemaire. Thesis subject: "Biological Data Integration. Towards a better global understanding of ascidian development". My work consisted first in data mining of biological data from ANISEED, the ascidian model system database. Following this analysis I integrated the data with information from the zebrafish database (ZFIN), in order to gain some insights regarding the evolution of gene expression programs within these chordate species (Sobral et al, 2009). In parallel, I extended ANISEED with knowledge structures and web interfaces to represent information regarding transcriptional gene regulation (Tassy et al, 2010). I have also collaborated in the development of a gene model collection based on a large *C. intestinalis* cDNA library. In this project I obtained functional annotations for the clustered cDNA-based models (eg. ortholog predictions, GO enrichment analysis, etc.). Finally, I have also contributed to a project running in the lab that aimed at extracting the DNA binding affinities for most *Ciona intestinalis* transcription factors via a robotized SELEX assay.

**2004-2005:** First Year Graduate Student – Gulbenkian Doctoral Program. This year consisted in weekly theoretical sessions covering many aspects of biology, including bioinformatics and computational biology.

**2002-2004:** MSc in Artificial Intelligence and Multimedia, at IST: "Development of Web-Based Educational 3D Virtual Storytelling Environments for Children". This work consisted in representing information regarding a specific domain (social coping for children). It also involved a lot of work in the design of Human-Computer Interfaces (web interfaces, physical interfaces like an electronic plush-toy) and programming (mostly in Java) to build virtual environments to run simulated social situations for children, in the context of a large European project.

**2002:** Licenciatura in Computer Engineering (five year degree), *Instituto Superior Tecnico (IST)*, Lisbon. Specialization in Artificial Intelligence.

### **Conferences, Workshops, Courses Attended**

**4-5 July 2011:** Advanced RNA-Seq and ChIP-Seq data analysis with Bioconductor.

**6-7 August 2009:** GMOD community meeting.

**21-25 June 2009:** 5<sup>th</sup> International Tunicate Meeting, Okinawa, Japan. Co-chaired the session on Genomics and Gene Networks.

**June 2007:** 4<sup>th</sup> International Tunicate Meeting, Villefranche-sur-Mer, France. Oral presentation: *"ANISEED data integration. Towards a global understanding of ascidian development"*.

**7-11 May 2007:** Workshop: Bioinformatics III, Ghent University, Belgium.

**May 2003:** World Education Market (WEM). Exhibitor in the Research Pavilion representing the Victec Project, Lisbon, Portugal; presented the "Sentoy" and the "Fearnot!" application.

### **Publications**

The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps.

Barroso-Batista J, Sousa A, Lourenço M, Bergman ML, **Sobral D**, Demengeot J, Xavier KB, Gordo I. PLoS Genet. 2014 Mar 6;10(3)

ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74.

Flicek et al., Ensembl 2011. Nucleic Acids Res. 2011 Jan;39(Database issue):D800-6. Epub 2010 Nov 2

Tassy O\*, Dauga D\*, Daian F\*, **Sobral D\***, Robin F, Khoueiry P, Salgado D, Fox V, Caillol D, Schiappa R, Laporte B, Rios A, Luxardi G, Kusakabe T, Joly JS, Darras S, Christiaen L, Contensin M, Auger H, Lamy C, Hudson C, Rothbacher U, Gilchrist MJ, Makabe KW, Hotta K, Fujiwara S, Satoh N, Satou Y, Lemaire P. The ANISEED database: Digital representation, formalization, and elucidation of a chordate developmental program. Genome Res. 2010 Oct;20(10):1459-68. Epub 2010 Jul 20.

**Sobral D**, Tassy O, Lemaire P. Highly divergent gene expression programs can lead to similar chordate larval body plans. Curr Biol. 2009 Dec 15;19(23):2014-9.

### **Other activities**

Webmaster of Grupo Flamingo, a non-governmental environment awareness association. I have created the group's domain, installed and configured Joomla in a shared server. I have also slightly tweaked some of Joomla's php code to adapt it to the group's needs.

### **Contacts for Recommendations**

Dr. Patrick Lemaire: [lemaire@ibdm1.univ-mrs.fr](mailto:lemaire@ibdm1.univ-mrs.fr) (my PhD supervisor)

Dr. Ian Dunham: [dunham@ebi.ac.uk](mailto:dunham@ebi.ac.uk) (my supervisor at Ensembl)

Dr. José Pereira-Leal: [jleal@igc.gulbenkian.pt](mailto:jleal@igc.gulbenkian.pt) (my supervisor at IGC)